ASTES

Advances in Science, Technology & Engineering Systems Journal

Special Issue

Innovation in Computing, Engineering Science & Technology

2023

www.astesj.com
ISSN: 2415-6698

EDITORIAL BOARD
(Special Issue)

Editor-in-Chief

Prof. Passerini Kazmerski
Pritzker School of Molecular Engineering, University of Chicago, USA

Guest Editors

Prof. Wang Xiu Ying
Chongqing University, China

Prof. Yu Xiao Yan
Chongqing Normal University, China

Prof. María Jesús Espinosa Trujillo
Universidad Tecnológica Metropolitana, Mexico

Prof. Ahmad Yusairi Bani Hashim
Universiti Teknikal Malaysia Melaka, Malaysia

Prof. Mohamed Abdelaziz Hassan Eleiwa
University of Hail, KSA

Prof. Nicolae Tudoroiu
John Abbott College, Canada

Editorial

The Special Issue on Innovation in Computing, Engineering Science & Technology (2023) in the *Advances in Science, Technology and Engineering Systems Journal (ASTES Journal)* reflects the continued evolution of technological innovation driven by the seamless integration of computational intelligence and engineering advancements. As digital transformation accelerates across industries, the convergence of computing and engineering disciplines has become a defining force in shaping modern scientific inquiry and practical applications. This issue presents a collection of high-quality contributions that highlight emerging trends, novel methodologies, and transformative solutions across interconnected domains.

A central focus of this issue is the advancement of intelligent, data-driven systems that are redefining engineering processes and decision-making frameworks. Several contributions explore the application of artificial intelligence, machine learning, and advanced analytics in optimizing system performance, enhancing automation, and enabling predictive and adaptive capabilities. These innovations are applied across a wide range of sectors, including smart manufacturing, communication networks, robotics, and intelligent infrastructure, demonstrating the versatility and impact of computational approaches in engineering science.

The issue also underscores progress in core engineering innovations, particularly in areas such as sustainable technologies, advanced materials, and energy-efficient system design. Authors present research that addresses critical global challenges, including resource optimization, environmental sustainability, and resilient infrastructure development. The integration of computing technologies with engineering systems is especially evident in emerging paradigms such as the Internet of Things (IoT), cyber-physical systems, and next-generation communication architectures, where interdisciplinary collaboration is essential for achieving scalable and robust solutions.

Methodological diversity remains a defining strength of this special issue. The included studies employ a combination of theoretical modeling, simulation techniques, experimental validation, and applied case studies. This comprehensive approach ensures that the research not only contributes to academic knowledge but also offers practical insights for industrial implementation and policy development. Many of the contributions emphasize system scalability, efficiency, and adaptability, reflecting the evolving demands of modern technological environments.

The 2023 context provides a backdrop of rapid innovation, increased reliance on digital ecosystems, and a growing emphasis on intelligent and autonomous systems. The research featured in this issue captures this momentum, highlighting how advancements in computing and engineering science are enabling new levels of connectivity, efficiency, and technological sophistication.

The editorial team extends its sincere appreciation to the authors for their valuable contributions and to the reviewers for their diligent and constructive feedback. Their collective efforts have ensured the scholarly rigor and relevance of this special issue, reinforcing the journal's commitment to advancing multidisciplinary research.

This special issue offers a comprehensive perspective on the innovations shaping computing, engineering science, and technology in 2023. By bringing together diverse research contributions

and forward-looking insights, it serves as a platform for fostering continued interdisciplinary collaboration and for driving progress in an increasingly complex and technology-driven world.

Guest Editor

Prof. Ahmad Yusairi Bani Hashim

ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

Special Issue **May 2023**

CONTENTS

A Smart Farming Management System based on IoT Technologies for Sustainable Agriculture
by Alioune Cisse, Ousmane Diallo, EL Hadji Malick Ndoye, Mamadou Sy, Ousseynou Sene and Joel José Puga Coelho Rodrigues

Control Program Generator for Vehicle Robot using Grammatical Evolution
by Firdaus Sukarman, Ryoma Sato and Eisuke Kita

Optimal Engagement of Residential Battery Storage to Alleviate Grid Upgrades Caused by EVs and Solar Systems
by Rafi Zahedi, Amirhossein Ahmadian, Chen Zhang, Shashank Narayana Gowda, Kourosh SedghiSigarchi and Rajit Gadh

Transmission of the CAP Protocol through the ISDB-T Standard
by Nelson Bolívar Benavides Cifuentes and Gonzalo Fernando Olmedo Cifuentes

Augmented Reality Based Visual Programming of Robot Training for Educational Demonstration Site
by Lucksawan Yutthanakorn and Siam Charoenseang

Double-Enhanced Convolutional Neural Network for Multi-Stage Classification of Alzheimer's Disease
by Pui Ching Wong, Shahrum Shah Abdullah and Mohd Ibrahim Shapiai

Verify of Left and Right Differences in Sleep Index using the Ring-type Sensor
by Yutaka Yoshida and Emi Yuda

Design and Manufacturing of Soft Grippers for Robotics by Injection Molding Technology
by Helmy Dewanto Bryantono, Melsiani Rosdiani Fillipin Saduk, Jiaqi Hong, Meng-Hsun Tsai and Shi-Chang Tseng

Improved Candidate-Career Matching Using Comparative Semantic Resume Analysis
by Asrar Hussain Alderham and Emad Sami Jaha

Feedback Controller for Longitudinal Stability of Cessna182 Fixed-Wing UAVs
by Veena Phunpeng, Wilailak Wanna, Sorada Khaengkarn and Thongchart Kerdphol

Tree-Based Ensemble Models, Algorithms and Performance Measures for Classification
by John Tsiligaridis

Social Media Text Summarization: A Survey Towards a Transformer-based System Design
by Afrodite Papagiannopoulou and Chrissanthi Angeli

Investigating the Impression Effects of a Teacher-Type Robot Equipped a Perplexion Estimation Method on College Students
by Kohei Okawa, Felix Jimenez, Shuichi Akizuki and Tomohiro Yoshikawa

Revolutionizing Robo-Advisors: Unveiling Global Financial Markets, AI-Driven Innovations, and Technological Landscapes for Enhanced Investment Decisions
by Marco I. Bonelli and Jiahao Liu

Strengthening LoRaWAN Servers: A Comprehensive Update with AES Encryption and Grafana Mapping Solutions
by Sheikh Tareq Ahmed, Annamalai Annamalai and Mohamed Chouikha

Condition Assessment of Medium Voltage Assets: A Review
by Eilin Gómez Mesino, Joaquín Caicedo, Miguel Mamaní, David Romero Quete, Andrés Cerón Piamba, Diego García Gómez, Guillermo Aponte Mayor, José Caicedo Erazo, Wilmar Moreno López, Edward Jay and Andrés Romero Quete

Infrastructure-as-a-Service Ontology for Consumer-Centric Assessment
by Thepparit Banditwattanawong and Masawee Masdisornchote

Analysis of Emotions and Movements of Asian and European Facial Expressions
by Ajla Kulaglic, Zeynep Örpek, Berk Kayı and Samet Ozmen

EEG Feature Extraction based on Fast Fourier Transform and Wavelet Analysis for Classification of Mental Stress Levels using Machine Learning
by Ng Kah Kit, Hafeez Ullah Amin, Kher Hui Ng, Jessica Price and Ahmad Rauf Subhani

The Graded Multidisciplinary Model: Fostering Instructional Design for Activity Development in STEM/STEAM Education
by Mauricio Flores-Nicolás, Magally Martínez-Reyes and Felipe de Jesús Matías-Torres

Comparative Study of J48 Decision Tree and CART Algorithm for Liver Cancer Symptom Analysis Using Data from Carnegie Mellon University
by Renhe Chi

Modeling Control Agents in Social Media Networks Using Reinforcement Learning
by Mohamed Nayef Zareer and Rastko Selmic

Enhancing the Network Anomaly Detection using CNN-Bidirectional LSTM Hybrid Model and Sampling Strategies for Imbalanced Network Traffic Data
by Toya Acharya, Annamalai Annamalai and Mohamed F Chouikha

Resonance Coil Design for a Novel Battery Cell Balancing with using Near-Field Coupling
by Juhyeon Jeon and Dongho Lee
MRI Semantic Segmentation based on Optimize V-net with 2D Attention
by Zobeda Hatif Naji Al-azzwi and Alexey N. Nazarov

Design and Prototyping of a 3DOF Worm-drive Robot Arm
by Ian Spencer Howard

Implementation of a GAS Injection Type Prefabricated Lifting Device for Underwater Rescue Based on Location Tracking
by Jong-Hwa Yoon and Dal-Hwan Yoon

Towards Real-Time Multi-Class Object Detection and Tracking for the FLS Pattern Cutting Task
by Koloud N. Alkhamaiseh, Janos L. Grantner, Ikhlas Abdel-Qader and Saad Shebrain

A Secure Medical History Card Powered by Blockchain Technology
by Samiha Fairooz, Shakila Yeasmin Miti, Zihadul Islam and Meem Tasfia Zaman

Landmarking Technique for Improving YOLOv4 Fish Recognition in Various Background Conditions
by Sutham Satthamsakul, Ari Kuswantori, Witsarut Sriratana, Worapong Tangsrirat and Taweepol Suesut

Bridging Culture and Care: A Mobile App for Diabetes Self-Care Honoring Native American Cultural Practices
by Wordh Ul Hasan, Kimia Tuz Zaman, Shadi Alian, Tianyi Liang, Vikram Pandey, Jun Kong, Cui Tao and Juan Li

Enhancing Cloud Security: A Comprehensive Framework for Real-Time Detection, Analysis and Cyber Threat Intelligence Sharing
by Fazalur Rehman and Safwan Hashmi

Accuracy Improvement-Based Wireless Sensor Estimation Technique with Machine Learning Algorithms for Volume Estimation on the Sealed Box
by Kitipoth Wasayangkool, Kanabadee Srisomboon, Chatree Mahatthanajatuphat and Wilaiporn Lee

Colorized iVAT Images for Labeled Data
by Elizabeth Dixon Hathaway and Richard Joseph Hathaway

Smart Agent-Based Direct Load Control of Air Conditioner Populations in Demand Side Management
by Pegah Yazdkhasti, Julian Luciano Cárdenas–Barrera and Chris Diduch

Dual Mode Control of an Inverted Pendulum: Design, Analysis and Experimental Evaluation
by Laura Álvarez-Hidalgo and Ian S. Howard

A Novel Metric for Evaluating the Stability of XAI Explanations
by Falko Gawantka, Franz Just, Marina Savelyeva, Markus Wappler and Jörg Lässig
Investigating Heart Rate Variability Index Classification in Macaca fascicularis and Humans: Exploring Applications for Personal Identification and Anonymization Studies
by Daisuke Hirahara, Itaru Kaneko, Junji Nishino, Junichiro Hayano, Oscar Martinez Mozos and Emi Yuda

Optimizing the Performance of Network Anomaly Detection Using Bidirectional Long Short-Term Memory (Bi-LSTM) and Over-sampling for Imbalance Network Traffic Data

by Toya Acharya, Annamalai Annamalai and Mohamed F Chouikha

Mathematical Model of Wind Turbine Simulator Based Five-Phase Permanent Magnet Synchronous Generator with Nonlinear Loads and Harmonic Analysis
by Peerawat Meesuk and Vijit Kinnares

Minimum Static VAR Compensation Capacity for Bad Voltage Drop Buses in Power Systems
by Hermagasantos Zein, Ahmad Deni Mulyadi and Achmad Mudawari

Waterfall: Salto Collazo. High-Level Design of Tokenomics
by Sergii Grybniak, Yevhen Leonchyk, Igor Mazurok, Oleksandr Nashyvan and Alisa Vorokhta

Simulation of Obstacle Detection Based on Optical Flow Images for Avoidance Control of Mobile Robots
by Mai Ngoc Anh

Advances in Science, Technology and Engineering Systems Journal Vol. 9, No. 1, 1-8 (2024)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

A Smart Farming Management System based on IoT Technologies for Sustainable Agriculture

Alioune Cisse[1,*], Ousmane Diallo[1], EL Hadji Malick Ndoye[1], Mamadou Sy[1], Ousseynou Sene[1], Joel José Puga Coelho Rodrigues[2]

[1]*Laboratoire d'Informatique et d'Ingénierie pour l'Innovation (LI3), Department of Informatics, University of Assane Seck, B. P. 523 Ziguinchor, Senegal*

[2]*COPELABS, Lusófona University, Lisbon, Portugal*

ARTICLE INFO

Article history:
Received: 20 October, 2023
Revised: 19 December, 2023
Accepted: 19 December, 2023
Online: 20 January, 2024

Keywords:
Internet of Things
Smart farming
Smart agriculture
Prototype deployment
Sustainable agriculture
Food self-sufficiency

ABSTRACT

Advances in Internet of Things (IoT) and wireless technologies are revolutionizing various sectors, including environment, education, healthcare, industry, etc. In the same dynamic, as the world population constantly evolves, solutions based on such technologies need to be proposed to improve the agricultural sector. Senegalese agriculture, primarily rain-fed and based on both cash crops and subsistence food crops, faces challenges from climate change, environmental impacts, and traditional methods. Lack of information on good practices, new efficient and low-cost farming methods, and access to seeds, along with the COVID-19 pandemic, further complicates the situation. This research work proposes a smart farming management system that uses IoT technology, named MbaïMi, whose main objective is to improve Senegalese agriculture through new digital technologies for sustainable agriculture, environmental benefits and food self-sufficiency. It provides a real- time decision support tool that offers farmers best seeds to choose at the beginning of work and helps optimize water resources. Additionally, the system includes features like sending SMS/advice, graphical visualizing of temperature and humidity changes for greenhouse cultivation control. The development of a web and mobile application, deployment of a prototype in southern Senegal, and validation case studies demonstrate the system's potential to improve farmers' yields and daily hard work to meet the challenges of sustainable agriculture, environmental benefits and food self-sufficiency.

1. Introduction

In general, IoT enables process automation by eliminating or reducing human interaction. In the automation process, IoT uses sensors to collect data, controllers to process the data, and actuators to complete the automation process [1]-[4].

To face the challenge of using IoT in various sectors of life in order to improve the living conditions of populations with sustainable development, researchers, practitioners and academics have proposed innovative solutions from the use of IoT technologies in several areas like healthcare, education, automation, industry and agriculture [5]-[7].

The term "smart agriculture" or "smart farming" has emerged as a result of the application of new technological innovations in the agricultural sector, including wireless sensors, Internet of Things (IoT), smart networks, location systems, data science, artificial intelligence, automation, robotics technologies, etc. [8], [9]. Therefore, the main goal of smart agriculture and smart farming is to automate all aspects of farming and agricultural methods to make the process more efficient and effective. This is a global challenge with applications as solutions to solve various problems, including the increased need for improving the quality and quantity of crops, the increased demand for food self-sufficiency, labor force shortages, and the aging of farmers, as well as the expansion of cutting-edge agricultural technology and the optimization of human labor [7]-[10].

Smart farming may be used for remote and automated control of soil quality, precision irrigation, precision plant nutrition, climate change as well as monitoring and proper growth of crops and livestock by taking advantage of efficient management of sensor-generated data. [11]-[13]. These sensor's data are transmitted via wired or wireless communication technology in database servers or cloud platforms to be processed and analyzed by the algorithms in place.

*Corresponding Author: Alioune Cisse, aliounecisse1995@gmail.com

https://dx.doi.org/10.25046/aj090101

Senegalese agriculture is mainly rain-fed and seasonal, based on both cash crops and subsistence food crops. The vast majority of agricultural producers and livestock managers are smallholders that use traditional approaches with many inefficiencies such as higher human interaction and labor cost and face the challenges of climate change, environmental impact and rudimentary means. Added to this the lack of information on good practices, new efficient and low-cost farming methods and technologies, accessing good seeds, the recently economic effects of the COVID-19 pandemic and monetary changes in view (ECO in the UEMOA zone). Efforts must be devoted to the search of scale savings in the name of competitiveness, in particular by increasing agricultural production volumes [14].

This research work proposes the deployment of a smart farming management system that uses IoT Technology for sustainable agriculture, called MbaïMi. The proposed idea employs soil pH, humidity and ambient air temperature, and soil moisture level to provide a real-time categorization of the best crops for a specific soil. Additionally, the system uses this information to automatically manage irrigation as well as manage ambient temperature and humidity for greenhouse farming. The main contributions of this work are:

- The proposal of a smart farming management system architecture that uses IoT technology to remotely and automatically control and increase crop quality and quantity while optimizing human labor.
- The proposal of underlying algorithms of the system.
- The suggestion of a real-world, IoT-based smart farming management system prototype.
- The launch of the smart farming management web and mobile application called MbaïMi, which classifies the best crops for a specific soil in real-time based on factors including soil pH, humidity and ambient air temperature, and soil moisture level. Moreover, the proposal allows to automatically manage irrigation and ambient temperature and humidity for greenhouse cultivation.
- Prototype deployment in the southern region of Senegal and some case studies for evaluation and validation. This shows that the proposal can efficiently help farmers to improve their yields, facilitate their daily hard works to meet challenges of sustainable agriculture, environmental benefits and food self-sufficiency

The rest of this paper is structured as follows. Section 2 presents the related work addressing the most relevant solutions available in the literature. Section 3 describes the proposed system architecture with its underlying algorithms, while the technologies used for the system deployment are presented in Section 4. The

Validation of the proposal via a case study is presented in Section 5. Finally, Section 6 concludes the paper and presents future research works.

2. Related Work

Nowadays, humanity is facing many phenomena such as climate change, floods, and threatening bush fires, which are exposing people, especially in parts of Africa, to famine. Hence, according to the arguments of the coordinator of the PIESAN (Project for Eco-Sustainable Intensification of Agriculture in the Niayes (a locality in Senegal)), the natural degradation of the soil resulting from salinization and acidification, the lack of application of good sustainable agricultural practices (BPAD), unsuitable irrigation and water quality are among the constraints of the agricultural sector. In addition, it adds to these constraints, the lack of control of climatic hazards, the low adoption of proven technical and technological innovations [15].

To overcome this situation, researchers proposed solutions to address food self-sufficiency challenges, aiming to improve agricultural systems through new effective methods.

The work in this paper [10] proposed the design of a smart farm prototype in Thailand, featuring an automatic water control system using IoT devices and sensors. The system consists of a Raspberry Pi board installed in a control box to collect data from the field using a DHT22 sensor and a web-based application. Performance evaluation on a rice farm in Thailand demonstrated its usefulness for Agricultural 4.0 and increased productivity.

In Senegal, E-Tolby has been introduced that is a mobile and web application aimed at improving agricultural output and field watering. It offers real-time decision-making, digital farmer profiles, irrigation water and fertilizer inputs management, and remote practices. Additionally, E-Tolbi offers farming organizations up-to-the-minute information on crop yields on fields and manufacturing facilities. However, the E-Tolbi's algorithm doesn't consider the knowledge of the best crops/seeds for a specific soil [16], [17].

A low-cost framework for food traceability is proposed in [18], using commercial and proprietary sensing devices to monitor air, water, soil parameters, and herbicide contamination during farming. The proposed framework includes an automated prototype for in situ glyphosate detection, monitoring indicators like temperature, humidity, and pH of the irrigation water, total Volatile Organic Compounds (VOCs) and equivalent CO2. The performance evaluation was conducted in an Italian olive farming company, proving its efficacy in food traceability.

The study in [19] proposed a low-cost intelligent system for smart irrigation using IoT technologies. It includes admin mode, onetime irrigation schedule estimation, neural-based decision making, and remote data monitoring. The system uses MQTT and HTTP to inform in real-time farmers about crop conditions and the performances evaluation is performed with a sample crop test-bed for presenting results that include irrigation schedule, neural net decision making and remote data viewing. The authors claim that their proposal proves beneficial with its intelligence, low cost and portability, making it suitable for farms and greenhouse.

AgroCares [20], a Dutch agricultural technology firm, has introduced a soil scanner in Kenya using NIR (near infrared technology) technology. The system provides farmers with immediate information on soil nutritional quality, enabling them to apply appropriate fertilizer. Agrocares offers various application services, including SoilCares, LeafCares, FeedCares, and InsectCares. The little Lab-in-a-Box is a portable detection laboratory for doing on-site analyses of micro and macronutrients. However, the solution requires daily presence on the property to update the samples and does not consider the knowledge of the best crops/seeds according to soil's condition.

FertiTogo, an interactive digital tool introduced during the 12th national Togolese peasant forum, helps farmers determine their agricultural soil fertility and provide precise fertilization recommendations. Using soil spectroscopy, it is a Togo-specific solution, lacking individual sensors and irrigation water control. However, updating the information requires constant implanting and pressing a button, which is not very practical [21].

The proposed smart agriculture model in [22] aims to develop a real-time monitoring system for soil properties, consisting of Farm, Server, and Client levels. It implements decision support advisory models for pest and disease forewarning, crop disease identification, and SMS alerts. The proposed network model implements connected objects with 802.15.4/Zigbee radio links, which therefore limits the range of the solution to small areas.

Based on the literature, the proposals share some similarities with the MbaïMi paper. However, as far as we know, none of the current proposed smart farming systems utilize soil pH, humidity, ambient air temperature, and soil moisture level measurements to provide large-scale farm managers with an effective real-time decision support tool for their agricultural practices, offering them best seed selection from the outset of their field work. Additionally, the proposal includes a prototype deployment, along with a web and mobile application.

3. Architecture of the Proposed Smart Farm Management System

The proposed architecture of the smart farm management system is depicted in Figure. 1. The goal is to facilitate communication between the various components of the system. This communication can be identified in several steps.

The key components of the system consist of an Arduino MEGA 2560 board, a water pump, sensors like soil pH sensor, soil moisture sensor, and temperature and humidity sensors (DHT11), a cloud service, and a web and mobile application.

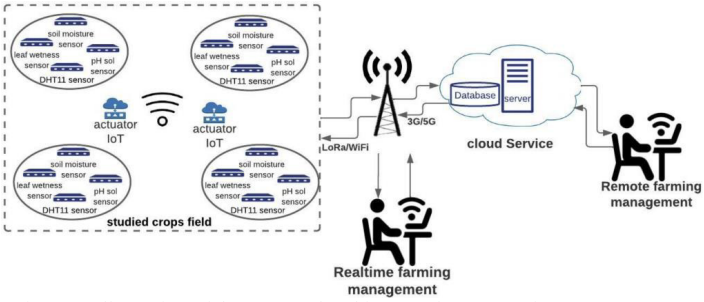


Figure 1: Illustration of the proposed architecture for a smart farm management system

3.1. Description of the Architecture

Studied crops field: The crop farm being studied incorporates IoT devices, including an Arduino MEGA 2560 board (refer to Figure. 2) and a water pump. It also utilizes sensors such as a soil moisture sensor (refer to Figure. 3. A), a soil pH sensor (refer to Figure. 3. B), and DHT11 sensors for temperature and humidity (refer to Figure. 3. C). These sensors gather essential data on soil moisture, acidity, ambient temperature, and air humidity, which are then transmitted through the Gateway for real-time management or to cloud computing services for in-depth analysis.

The network technologies considered for transmitting this data include:

Wi-Fi : between farm equipment and the Gateway;

G/5G : between the Gateway and the Cloud service.

Figure 2: Arduino Mega 2560 board.

Figure 3: (A) The soil moisture sensor,

Figure 3: (B) The pH sensor,

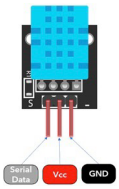


Figure 3: (C) The DHT11 sensor.

Real-time farming management: A web or mobile application allows farmers to access real-time information about the farm under study. The data is presented in different formats, including graphical representations of ambient humidity and temperature, with other data shown in percentages. Additionally, the algorithm proposed in this work provides a classification of the most suitable crops or seeds based on soil pH to increase yield. Furthermore, some tasks such as watering can be managed through the application.

Cloud computing services: It provides administration services for the proposed system's applications, database, and operating algorithm. Then, users may remotely access these data using a computer or a smartphone. The cloud service will enable additional analysis of the data gathered from the farm being studied.

Remote farming management: the farm management services of the proposed system can also be accessed remotely via a web or mobile application housed in the cloud and especially for further analysis of the data captured.

3.2. Proposed Soil pH Check Algorithm

The proposed algorithm is based on the soil's pH level. The pH level captured is recorded in the variable denoted v. Therefore, the value in v undergoes verification in intervals of values corresponding to the acidity or alkalinity (basic) level of the soil (see Figure. 4).

After determining the level of alkalinity or acidity of the soil, the algorithm will check the adaptability of the crops or seeds selected in percentage based on the predefined ranges for each crop or seed.

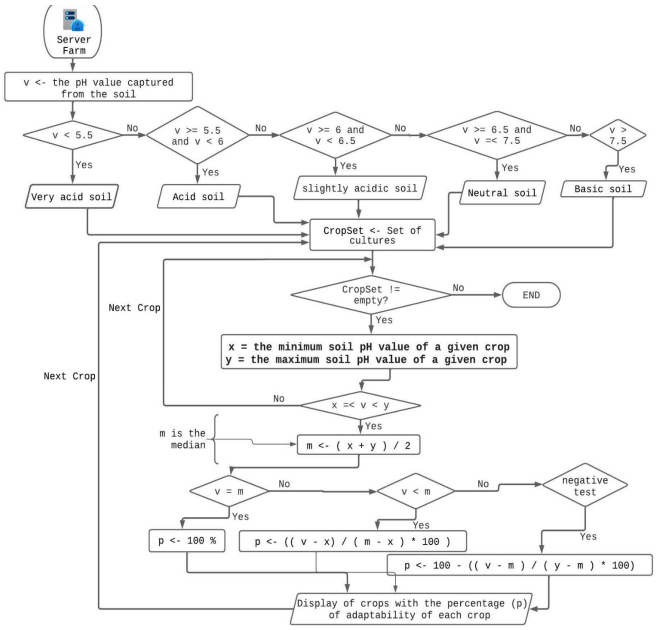


Figure 4: Algorithm for checking soil's pH.

3.3. Proposed Automatic Plant Watering Algorithm

Automatic plant watering is achieved by monitoring soil moisture levels and setting threshold values for each type of crop. The proposed algorithm, shown in Figure. 5), automates the watering process by comparing the soil humidity named *"hSol"* to the defined threshold values for a specific crop *[hSolMinC; hSolMaxC]*. "*hSolMinC"* represents the minimum soil moisture value, while *"hSolMaxC"* represents the favorable maximum soil moisture value for a particular crop.

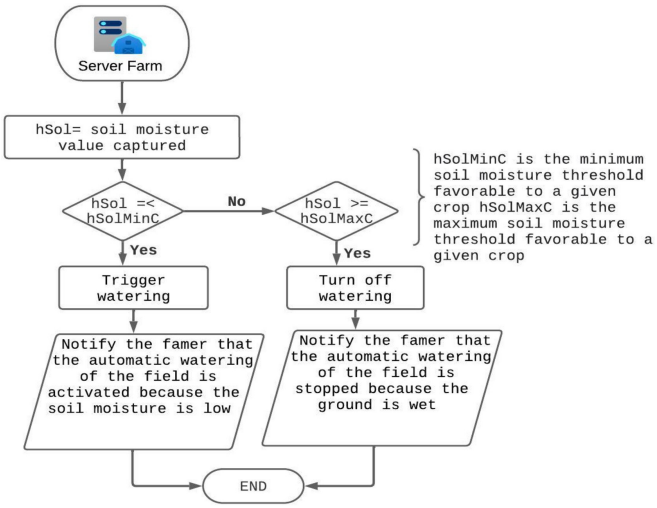


Figure 5: Automatic Plant Watering Algorithm.

3.4. Proposed Ambient Humidity and Temperature Management Algorithm

Humidity and temperature parameters affect the proper growth of plants in a farm. Each crop has its preferred climate range for optimum quality assurance in production. This is why threshold values are provided to study the frequency of optimum conditions and then alarm the farmer. The study of these parameters is especially useful for greenhouse agriculture. The proposed algorithm is shown in Figure. 6.

- Let *T* be the collected ambient temperature and *[Tx; Ty]* represents the defined range conducive to optimal crop growth;
- Let *H* be the collected ambient humidity and *[Hx; Hy]* represents the defined range conducive to good crop growth.

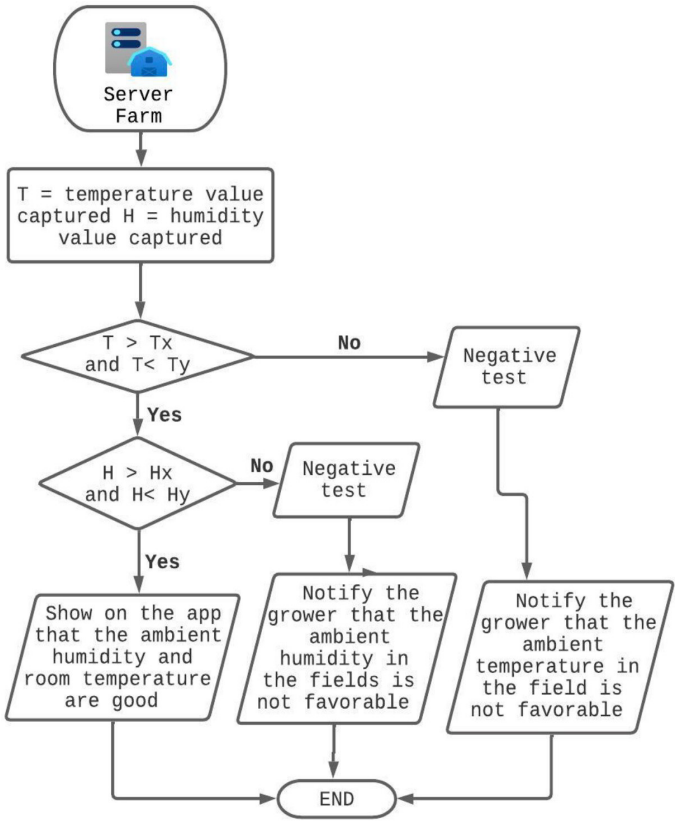


Figure 6. Ambient Humidity and Temperature Management Algorithm.

4. Used Technologies for the System Deployment

The paper's work is based on a monolithic application that depicts a multi-level software application, where various components are integrated into a single program from a unified platform.

Choosing the right technology is a crucial step in developing an application, as its performance, durability, security, and time efficiency all depend on this choice. The Table 1 below provides a summary of the technologies used in this work.

Table 1: Summary of the used technologies.

Technologies	Target/Use	characteristics
Visual studio code frameworks [23]	implementation IDE of the web application	- integrate a debugging console - integrate a terminal for launching server commands - features a robust code auto-completion engine (IntelliSense)
ReactJS and AngularJS frameworks [24], [25]	front-end web application programming language	- Free developed JavaScript library - develops native Android and iOs applications

NestJS [25]	Language used for the back end.	- Integrate TypeScript, object-oriented and reactive functional programming - server-side technology based on Node.js - API NEST access
Arduino Software [26]	coding the microcontroller	- open-source Arduino Software for writing and uploading code to Arduino boards - provides autocompletion, code navigation and live functionality - debugger
MongoDB [27]	Manages smart farming data	- version 6.0.1. - Manges NoSQL databases

5. System Evaluation and Validation

5.1. Electronic Diagram of The Smart Farm System Prototype based on IoT

Figure. 7 displays the circuit diagram of the proposed prototype system, which was generated using Fritzing software [28]. In addition to the sensors mentioned in section 3.1, the system prototype includes:

MEGA 2560 microcontroller

The Arduino Mega 2560 board, powered by an ATMega2560 running at 16 MHz, offers 54 I/O, comprising 14 PWM, 16 analog, and 4 UARTs [29]. It includes ports for extra modules and a bootloader for programming alterations. While it is better suited for intricate projects than the Uno board [30], the Arduino MINI [31] is more suitable for farming use.

A Sprinkler Pump

In the prototyping process, a small water pump is utilized as a sprinkler pump. It is equipped with a small water output pipe and requires a 5V power supply.

A relay module

The Arduino board outputs 5V, necessitating a relay to alternate between the Arduino and a 12V or higher sprinkler pump, allowing for the pump to be powered on or off.

A SIM800L module

This mini-module supports 2G GSM and GPRS data, requires minimal power (3.4V to 4.4V), supports advanced AT instructions, and uses a serial communication interface, compatible with Arduino boards.

An ESP 32 module

The prototype employs a 38-pin ESP32 module, which is a development board featuring the SMD ESP32-WROOM-32 microcontroller, to test system functionalities [32]. This setup enables the control of sensors, modules, and actuators through Wi-Fi 802.11 b/g/n and dual mode Bluetooth version 4.2. The module is equipped with two cores, 38 GPIO pins, 16 PWM channels, and 4MB flash memory. It utilizes a micro USB Type B for power and ESP32 programming, integrating the USB controller to UART CP2102.

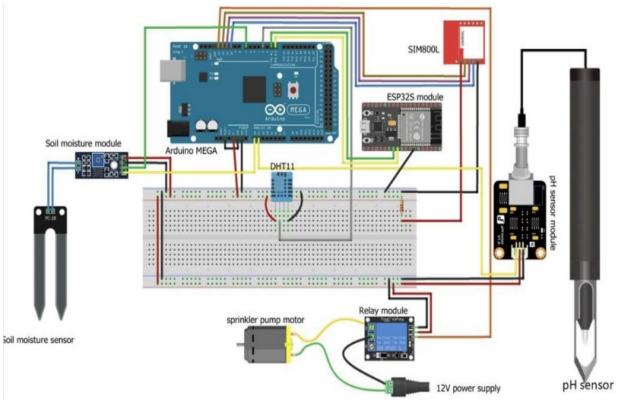


Figure 7: The electronic prototype diagram designed using Fritzing software [28].

5.2. Smart Farm Management System Prototype Deployment

The smart farm management system, utilizing IoT technologies, was created to offer real-time guidance to farmers, empowering them to make informed decisions in their agricultural practices. The system also assists in the management of water resources. Figure. 8 depicts the actual prototype's deployment at a farm in Senegal's southern area, serving as a case study for testing and validation.

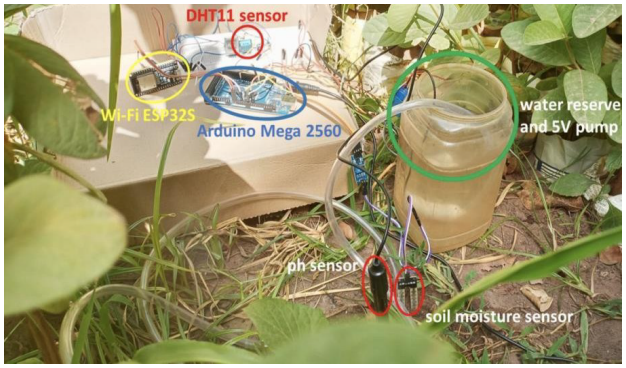


Figure 8. Deployment Capture of the actual prototype for testing and validation.

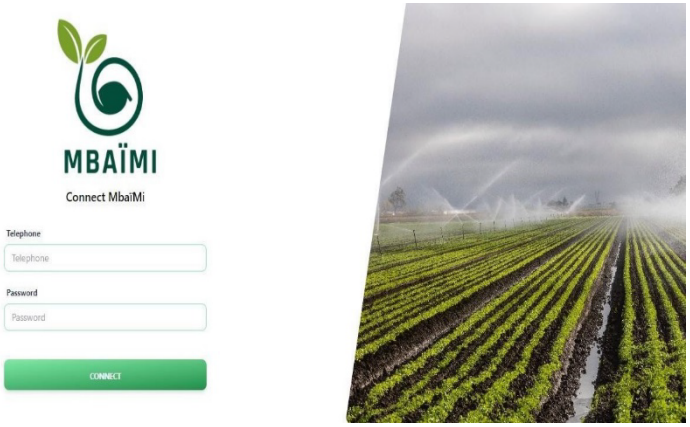


Figure 9: Web application access page.

5.3. The Smart Farm Web App

The *MbaïMi* smart agriculture Web application uses sensor data like readings of the pH of the soil, the humidity and temperature, and the soil moisture level to categorize in real-time crops for specific soils, assisting farmers in selecting seeds that are best suited to the soil, thereby increasing production.

5.4. Web application access interface

The home page is user-friendly, but after opening the program, the administrator or farmer has to authenticate themselves to use its features (see Figure. 9).

5.5. Interface for users setting

This *MbaïMi* system allows administrators and farmers to manage users by assigning roles, entering user information (first name, last name, phone number) and receiving an automatic login password via SMS or email (Figure. 10).

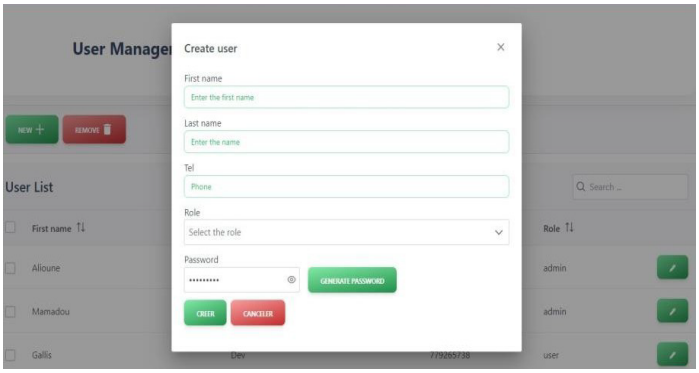


Figure 10: Interface for users setting.

5.6. Interface for setting the studied farm

An administrator or farmer navigates to a menu-rich page to manage different farms (see Figure. 11), managing operations like creation, visualization, update, and deletion, and identifying farms by name, owner, and area.

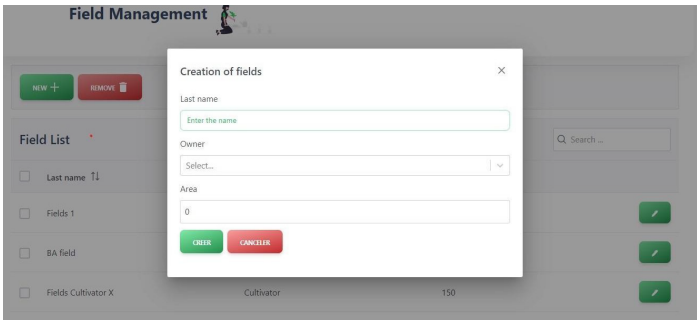


Figure 11: Interface for setting the studied farm.

5.7. Interface for setting the used sensors

An administrator or farmer conFigureures various sensors and actuators, providing information about their names, types, and connected nodes on a page like Figure. 12.

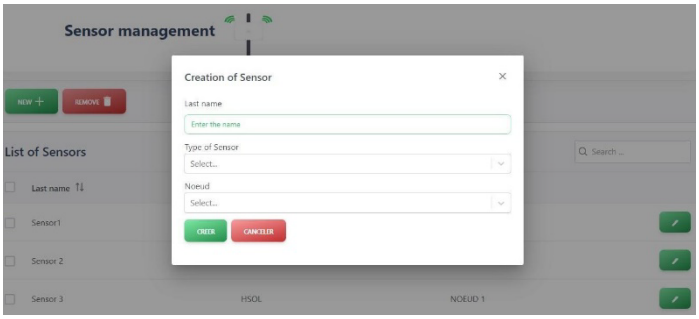


Figure 12: Interface for setting the sensors.

5.8. Interface for crop settings

The interface below (see Figure. 13.) allows users to set thresholds for investigated factors like pH, soil moisture, and leaf humidity, and includes descriptions of each crop to improve seed information in the database.

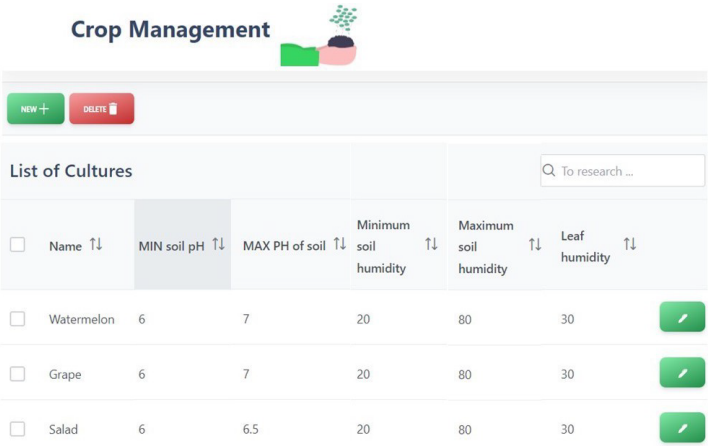


Figure 13: Interface for setting crops.

5.9. Interface showing real-time categorization of best crops /seeds for a studied soil

In the web or mobile application, crops/seeds are displayed along with farm's adaptation percentages using ergonomic graduated color bands (see Figure. 14). For example, watermelon seeds are favorable for the studied farm with a percentage of 95.59%, so the green color band, whereas the passion fruit seeds are unfavorable with a percentage of 8.83%, hence the red band. By choosing the best crops /seeds for a specific soil one may reach challenge sustainable agriculture, reduces environmental degradation and waste seeds and products through more efficient use land, water, fertilizers and pesticides.

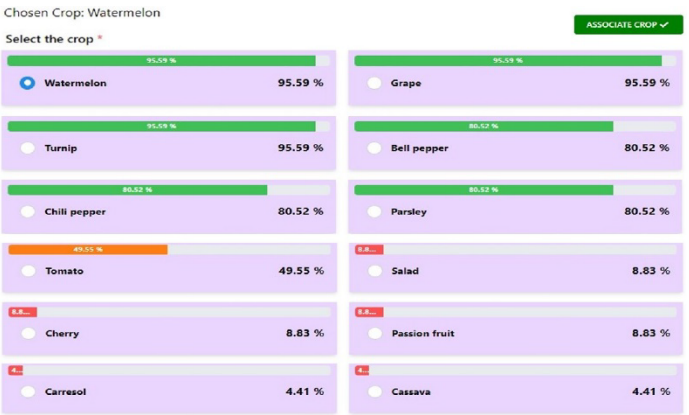


Figure 14: Real-time categorization of the best crops for a given soil.

5.10. Interface for displaying the statistical data

Administrators can monitor sensor data collected on each farm node and visualize temperature and humidity evolution diagrams on the statistics page (see Figure. 15.).

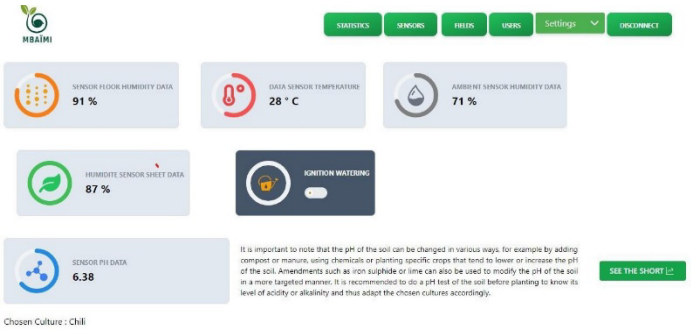


Figure 15:. Display of the statistical data.

5.11. The Smart Farm Mobile Application

As shown in Figure. 16., the user initiates the mobile application and inputs their personal information to access their user space.

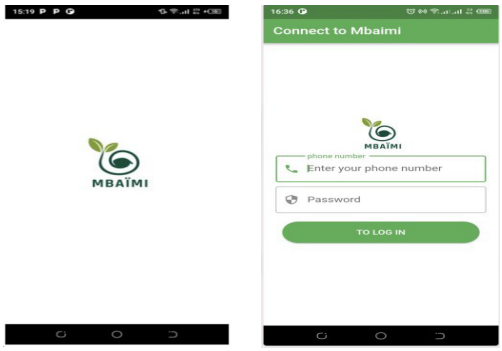


Figure 16: Mobile App Authentication.

The mobile application also assists farmers to consult adapted crops by monitoring sensor data collected every 10 minutes as well as the temperature and humidity diagrams of the farm (see Figure. 17). It also allows famers to manually control watering of plants and send automatic SMS messages, since they may not have a constant connection to the internet.

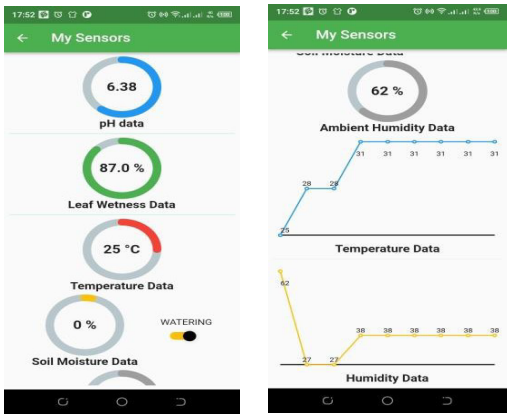


Figure 17: Display of the mobile application sensor data.

6. Conclusions and Future Works

This paper introduces *MbaïMi*, an implementation of a smart farm management web and mobile application with a real hardware prototype that uses IoT technology, which provides farmers an effective decision- support tool for their agricultural practices. The system uses soil pH, humidity, temperature, and moisture levels to determine the soil acidity used to provide a real-time categorization of best crops for specific soils, aiding in seed selection and increasing production. The system also offers autonomous irrigation and temperature control, optimizing water resource utilization and encouraging greenhouse farming.

The *MbaïMi* system is a promising decision support tool for farmers in Senegal. The development of web and mobile

applications and the deployment of the prototype in southern

Senegal show that the system provides real-time categorization of the best crops /seeds, allows automatic irrigation and automatic management of ambient temperature and humidity of the studied soil. This demonstrates the system can improve farmer's yields and daily hard work to meet the challenges of sustainable agriculture, environmental benefits and food self-sufficiency

LoRa technology is ideal for wireless communication in southern Senegal due to its 20 km transmission range and private mode conFigureuration. Therefore, for future works we plan to include the LoRa technology for communication between the IoT devices in farms and the gateways.

7. Acknowledgements

This work is partially funded by Brazilian National Council for Scientific and Technological Development - CNPq, via Grant No. 306607/2023-9, and the LI3 laboratory of the Assane Seck University of Ziguinchor.

References

[1] S.R.J. Ramson, S. Vishnu, M. Shanmugam, "Applications of Internet of Things (IoT) – An Overview," in IEEE 2020 5th International Conference on Devices, Circuits and Systems (ICDCS), 92–95, March 2020, doi:10.1109/ICDCS48716.2020.243556.

[2] A.R. Madushanki, M.N Halgamuge, W. A., A. Syed, "Adoption of the Internet of Things (IoT) in Agriculture and Smart Farming towards Urban Greening: A Review," Int. Journal of Advanced Computer Science and Applications, **10**(4), 2019, doi:10.14569/IJACSA.2019.0100402.

[3] Z. Eghbali, M.Z. Lighvan, "A hierarchical approach for accelerating IoT data management process based on SDN principles," Journal of Network and Computer Applications, **181**, 103027, 2021, doi:10.1016/j.jnca.2021.103027.

[4] A.A. Sadri, A.M. Rahmani, M. Saberikamarposhti, M. Hosseinzadeh, "Fog data management: A vision, challenges, and future directions," Journal of Network and Computer Applications, **174**, 102882, 2021, doi:10.1016/j.jnca.2020.102882.

[5] A. Khanna, S. Kaur, "Evolution of Internet of Things (IoT) and its significant impact in the field of Precision Agriculture," Computers and Electronics in Agriculture, **157(2)**, 218–231, 2019, doi:10.1016/j.compag.2018.12.039.

[6] J. Muangprathub, N. Boonnam, S. Kajornkasirat, N. Lekbangpong, A. Wanichsombat, P. Nillaor, "IoT and agriculture data analysis for smart farm," Computers and Electronics in Agriculture, **156**, 467–474, 2019, doi:10.1016/j.compag.2018.12.011.

[7] M. Raj, S. Gupta, V. Chamola, A. Elhence, T. Garg, M. Atiquzzaman, D. Niyato, "A survey on the role of Internet of Things for adopting and promoting Agriculture 4.0," Journal of Network and Computer Applications, **187**, 103107, 2021, doi:10.1016/j.jnca.2021.103107.

[8] T. Talaviya, D. Shah, N. Patel, H. Yagnik, M. Shah, "Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides," Artificial Intelligence in Agriculture, **4**, 58–73, 2020, doi:10.1016/j.aiia.2020.04.002.

[9] E.H.M. Ndoye, O. Diallo, N. Hakem, E.N. Cabral, "Interference-Aware Nodes Deployment of a LoRa-Based Architecture for Smart Agriculture in the Southern Region of Senegal," Advances in Science, Technology and Engineering Systems Journal, **7**(6), 248–255, 2022, doi:10.25046/aj070628.

[10] P. Suanpang, P. Jamjuntr, "A Smart Farm Prototype with an Internet of Things (IoT) Case Study: Thailand," Journal of Advanced Agricultural Technologies, **6**(4), 241–245, 2019, doi:10.18178/joaat.6.4.241-245.

[11] S. Kim, M. Lee, C. Shin, "IoT-Based Strawberry Disease Prediction System for Smart Farming," Sensors, **18**(11), 4051, 2018, doi:10.3390/s18114051.

[12] V.R. Pathmudi, N. Khatri, S. Kumar, A.S.H. Abdul-Qawy, A.K. Vyas, "A systematic review of IoT technologies and their constituents for smart and sustainable agriculture applications," Scientific African, **19**(11), e01577, 2023, doi:10.1016/j.sciaf.2023.e01577.
[13] A. Mukherjee, S. Misra, N.S. Raghuwanshi, "A survey of unmanned aerial sensing solutions in precision agriculture," Journal of Network and Computer Applications, **148**, 102461, 2019, doi:10.1016/j.jnca.2019.102461.
[14] S.S.A. Emira, K.Y. Youssef, M. Abouelatta, "Simulated IoT Based Sustainable Power System for Smart Agriculture Environments," Advances in Science, Technology and Engineering Systems Journal, **6**(1), 1030–1039, 2021, doi:10.25046/aj0601114.
[15] PIESAN, Newsletter Project information of Eco-Sustainable Intensification of Agriculture in the Niayes, N°01, 2021.
[16] S. Abdelkrim, L. Daniel, "Agritech : A la découverte de « Tolbi », la startup qui modernise l'agriculture sénégalaise," StartupBRICS. StartupBRICS L'actu Tech & Startup Des Emergents, 2022, https://startupbrics.com/.
[17] TOLBI, https://tolbico.com, 2022.
[18] A.V. Radogna, M.E. Latino, M. Menegoli, C.T. Prontera, G. Morgante, D. Mongelli, L. Giampetruzzi, A. Corallo, A. Bondavalli, L. Francioso, "A Monitoring Framework with Integrated Sensing Technologies for Enhanced Food Safety and Traceability," Sensors, **22**(17), 6509, 2022, doi:10.3390/s22176509.
[19] N.K. Nawandar, V.R. Satpute, "IoT based low cost and intelligent module for smart irrigation system," Computers and Electronics in Agriculture, **162**, 979–990, 2019, doi:10.1016/j.compag.2019.05.027.
[20] AgroCares. "Smart Farming | Nutrient Testing – AgroCares. AgroCares nutrient intelligence," 2022, https://www.agrocares.com/.
[21] FertiTogo, "Plate-forme FertiTogo, système d'aide à la décision et à l'information," FertiTogo, 2022.
[22] K.A. Patil, N.R. Kale, "A model for smart agriculture using IoT," in 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), IEEE: 543–545, 2016, doi:10.1109/ICGTSPICC.2016.7955360.
[23] StudioVisual, Documentation for Visual Studio Code, 2022.
[24] Ionic, https://ionicframework.com/docs., Introduction to Ionic, 2022.
[25] NestJS., https://docs.nestjs.com, Documentation, 2022.
[26] Arduino, https://docs.arduino.cc/., Documentation., 2022.
[27] MongoDB, https:/www.mongodb.com/., 2023.
[28] Fritzing Software, https://fritzing.org/, 2023.
[29] Arduino Mega2560:, "https://store.arduino.cc/products/arduinomega-2560-rev3," 2023.
[30] Arduino UNO, https://docs.arduino.cc/hardware/uno-rev3., 2023.
[31] Arduino Pro, https://docs.arduino.cc/retired/boards/arduino-pro-mini., 2023.
[32] ESP32, "http://esp32.net/.," 2023.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 8, No. 6, 01-07 (2023)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Control Program Generator for Vehicle Robot using Grammatical Evolution

Firdaus Sukarman[1,3], Ryoma Sato[2], Eisuke Kita[*,1]

[1]*Graduate School of Informatics, Nagoya University, Nagoya, 464-8601, Japan*

[2]*Graduate School of Information Science, Nagoya University, Nagoya, 464-8601, Japan*

[3]*Faculty of Mechanical Engineering, Universiti Technologi MARA Selangor, 40450 Shah Alam, Malaysia*

ARTICLE INFO

Article history:
Received: 16 May, 2023
Accepted: 10 September, 2023
Online: 30 November, 2023

Keywords:
Evolutionary Computation
Robot Control
Grammatical Evolution
Vehicle Robot
Program Generation

ABSTRACT

A robot development has spread widely for various purposes. It is difficult to create a control program for an autonomous mobile robot manually. Therefore, an automatic design of the control program for an autonomous mobile robot is proposed in this research. The autonomous mobile robot is created with LEGO MINDSTORMS EV3, and the control program for the autonomous mobile robot is designed using Grammatical Evolution (GE). Grammatical Evolution (GE), which is one of the evolutionary computations, is designed to generate a program or a program fragment satisfying the design objective. PyBullet is used with GE to simulate the behavior of the robot. A robot traveling along a trajectory was considered as an example. GE can generate the control program of the robot behavior of a robot vehicle traveling along a trajectory. The computer simulation reveals the robot can travel along a designated line. Since there is a reality gap between the simulator and the real environment, the parameters of the vehicle robot such as produced power and sensor sensitivity are calibrated to reduce the gap. Comparison of the computer simulation and the experimental result shows that the reproducibility of the vehicle trajectory in the real environment is high.

1 Introduction

Evolutionary algorithm (EA), which is one of the heuristic search methods, is widely applied for complex and continuous optimization problems [1]. Specially, it is very effective for solving problem with too many design variables. Evolutionary Computation includes Genetic Algorithm (GA) [2, 3], Genetic Programming(GP) [4, 5], Evolutionary Strategy (ES), Evolutionary Programming (EP) and so on.

Genetic Algorithm is widely applied for the complex optimization problem. Candidate solutions of the problem are defined as the individuals [2, 3]. Each individual has the chromosome which defines binary design variables. Population is constructed by the group of the individuals. By applying the genetic operators such as selection, crossover and mutation, individuals evolve into new individuals so that the objective function is minimized. Genetic Programming(GP) is also well-known evolutionary algorithm [4, 5]. Genetic Algorithm aims at finding the solutions of functions, whereas Genetic Programming aims at designing functions and programs [6, 7]. Individuals are defined in a binary tree structure, which is very different from them in Genetic Algorithm. Since the individuals are defined in a binary tree structure, genetic operators for GP are very different from them for GA.

Grammatical Evolution (GE), which was presented in 1998 by O'Neill and C. Ryan, is the evolutionary algorithm for determining the function or program which satisfies the design objective [8]–[10]. The aim of GE is the same as GP. Its algorithm, however, is slightly different. Individuals of GE are defined in the binary or string, like GA. The translation from the binary or integer numbers to function or program is performed according to Backus-Naur form (BNF), which is defined by a user in advance. Authors' colleagues applied Grammatical Evolution to symbolic regression problem, stock price prediction problem and generation of control program of artificial ant in computer simulation [11]–[14]. In this study, Grammatical Evolution is applied for generating the control program of a real vehicle robot. A vehicle robot is made of LEGO MINDSORM EV3 [15, 16].

PyBullet is utilized for simulation of the actual robot motions [17, 18] . This environment is widely utilized as a robot learning environment for manipulation due to its portability and light weight

[*]Corresponding Author: Eisuke Kita, Email: kita@i.nagoya-u.ac.jp

for variety of machine learning tasks [19]. After the program is created in the simulation environment, it is applied to control the actual robot. Since there are differences in sensor sensitivity and motor output between the actual robot and the robot in the simulation environment, the program is modified to compensate for the differences. The results of the robots in the actual and the simulation environments are compared for discussing the validity of the generated program.

The remaining part of this paper is organized as follows. The evolutionary computations are introduced in Section 2. In Section 3, experiments and discussion of control program design using GE are performed. In Section 4, the designed control program is applied to the actual vehicle robot. The conclusion and future issues are summarized in Section 5.

2 Grammatical Evolution

2.1 Outline

Grammatical Evolution is designed to find a function or a program satisfying the design objective. GE composes the initial population by the individuals with randomly generated bit-strings or the sequence of integer numbers [20]. The translation from the integer-string to the program is performed according to the translation rules of Backus-Naur Form (BNF) grammar [21]. After the integer-string of each individual is translated into a function or a program, its fitness is evaluated. The individuals are updated by using genetic operators such as selection, crossover, and mutation with the individual fitness. These process are repeated till the design objective is satisfied. The process is summarized as follows.

1. Define translation rules based on targeted task, which translates integer string to a function or a program.
2. Generate individuals from integer string randomly to define an initial population.
3. Translate the integer-string of each individual into a program.
4. Select parent individuals from population according to fitness.
5. Apply genetic operators such as selection, crossover, and mutation to parent individuals to generate offspring individuals.
6. Update population.
7. If the individual satisfying the design objective can be found, the results are output. If not so, go back to Step 3.

2.2 Translation from Integer String to Program

The translation rules are defined with the set of non-terminal symbols N , the set of terminal symbols T , and the start symbol S. Non-terminal symbols are replaced with the other non-terminal or terminal symbols according to the translation rules. Terminal symbols, on the other hand, is no longer replaced. The sets of the symbols are summarize as follows.

N = { <code>,<op>,<var> }
T = { +,-<*,/,X }
S = { <code> }

The translation rule is shown in Table 1. The symbol "|" means "or". The symbol <code> (A) is replaced with one of three candidates; <code><code> (A0), <op> (A1), and <var> (A2). The symbol <op> and <var> have four candidates and one candidate, respectively.

Table 1: Example of Simple Translation Rules

(A)	<code>	::=	<code><code>	(A0)
		|	<op>	(A1)
		|	<var>	(A2)
(B)	<op>	::=	+	(B0)
		|	-	(B1)
		|	*	(B2)
		|	/	(B3)
(C)	<var>	::=	x	(C0)

Table 2: Evolution of Symbols

S_n	r_A	m	A	Selected symbols	Replaced symbol
			<code>		
3	3	0	<code>	<code><code>	<code><code>
0	3	0	<code>	<code><code>	<code><code><code>
2	3	2	<code>	<var>	<var><code><code>
			<var>	x	x<code><code>
1	3	1	<code>	<op>	x<op><code>
2	4	2	<op>	*	x*<code>
2	3	2	<code>	<var>	x*<var>
			<var>	x	x*x

The translation process is summarized as follows. For example, assuming that the individual is defined as "302122", the translation of "302122" according to Table 1 is shown in Table 2. The start symbol is α = <code>. The leftmost not-used gene is n_0 = 3. The symbol α = <code> has three potential symbols, $n_0 = 3$. Since the remainder of $n_0 = 3$ divided by $n_\alpha = 3$ is $n_r = 0$, the symbol α = <code> is replaced with 0-th symbol of the candidates <code><code>.

Next, the leftmost symbol α = <code> of the symbols <code><code> is replaced as follows. The second leftmost not used number of the individual is $n_0 = 0$. The symbol α = <code> has three candidates and thus, $n_0 = 3$. Since the remainder of $n_0 = 0$ divided by $n_\alpha = 3$ is $n_r = 0$, the symbol α = <code> is replaced with the symbol <code><code> and then, the symbols <code><code> becomes the symbols <code><code><code>.

Next, the leftmost symbol α = <code> of the symbols <code><code><code> is replaced as follows. The third leftmost not-used number of the individual is $n_0 = 2$. The symbol α = <code> has three candidates and thus, $n_\alpha = 3$. Since the remainder of $n_0 = 2$ divided by $n_\alpha = 3$ is $n_r = 2$, the symbol α = <code> is replaced with the symbol <var> and then, the symbols <code><code><code> becomes the symbols <var><code><code>. According to the similar process, the individual "302122" results in $x * x$. The whole process is shown in Table 2.

3 Design of Control Program

3.1 Robot Vehicle

The robot vehicle is created by LEGO MINDSTORMS EV3 (Fig.1). LEGO MINDSTORMS is an educational robot kit jointly developed by LEGO and Massachusetts Institute of Technology. LEGO MINDSTORMS EV3, which was released in 2013, contains an intelligent block with a 32bit ARM9 microprocessor which allows the robot to operate autonomously by downloading and executing

the program. Several sensor such as an ultrasonic sensor, a color sensor, a gyro sensor, and a touch sensor can be attached to the input port and then, the values of each sensor can be used to control the servo motor connected to the output port.

For rapid software development, the robot is loaded with MicroPython firmware for complete integration with a Python-based simulation environment. For uploading and debugging programs on robots, a standard IDE such as Visual Studio Code is utilized. Actual course used in this experiment is shown in Fig.2.

Figure 1: Lego Mindstorms EV3

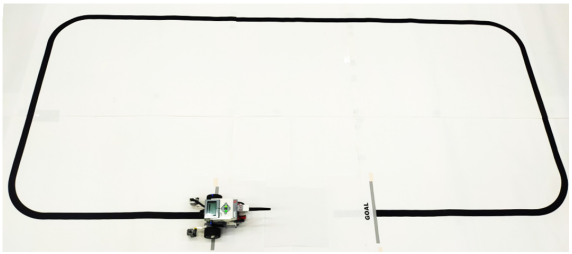


Figure 2: Actual Course

robot, which determines where and in what direction to place the robot first. The GOAL tile is the end of the robot motion. When the center of the robot's axle enters the GOAL tile, the simulation is terminated. The CHECKPOINT tile is functioned as Checkpoint for the robot motion. In Fig. 4, the tiles numbered with $i = 1, \cdots, 18$ denote the CHECKPOINT tiles. A robot has to pass all CHECKPOINT tiles when a robot moves along a line.

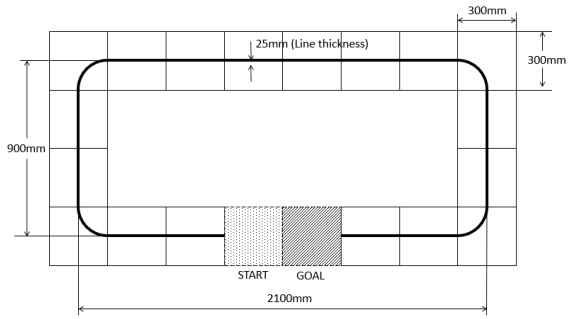


Figure 3: Dimension of Course

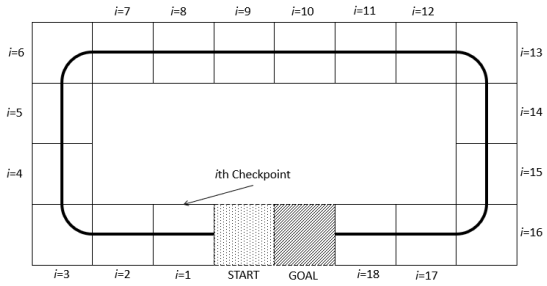


Figure 4: Course Checkpoint

3.2 Simulation Environment

PyBullet is utilized for simulation due to its capability to replicate actual robot motions using Bullet Physics Engine and stability in repetitive machine learning calculations [17] . This environment is widely utilized as a robot learning environment for manipulation due to its portability and light weight for variety of machine learning tasks.

Robot and track model is generated using standard Unified Robot Description Format (URDF) used in Robotics Operating System (ROS) [22]. Object physic characteristics can be defined deliberately in the XML-styled format thus comparatively fasten simulation environment with accurate parameters. Center of axle is configured as measurement point for positional analysis.

Detailed dimensions of the course is shown in Figure 3. The course design is performed by placing 30cm x 30cm tiles on a 240cm x 120cm field. Lines center is located at the center of the tiles with thickness 25mm to match attachment height of the reflection sensor. This setup will enable the reflection sensor to completely sense whether the surface reflect light or not. Tiles with various shapes are prepared, and it is possible to design a course according to the purpose. START, GOAL, and CHECKPOINT can be set for tiles as their specifications. The START tile is the starting position of the

3.2.1 Control Program Translation rules

Grammatical Evolution can generate the best control program from the randomly defined candidate programs according to translation rules known as Backus-Naur Form (BNF) Grammar.

Table 3 shows the translation rule defined for this experiment. In defining the translation rule, the control statements of LEGO MINDSTORMS EV3 are introduced. This statements consists control statement used in PyBullet and LEGO MINDSTORM EV3, and can be controlled with identical control statement both in simulator and real machine.

The translation rules are defined with the set of non-terminal symbols N, the set of terminal symbols T, and the start symbol S. The sets of the symbols are summarize as follows.

N = { <code>,<op>,<num> }
T = { if(SensorValue > (100+)/2): else:,
motor(, ,),-100,-90,-80,-70,-60,
-70,-60,-50,-40,-30,-20,-10,0,10,20,
30,40,50,60,70,80,90,100 }
S = { <code> }

The symbol <code> (A) can be replaced with one of three candidates; <code><code> (A0), <op> (A1) and if(SensorValue > (100+<num>)/2): <op> else:

<op>(A2). The symbol <op> and <num> have one candidate and 21 candidates, respectively.

Table 3: Translation rules

(A)	<code>::=	<code><code>	(A0)
	|	<op>	(A1)
	|	if SensorValue>(100+<num>)/2{<op>} else{<op>}	(A2)
(B)	<op>::=	motor(<num>,<num>,100+<num>)	(B0)
(C)	<num>::=	-100	(C0)
	|	-90	(C1)
	|	-80	(C2)
	|	-70	(C3)
	|	-60	(C4)
	|	-50	(C5)
	|	-40	(C6)
	|	-30	(C7)
	|	-20	(C8)
	|	-10	(C9)
	|	0	(C10)
	|	10	(C11)
	|	20	(C12)
	|	30	(C13)
	|	40	(C14)
	|	50	(C15)
	|	60	(C16)
	|	70	(C17)
	|	80	(C18)
	|	90	(C19)
	|	100	(C20)

At translation rule (A), the statement <code><code> in Rule (A0) adds the evolvable mechanism. Direct selection of motor control is added using the statement <op> in Rule (A1) that enable the selection of motor movement directly. For sensor input, if SensorValue>(100+<num>)/2{<op>}else{<op>} is used as *if*-statement to control the motor movements upon receiving the sensor value. The statement SensorValue get the output from color sensor which is integer between 0 to 100 and the conditional value (100+<num>)/2 is used to match the color sensor input in increments of 5. This conditional statement produces two executable statements <op>, first statement is executed when the sensor value is larger than selected value, otherwise the second statement is executed. This allows the color output of the sensor change the pattern of the motor input.

At translation rules in (B), the statement motor(<num>,<num>, 100+<num>) is the control statement for the motors, where the arguments are for left motor, right motor and waiting time for the movement to complete. The statement <num> for first and second argument is used to select the speed input range from -100 to 100 in 10 increments. Negative value means the motor will run backwards with percentage ratio. The statement 100+<num> is a control argument that sets the execution time (ms) of motor power input. Since the argument started with value of 100, only positive value will be produced from 0 to 200 in increments of 10.

At translation rule C, a common usable parameter with a value between -100 and 100 is set so that it can be used for sensor and motor input conditional values.

3.2.2 *Fitness Function and Parameters*

Each generated program is evaluated using a fitness function capable of achieving the goal by passing all checkpoints in less than 60 seconds. The fitness function is defined such that lower values indicate a better solution, so that the control program generation problem is appointed as a minimization problem. The objective function is defined as below.

$$\min f(x) = \begin{cases} T_{Goal} & (if N_{checkpoints} = 18, T_{Goal} \leq 60) \\ 10^4 & (otherwise) \end{cases} \quad (1)$$

where $N_{checkpoints}$ and T_{Goal} represent number of checkpoints tiles,i-th and total time taken,T (s) to reach the GOAL tile, respectively. Only individuals that pass through within all 18 checkpoints are evaluated and beyond that the fitness is penalized with very large number(10^4). Large number is used as fitness scaling to control diversity to prevent the populations converge to early toward a single optimal solution[3]. Solutions fitness that go beyond the tiles will also be deducted until the trajectories of robot follow the line and finally reach the determined target. Parameters used is shown in Table 4

Table 4: GE Parameters

Population size	200
Maximum generation	50
Gene length	1000
Number of elites	1
Selection	Tournament
Tournament size	5
Crossover type	One-point crossover
Crossover rate	0.4, 0.5, 0.7
Mutation rate	0.03, 0.05, 0.07

3.3 *Simulation Results*

In Figure 5 , we perform five simulations with different initial values and show the average value. This attempts to remove the influence of how the initial values were selected. The horizontal and the vertical axes represent the number of generations and the arrival time to the GOAL tile, respectively. The result of the crossover rate 0.5 and the mutation rate 0.05 demonstrates the quickest convergence during the initial generation. Table 5 shows the average fitness of the best individual in the final generation for different crossover rates and mutation rates. When the crossover rate is 0.4, the simulation did not converge for mutation rate 0.03 and 0.05, but shows lowest rate of convergence when mutaton rate is 0.07. The best fitness is obtained when the crossover rate is 0.5 and the mutation rate is 0.05.

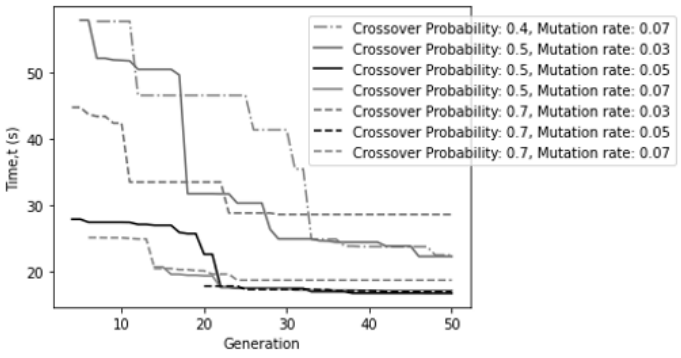


Figure 5: Convergence history travel time over generation

Table 5: Effect of Crossover and Mutation rate

Crossover rate \ Mutation rate	0.03	0.05	0.07
0.4	-	-	22.51
0.5	22.32	16.74	17.20
0.7	28.64	16.98	18.74

```
if(SensorValue > 80):
    motor(80,80,50)
else:
    motor(80,0,160)
```

Figure 6: Generated program

Fig.6 shows the program generated based on the best individual of the final generation. It is notice that the best generated program is a simple line tracing program. When the color sensor of a robot vehicle notices white on the course, the vehicle moves forward with the maximum motor output. When the color sensor notices black, the vehicle rotates to the right on the spot. At the start, the vehicle moves forward because the color sensor is positioned on the white field. When the color sensor enters the black line, the vehicle rotates to the right. After completing the rotation, the vehicle moves forward because the color sensor is positioned on the white field. After repeating forward and right turns alternately four times in this way, a robot vehicle goes straight and reach the GOAL tile.

4 Experiment

4.1 Performance Estimation of Robot Vehicle

LEGO MINDSTORMS EV3 robot can be controlled using motor power input (Duty Cycle) in percentage and speed input. When the motor is assigned with increments of speed in 2s interval time, rotational speed increases and reached the limit of 800degree/s as shown in the Fig. 7 which is equivalent to the maximum of 80% of the power input. Motor power input is capped to 80% and the limitations are also applied to the simulation.

Motor power input is a simple method to control the movement of the robot by supplying current to the motor in ratio. The estimation of motor speed require actual power input when the robot is under actual load when moving around the track with surface friction. Fig.8 shows the relation between power and rotational speed of motor in degree/s.

4.2 Kinematics of Robot

Grammatical Evolution program uses speed input to vary the velocities of the two wheels, thus determine the trajectories of the robot. Actual robot used in this experiment utilize Differential-drive Robot concept where two independent motor with wheel radius,r rotate about the same axis and low-friction caster wheel is used to keep the robot horizontal.

For left and right motor linear velocity, v_l and v_r, where R is distance from center of curvature to the center of the axle and l is effective distance between wheel (axle), kinematics equation based on Instantaneous center of curvature concept for the system is shown below;

$$v_l = \omega(R + l/2)$$
$$v_r = \omega(R + l/2)$$
$$\omega = \frac{v_l + v_r}{2} \quad (2)$$

By varying the motor speed for left and right, trajectories of the robot can be determined based on the turning rate, ω.

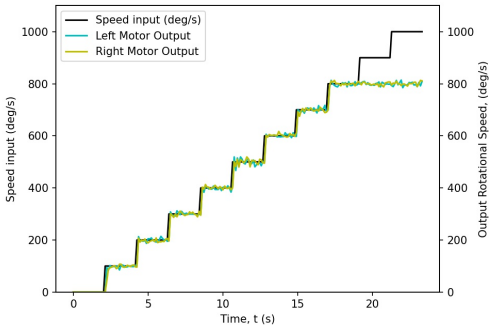


Figure 7: Relation between Speed input and Rotation Speed (degree/s)

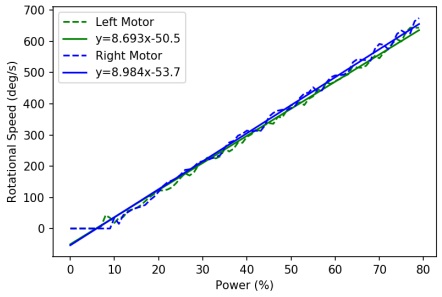


Figure 8: Relation between Power and Rotation Speed (degree/s)

4.3 Actual Robot Experiments

Evaluation of the simulation is done by using the actual robot with identical dimensions to prove the effectiveness of the program generated. Program generated from GE is executed to check the effectiveness.

Figure 9 shows the trajectory of the vehicle. A robot vehicle starts from the position of X=0 and Y=0 and then, makes one lap around the course clockwise. A solid line shows the computer simulation result and broken lines denote actual experiment results of five trials of vehicle robot, respectively. Table 6 shows the arrival time for 5 trials performed. Maximum, minimum, and mean arrival times are 18.71 (s), 19.92 (s) and 19.30 (s), respectively. Since the arrival time of the simulation is 16.74 (s), the errors of maximum, minimum, and mean arrival times are 11.8 (%), 19.0 (%) and 15.4 (%), respectively. In experiments using actual robots, it is expected that the arrival time becomes longer than that in simulations due to factors such as robot acceleration and control delay time.

This difference in time elapsed is caused by the actual robot movement take a longer path to make turn for each line corner. Longer sensor responses and efficiency of the motor are two main reasons which create a tiny small lag for the robot to make turns. Even there is a small time difference, the robot control program able to accomplish the main objective which is following the line. In addition, Figure 10 shows the difference between the vehicle position in the simulator and in the real environment at each time step (0.01s). The error fluctuates up and down with time, but basically increases with time.

The correlation coefficient between the vehicle coordinates in the simulator and the experiment is evaluated for the trajectory of the vehicle, and the result is show in Table 7. A high correlation is shown in both the x-coordinates and y-coordinates. It can be seen that the correlation coefficient for the y-coordinate is lower than that for the x-coordinate. From Fig. 9, it is expected that the deviation of the y-coordinate between the simulation and the experiment is relatively large.

Table 6: Running Time

Experiment	Arrival time (s)	Error (%)
Trial 1	18.98	13.4
Trial 2	19.51	16.5
Trial 3	18.71	11.8
Trial 4	19.92	19.0
Trial 5	19.47	16.3
Average	19.32	15.4

Arrival time at simulation : 16.74 (s)

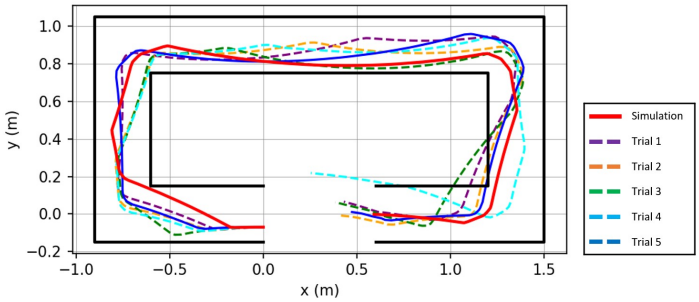


Figure 9: Robot path comparison between simulation and actual

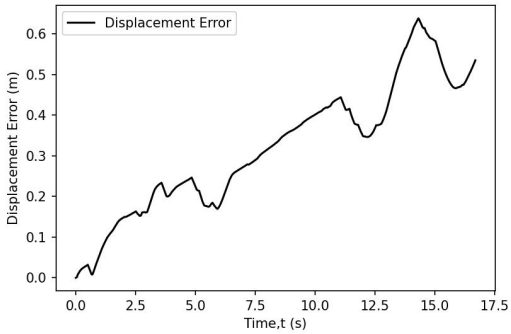


Figure 10: Displacement between vehicle position in simulator and real world

Table 7: Correlation coefficient between vehicle coordinates in simulation and experiment

x-coordinate	0.9553
y-coordinate	0.7743

5 Conclusion

A method of designing a control program for an autonomous mobile robot using Grammatical Evolution (GE) was proposed in this research. The autonomous mobile robot was created with LEGO MINDSTORMS EV3, and the control program for the autonomous mobile robot was designed using Grammatical Evolution (GE). Robot Virtual Worlds was used to simulate the behavior of the robot. A robot traveling along a trajectory was considered as an example. As a result, the fitness converged toward the optimal solution, and the running trajectory of the robot according to the designed control program was appropriate. A real machine experiment was performed by applying the control parameters that adjust the gap between the simulator and the real environment. Comparing the computer simulations and experimental results shows that the target achievement time in the running experiment was equivalent to that of the simulator. In addition, it was shown that the reproducibility of the vehicle trajectory in the real environment was high.
Finally, future issues is summarized as follows. First, to reduce the computational time, an algorithm should be revised to improve the convergence performance. Secondly, the experiments with complicated courses and conditions should be performed to verify the effectiveness of the proposed method for many problems.

References

[1] H.-P. P. Schwefel, Evolution and Optimum Seeking: The Sixth Generation, John Wiley & Sons, Inc., USA, 1993.

[2] J. H. Holland, "Adaptation in natural and artificial systems," University of Michigan press, Ann Arbor, MI, **1**, 1975.

[3] D. E. Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning," NN Schraudolph and J., **1**, 1989.

[4] J. R. Koza, Genetic Programming II: Automatic Discovery of Reusable Programs, volume 1, MIT Press, 1994.

[5] J. R. Koza, Genetic Programming lll: Darwinian Invention and Problem Solving, Morgan Kaufmann, 1999.

[6] V. de Carvalho Santos, C. F. M. Toledo, F. S. Osorio, "An exploratory path planning method based on genetic algorithm for autonomous mobile robots," 62–69, IEEE, 2015, doi:10.1109/CEC.2015.7256875.

[7] R. Kala, "Multi-robot path planning using co-evolutionary genetic programming," Expert Systems with Applications, **39**, 3817–3831, 2012, doi:10.1016/j.eswa.2011.09.090.

[8] C. Ryan, J. Collins, M. O. Neill, Grammatical evolution: Evolving programs for an arbitrary language, volume 1391, 83–96, 1998, doi:10.1007/BFb0055930.

[9] M. ONeil, C. Ryan, "Grammatical evolution: Evolutionary automatic programming in an arbitrary language," Norwell, MA, **10**, 1–978, 2003.

[10] C. Ryan, M. ONeill, J. J. Collins, Introduction to 20 years of grammatical evolution, 2018, doi:10.1007/978-3-319-78717-6_1.

[11] T. Kuroda, H. Iwasawa, E. Kita, "Application of advanced Grammatical Evolution to function prediction problem," Advances in Engineering Software, **41**, 1287–1294, 2010, doi:10.1016/j.advengsoft.2010.09.005.

[12] E. Kita, Y. Zuo, H. Sugiura, T. Mizuno, "Acceleration of Grammatical Evolution with Multiple Chromosome by Using Stochastic Schemata Exploiter," Proceedings - 2017 4th International Conference on Mathematics and Computers in Sciences and in Industry, MCSI 2017, **2018-Jan**, 190–195, 2018, doi:10.1109/MCSI.2017.40.

[13] H. Sugiura, M. Nagao, Y. Zuo, E. Kita, "Grammatical evolution using two-dimensional gene for symbolic regression: An advanced improvement with conditional statement grammar," International Journal of Critical Infrastructures, **13**, 2017, doi:10.1504/IJCIS.2017.083634.

[14] E. Kita, H. Sugiura, Y. Zuo, T. Mizuno, "Application of grammatical evolution to stock price prediction," Computer Assisted Methods in Engineering and Science, **24**, 2017.

[15] L. Group, "LEGO MINDSTORMS EV3," 2023.

[16] L. G. 2019-2020, "LEGO MINDSTORMS Education EV3 MicroPython," 2020.

[17] E. Coumans, Y. Bai, "PyBullet, a Python module for physics simulation for games, robotics and machine learning," 2017.

[18] F. Sukarman, E. Kita, "Auto-generated Control Program in Mobile Robot using Grammatical Evolution," 1–5, Association for Computing Machinery, 2022, doi:10.1145/3573910.3573921.

[19] X. Yang, Z. Ji, J. Wu, Y.-K. Lai, "An Open-Source Multi-goal Reinforcement Learning Environment for Robotic Manipulation with Pybullet," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **13054 LNAI**, 14–24, 2021, doi:10.1007/978-3-030-89177-0_2.

[20] M. Fenton, J. McDermott, D. Fagan, S. Forstenlechner, E. Hemberg, M. O'Neill, "PonyGE2: Grammatical evolution in python," GECCO 2017 - Proceedings of the Genetic and Evolutionary Computation Conference Companion, 1194–1201, 2017, doi:10.1145/3067695.3082469.

[21] G. Dick, P. A. Whigham, "Initialisation and grammar design in grammar-guided evolutionary computation," 534–537, ACM, 2022, doi:10.1145/3520304.3529051.

[22] O. R. 2021, "ROS - Robot Operating System," 2021.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 9, No. 2, 01-08 (2024)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Optimal Engagement of Residential Battery Storage to Alleviate Grid Upgrades Caused by EVs and Solar Systems

Rafi Zahedi[*,1], Amirhossein Ahmadian[1], Chen Zhang[1], Shashank Narayana Gowda[1], Kourosh SedghiSigarchi[2], Rajit Gadh[1]

[1] *Smart Grid Energy Research Center, University of California, Los Angeles, Los Angeles, 90095, USA*

[2] *ECE Department, California State University, Northridge, Northridge, 91330, USA*

ARTICLE INFO

Article history:
Received: 16 October, 2023
Revised:12 January, 2024
Accepted: 13 January, 2024
Online: 22 March, 2024

Keywords:
Electric Vehicle
Solar Energy
Distributed Energy Resources
Optimization

ABSTRACT

The integration of distributed energy resources has ushered in a host of complex challenges, significantly impacting power quality in distribution networks. This work studies these challenges, exploring issues such as voltage fluctuations and escalating power losses caused by the integration of solar systems and electric vehicle (EV) chargers. We present a robust methodology focused on mitigating voltage deviations and power losses, emphasizing the allocation of a Permitted Percentage (PP) of battery-based solar systems within residential areas endowed with storage capabilities.

A key facet of this research lies in its adaptability to the changing landscape of electric transportation. With the rapid increase of electric trucks on the horizon, our proposed model gains relevance. By tactically deploying PP to oversee the charging and discharging of batteries within residential solar systems, utilities are poised not only to assist with grid resilience but also to cater to the upcoming demands spurred by the advent of new EVs, notably trucks.

To validate the efficacy of our proposed model, rigorous simulations were conducted using the IEEE 33-bus distribution network as a designed testbed. Leveraging advanced Particle Swarm Optimization techniques, we have deciphered the optimal charging and discharging commands issued by utilities to energy storage systems. The outcomes of these simulations help us understand the transformative potential of various PP allocations, shedding light on the balance between non-battery-based and battery-based solar residences. This research underscores the need for carefully crafted approaches in navigating the complexities of modern grid dynamics amid the anticipated increase in electric vehicles.

1. Introduction

High penetration of distributed energy resources (DER), such as solar systems could result in power quality issues on the distribution network [1,2]. This will be impacted more severely as the number of Electric Vehicles (EVs) grows [3].

Moreover, the emergence of electric trucks introduces distinct challenges and prospects for the distribution network. As highlighted in [4,5], if utilities neglect the increasing prevalence of electric trucks, the distribution network experiences notable voltage fluctuations. They suggested that with an increase in charging loads, there's a heightened probability for upgrades needed for the distribution lines offering an expensive and time-consuming solution with extensive planning to implement [4]. However, strategies for utilities to bear the costs remains an open-ended important question. Furthermore, a high penetration of Non-Battery-Based Solar Residence (NBBSR) elevates the likelihood of voltage fluctuations on distribution lines, particularly when a significant amount of solar power is injected into the feeders

In accordance with FERC Order 2222, regulatory bodies in the United States have been actively promoting the integration and adoption of various Distributed Energy Resources (DERs). This regulatory framework also facilitates the development of models for aggregated resource participation, enabling Behind-the-Meter (BTM) assets such as BTM Battery Energy Storage Systems (BESS) to provide a comprehensive range of power system services for which they possess the technical qualifications [6].

[*]Corresponding Author: Rafi Zahedi, rafi73@g.ucla.edu

https://dx.doi.org/10.25046/aj090201

Moreover, in the context of bring-your-own-device business models, utility customers who are also energy producers, such as residential households, are increasingly investing in Battery-Based Solar Residence (BBSR) systems. These prosumers often utilize only a fraction of their storage capacity. In such scenarios, prosumers may consider sharing their unused storage capacity with third parties [7,8]. Additionally, through Energy-Storage-as-a-Service Arrangements, developers or utilities assume the initial costs associated with BBSR systems and subsequently own and manage these systems in return for a fee [9,10]. Both of these aforementioned business models enable utilities to expand the adoption of BBSR systems and exercise varying degrees of ownership and management as needed. For instance, New York ISO has designed a model that will let DERs such as BBSRs provide services to consumers, utilities, and the wholesale energy market to mainly ensure bulk power reliability and accessibility to all grid parties. However, they have not prioritized the power quality problems relating to the high penetration of NBBSR and EVs [11]. To mitigate the power quality challenges caused by DERs including EVs and solar systems, [12,13] provide more in-depth discussion. They suggested that utilizing storage systems could be potentially more cost-effective and merit more investigation for power quality and voltage deviation issues.

To underscore the importance of this study, a unique optimization approach is introduced and assessed within the context of the IEEE 33-bus system.

This approach aims to enhance the power quality of the grid, specifically focusing on mitigating active power losses and voltage deviations, all without the need for additional infrastructure upgrades such as storage systems, and the incorporation of new energy resources. To reduce the number of deciding factors without sacrificing its general applicability, the IEEE 33-bus system is divided into seven distinct sectors. Within each sector, there are multiple instances of two types of solar systems: BBSR and NBBSR. Notably, only NBBSR is authorized to supply solar energy to the grid, but a designated Permitted Percentage (PP) of the BBSR capacity is allocated to enable the utility to access it through charging and discharging commands as needed. Through exploration of various combinations of BBSR battery PPs and different ratios of NBBSR and BBSR at each bus, the study identifies the optimal charging and discharging commands from the grid for the existing storage capacity within each sector. This approach empowers utilities to determine the necessary PP levels by engaging in negotiations with BBSRs, taking into account the actual NBBSR and BBSR ratio and their desired power quality standards.

The remainder of this paper is categorized into four sections. In Section 2, details of the proposed methodology is investigated. A multi-objective optimization model is formulated in Section 3. The whole optimization flowchart is also introduced in this Section. The test case is analyzed in Section 4, and the results are demonstrated. Finally, the major contributions of the present work discussed in Section 5.

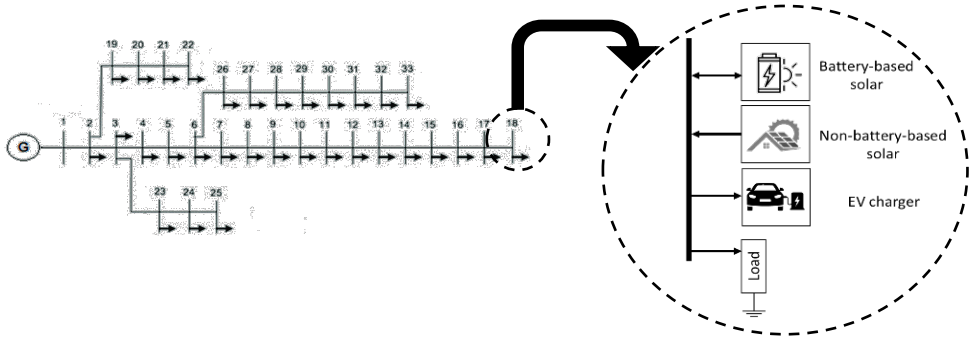


Figure 1: A typical structure of the Bus arrangement used in this study

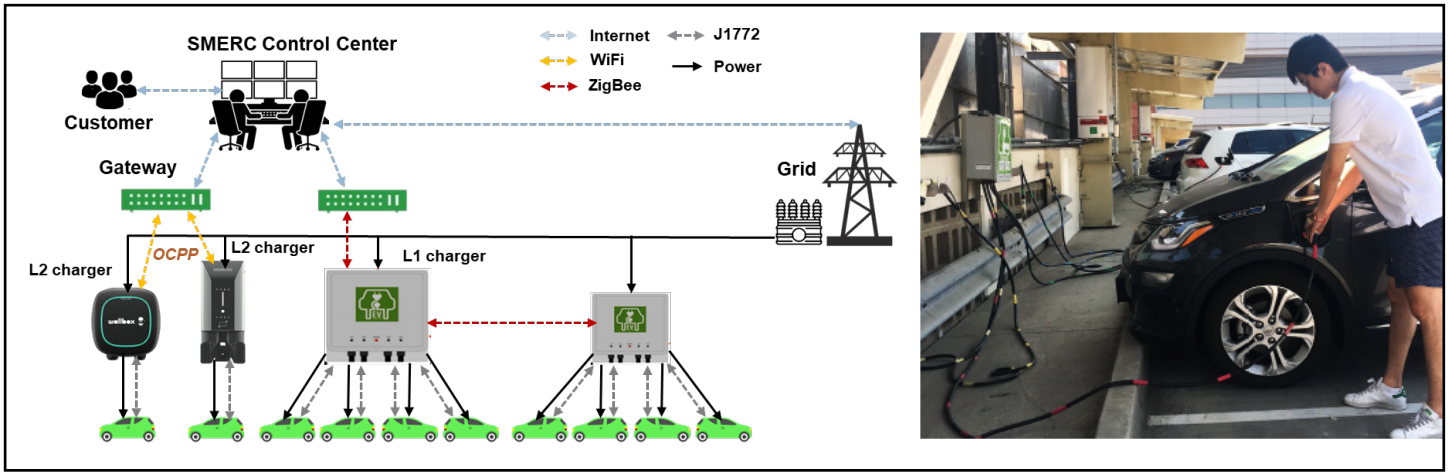


Figure 2: UCLA Smart Grid Energy Research Center charger network architecture

2. System Model

EV energy usage data was gathered from UCLA Smart Grid Energy Research Center charging station as illustrated in Figure 2. The parking structure has different level 1 and 2 chargers. The level 1 chargers talk among themselves and the gateway via ZigBee, and level 2 chargers talk to the gateway via OCPP.

The primary objective of this study is to address the challenges posed by the increasing adoption of EV chargers and solar systems within distribution networks, specifically targeting power losses and deviations in bus voltage. To achieve this goal, the proposed model has been developed in accordance with the IEEE-33 bus radial distribution standard based on [14]. It is assumed throughout this study that every residential dwelling connected to each bus incorporates a solar power generation system. Figure 1 presents an overview of the distribution system encompassing all buses adhering to the IEEE-33 [15] bus standard, encompassing EV chargers, two types of solar systems (BBSR and NBBSR), as well as residential load. Additionally, we introduce the concept of "PP," denoting a percentage of BBSR's capacity that utilities can employ to issue charging and discharging commands to/from the grid as needed.

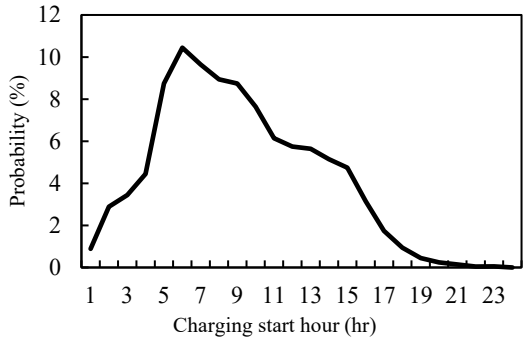


Figure 3: Distribution of the charging start hour for weekdays

Figure 3 shows our assumption on the average distribution of a charging start hour during weekdays, derived from [16].

Batteries serve as short-term energy storage devices and are commonly utilized in conjunction with solar systems, referred to as BBSR [17]. Furthermore, the energy available for each bus can be determined using the following calculation [17]:

$$E_{BT,max} = \beta \times N_{User} \times E_{BT,user} \quad (1)$$

where β, N_{User}, and $E_{BT,user}$ represent the PP of BBSR's capacity, number of residential houses per each bus, nominal storage capacity of each BBSR (kWh), respectively. Additionally, for each time interval, the calculation of available energy per bus is as follows [17]:

$$E_{BT,i}(t) = E_{BT,i}(t-1) + \sum_{t=1}^{T} P_{BT(t)} \cdot \Delta(t) \quad (2)$$

where T, $E_{BT,i}$ (kWh) and $P_{BT,i}$ (kW) are number of hours of operation, the stored energy, and the dispatched power from the utility determined by the optimization algorithm at time interval t for bus i. Finally, the state of charge of BBSR's batteries (SOC_{BT}) calculated as [17]:

$$SOC_{BT,i}(t) = 100 \times \frac{E_{BT,i}(t)}{E_{BT,max}} \quad (3)$$

3. Proposed Methodology

The proposed approach is designed to reduce the total active power loss and voltage deviations over a 24-hour operational period, enabling the utility to implement optimal charging and discharging commands through Particle Swarm Optimization (PSO). Figure 4 illustrates the flowchart outlining this methodology [14].

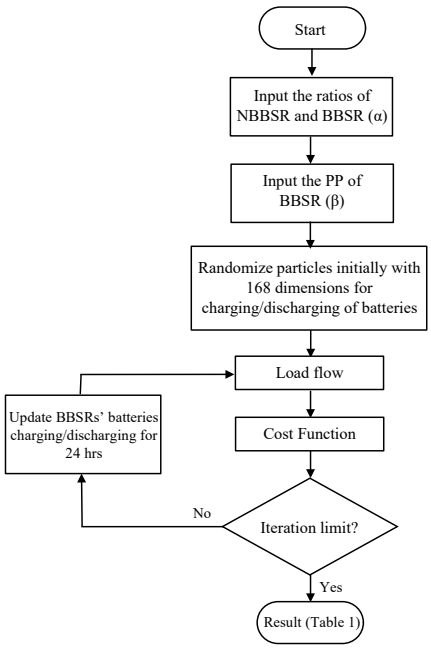


Figure 4: Flowchart of overall proposed optimization procedure

3.1. Cost Function

The cost function in this study can be expressed as the weighted combination of both active power loss and voltage deviation. Active power loss represents the amount of energy dissipated as heat in the distribution network, while voltage deviation measures the deviation of bus voltages from their nominal values. Voltage deviation exceeding a limit can affect electric equipment's lifetime and trigger the protection relays. They can be represented as follows [15]:

$$P_{loss} = \sum_{t=1}^{T} \sum_{k=1}^{N} I_{t,k}^2 R_{t,k} \quad (4)$$

$$V_{dev} = \sum_{t=1}^{T} \sum_{k=1}^{N} |V_{rated} - V_{t,k}| \quad (5)$$

$$\text{Cost Function} = \min\left(w_1 . P_{loss} + w_2 . V_{dev}\right) \quad (6)$$

where V_{rated}, V_K , V_{dev}, I_K , R_K , P_{Loss} , T and N are rated voltage, the voltage of each bus, sum of all buses' voltage deviation, feeder current loading, line resistance, total active power loss, the total number of hours of operation and the total number of lines in the radial distribution network, respectively.

3.2. Particle Swarm Optimization

PSO is an optimization technique that was created by Kennedy and Eberhart [15]. It is inspired by the behavior of bird flocking and fish schooling. PSO is composed of a set of particles that make up a group. Each particle searches its local space to find the local minimum or maximum. The velocity and position of each particle can be updated according to its best experience and the best experience of its neighbors [15]. The variables of each PSO's particle are important factors to guarantee the optimal solution [16]. Using a revolutionary optimization algorithm for a high number of variables can entail significant drawbacks, including the potential for high computational costs and the curse of dimensionality, where the exponential growth in search space makes efficient exploration challenging. Additionally, understanding algorithm behavior and diagnosing issues can be harder, leading to overfitting, generalization problems, and difficulty in visualizing the optimization landscape. Finally, there's an increased risk of premature convergence to suboptimal solutions, making optimization in high-dimensional spaces particularly challenging.

The application of PSO in this study involves optimizing the charging and discharging commands of residential battery storage systems within the distribution network. PSO is employed to find the optimal values for parameters that govern the charging and discharging process, ensuring the reduction of active power losses and voltage deviations. Specifically, the PSO algorithm iteratively refines the charging and discharging commands based on the performance metrics defined by the cost function, converging towards an optimal solution that minimizes power losses and maintains grid voltage stability.

In distribution systems, the number of buses can be relatively high, which leads to weak optimization results. In this study, the IEEE-33 bus has been subdivided into seven sectors, as depicted in Figure 5, to effectively address the challenge posed by a high number of variables. This partitioning of the area is based on presumed demographic conditions and charging demands. Therefore, the charging/discharging command can be strategically allocated to each of these sectors to reduce the number of decision variables for the optimization algorithm (see Figure 5).

3.3. Load Flow

The backward/forward sweep method is used for load flow. The backward/forward sweep is an iterative method in which, at each iteration two computational stages are performed. It is one of the most effective methods for load flow of radial distribution systems [18]. The backward/forward sweep method was chosen for its effectiveness in analyzing radial distribution systems. This method allows for an iterative calculation of currents and voltages, starting from the load buses and moving towards the substation and vice versa. Its suitability for radial systems simplifies the load flow analysis, and its step-by-step approach facilitates understanding.

Below's a brief overview of how the backward/forward sweep load flow method works:

Backward Sweep:

- Start from the load buses (the farthest points from the substation) and work your way back toward the substation.
- At each load bus, calculate the current injected into the bus by the load.
- Use Kirchhoff's Current Law to calculate the current leaving the bus toward the substation.

Forward Sweep:

- Start from the substation and move outward to the load buses.
- At each bus (except the substation), calculate the current entering the bus based on the current calculated in the backward sweep.
- Use Ohm's Law to calculate the voltage at each bus based on the current and the impedance (resistance and reactance) of the transmission lines and transformers.

Active power consumption per bus is the variable for this method in each iteration, which is obtained from the following equations

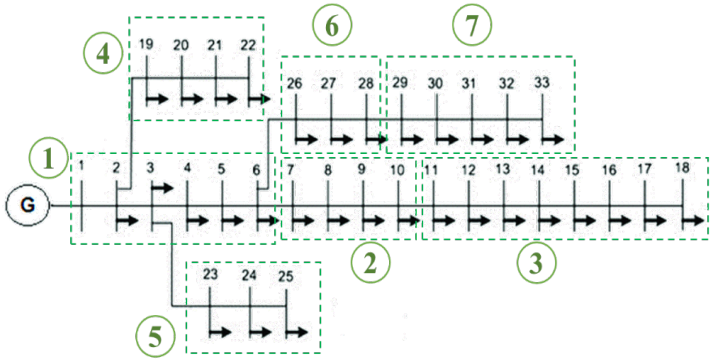


Figure 5: Modified IEEE-33 bus with 7 sectors

$$P_{Load,i}(t) = P_{T,Solar,i}(t) + P_{EV,i}(t) + P_{Res,i}(t) \quad (7)$$

$$P_{T,Solar,i}(t) = \alpha \times N_{B,i} \times P_{Solar,i}(t) \quad (8)$$

$$\alpha = \frac{N_{NBBSR}}{N_{NBBSR} + N_{BBSR}} \quad (9)$$

where $P_{Load,i}$, $P_{T,Solar,i}$, $P_{Solar,i}$, $P_{EV,i}$, $P_{Res,i}$, α, $N_{B,i}$, N_{NBBSR}, and N_{BBSR} are load, total generated solar energy by NBBSR, generated solar per NBBSR unit, EV load, residential load, proportion coefficient, number of residences, number of NBBSR, and the number of BBSR at bus i, respectively.

4. Results

The IEEE-33 bus radial distribution system has been meticulously implemented as the test case using MATLAB 2022b, comprising 33 buses and an intricate network of 32 distribution lines. These lines exhibit varying current-carrying capabilities, with lines connecting node-1 to node-9 accommodating robust capacities of 400 A, while the remaining lines exhibit 200 A capacity. In this meticulously crafted system, the combined losses in active and reactive power are quantified at 281.58 kW and 187.95 kVAR, respectively. It's imperative to note that the baseline voltage is standardized at 11 KV.

The empirical foundation for this research derives from real-world data collected from the UCLA charging station, particularly focusing on EV charging behaviors. Furthermore, solar energy generation patterns for each NBBSR have been meticulously modeled, drawing inspiration from the empirical distributions captured in Figure 6. This data stems from the UCLA 35 kW solar plant, albeit downscaled to a maximum of 10 kWs to align with the research's temporal requirements. Complementing this, the EV load distribution per individual bus is aptly visualized in Figure 7, providing a granular perspective on energy consumption patterns.

The implementation of seven distinct control sectors is a pivotal aspect of this study. Remarkably, this strategic segmentation substantially reduces the number of optimization variables from an initial count of 768 (32 buses × 24 hours) down to 168 (7 sectors × 24 hours). It is essential to note that the number of residences for every bus is the same and equal to 92.

Through the deployment of the PSO algorithm, a comprehensive exploration of various combinations of the optimization parameters α and β is undertaken. The outcomes of this rigorous experimentation are shown in Figure 8, where it is observed that that an insightful pattern emerges. Voltage deviations from the nominal values and power losses exhibit decreasing trends with the progressive augmentation of β for each α. Moreover, the cost function showcases a dynamic response, reducing as α escalates from 0% to 80% and then increasing as α surpasses the 80%. Also, figure 8 indicates that the lowest cost function happens where α equals 70% and β equals 30%.

The numerical values for 77 evaluated combinations of α and β are summarized in Table 1. To show the significance of these findings, it should be noted that the cost function for an unoptimized network is 80. Notably, the achieved cost functions marked improvements in voltage stability and substantial reductions in power loss across the distribution network.

In alignment with the study's objectives, voltage limits within an acceptable range are strategically set at ±10% of the nominal voltage, as articulated in [19].

In radial distribution networks, the end bus of each branch usually faces the highest voltage drop. To evaluate the proposed methodology's efficacy, the voltage profiles of buses 18 and 33, which are the end bus of their own branches are analyzed. Figure 9 illustrates the voltage profiles of buses 18, representing the end buses of their branches. The profiles correspond to the configuration where α equals 70% and β equals 30%, achieving the lowest cost function as highlighted in Table 1. As seen in Figure 9, the proposed methodology consistently meets the specified voltage limits, showcasing its effectiveness in maintaining grid stability. While in other scenarios the voltage profile violates the 10% deviation. A similar pattern is observed in Figure 10 for the voltage profiles of bus 33, reinforcing the robustness of our approach in different scenarios.

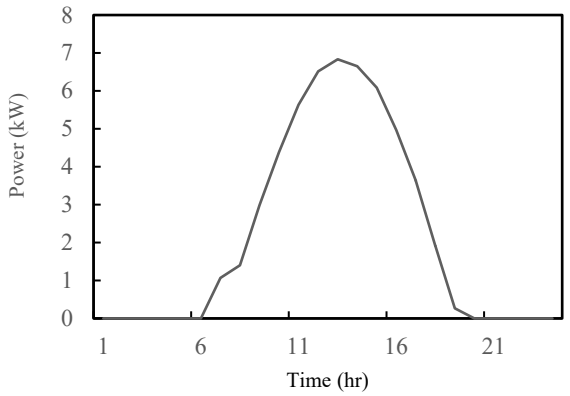


Figure 6: Generated solar

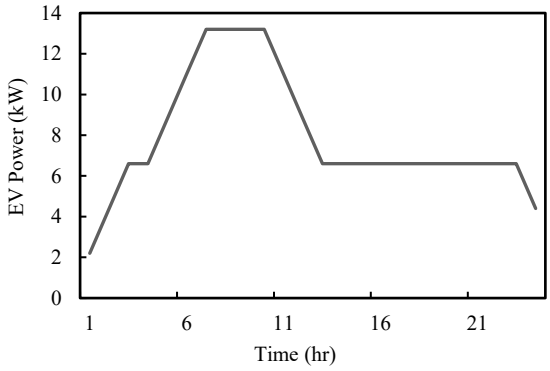


Figure 7: EV load per bus

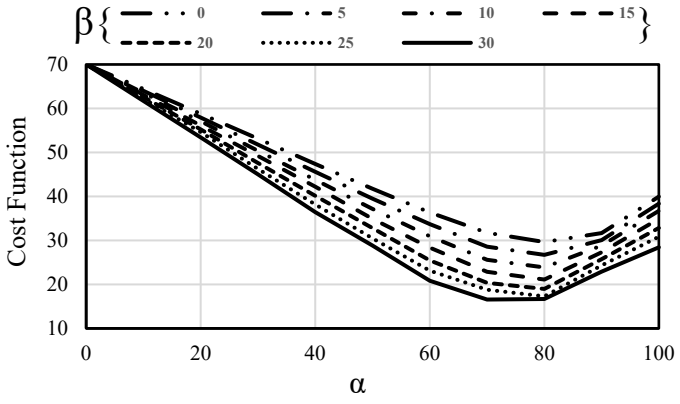


Figure 8: Cost Function value for different combination

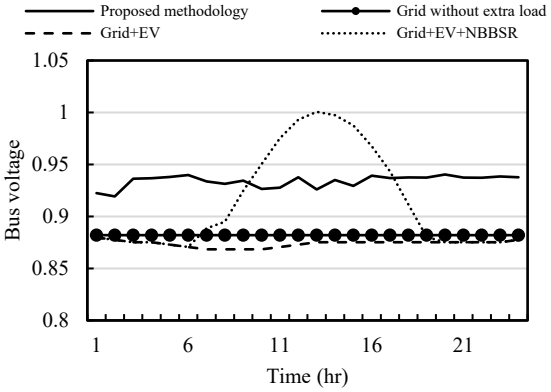


Figure 9: Bus 18 voltage for different scenarios during 24 hr of operation

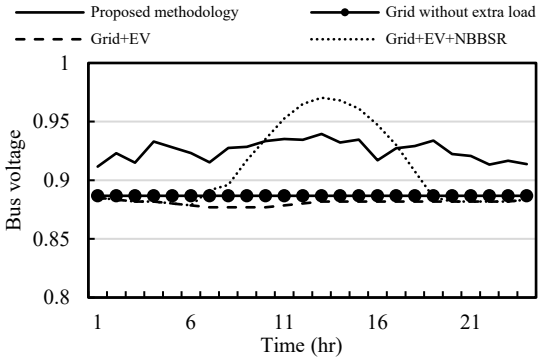


Figure 10: Bus 33 voltage for different scenarios during 24 hr of operation

Table 1 provides numerical values of Figure 8 for various combinations of α and β, showcasing the impact of these parameters on voltage stability, power losses, and the overall cost function. Table 2 further compares the proposed methodology with alternative scenarios. It provides a comprehensive comparative analysis among the proposed methodology with α set at 70% and β at 30% and alternative scenarios, considering key metrics such as total active power loss, average voltage deviation across all 33 buses, and cost function values.

The outcomes of our analysis reveal remarkable improvements in various performance metrics. Specifically, we observe a substantial 47.8% reduction in active power losses when compared to Grid without extra loads. Furthermore, the reduction in active power loss is a notable 53.2% compared with the Grid+EV scenario. Also, the proposed methodology registers a commendable 33.6% decrease in active power losses in comparison with GRID+EV+NBBSR scenario. These findings underscore the remarkable efficacy of our proposed methodology in enhancing the overall efficiency of the distribution network.

Table 1: Cost Function (equation 6) for different combination of α and β

α \ β	0%	5%	10%	15%	20%	25%	30%
0%	70.01663	70.01663	70.01663	70.01663	70.01663	70.01663	70.01663
10%	64.44564	63.98924	63.53546	63.09185	62.60636	62.22017	61.70375
20%	58.81144	57.91499	57.07909	56.09525	55.22481	54.61636	53.40737
30%	53.11391	51.76321	50.36399	49.20472	47.59238	46.2408	45.02529
40%	47.39317	45.58555	43.77159	42.20241	40.14026	38.16705	36.43257
50%	41.75065	39.50142	37.14781	34.90428	32.78993	30.47759	28.8417
60%	36.36633	33.71627	30.93126	28.34218	25.45428	23.10252	20.82519
70%	31.73868	28.55686	25.57219	22.87464	20.36441	18.78012	**16.55407**
80%	29.62956	26.71179	23.85455	21.09941	18.96617	17.28412	16.66144
90%	31.67723	30.07172	28.833	27.30397	25.7509	24.32898	22.90161
100%	39.96917	38.34332	36.70755	34.89518	32.83336	30.8411	28.44189

Table 2: Comparison between the proposed methodology and other scenarios for 24 hr of operation

	Grid without extra load	**Grid+EV**	**GRID+EV+NBBSR**	**Proposed methodology**
Active power loss (kW)	6758.1	7541.6	5314.7	3527.4
Average voltage deviation (%)	6.99	7.41	5.33	4.34
Cost function	26.5094	28.6454	21.1694	16.55407

Our analysis demonstrates an impressive average voltage deviation reduction of 37.9% when contrasted with the Grid without extra loads. The Grid+EV scenario also benefits significantly, with a 41.4% reduction in average voltage deviation. Also, the proposed methodology can achieve 18.6% reduction in average voltage deviation of GRID+EV+NBBSR scenario. These results highlight the transformative potential of our approach in achieving enhanced grid performance and reliability across various scenarios.

The results are striking, revealing a substantial 37.55% reduction in the cost function when compared to Grid without extra loads, a remarkable 42.21% decrease relative to the Grid+EV scenario, and a noteworthy 21.8% deduction in comparison with the GRID+EV+NBBSR scenario.

5. Conclusion

In this study, we have presented an innovative approach to enhance grid power quality without extensive infrastructure upgrades. Our focus was on mitigating active power losses and reducing voltage deviations, critical for ensuring a reliable and stable distribution network. Utilizing the widely recognized IEEE 33-bus system as our testing ground, we explored various scenarios involving PP of BBSR and different ratios of NBBSR to BBSR at each bus. Through optimization, we determined efficient charging and discharging commands, minimizing the need for costly modifications.

Our research demonstrated that several combinations effectively met voltage drop limitations, highlighting the potential for utilities to proactively address challenges posed by DERs. Additionally, electric trucks, while environmentally promising, pose unique charging infrastructure challenges. Our methodology offers a solution by optimizing charging commands, reducing strain on local grids, minimizing disruptions, enhancing electric truck operations, and ensuring overall grid stability. This adaptable approach stands as a strategic tool for utilities and fleet operators, fostering efficiency and sustainability in the face of evolving energy demands.

The outcome of preventing the unnecessary upgrade of the grid from this study not only minimizes power losses and stabilizes the grid but also plays a pivotal role in reducing carbon emissions, particularly in regions where the energy grid relies on a mix of fossil fuels. Implementing the proposed methodology encourages both BBSRs and NBBSRs to expand their solar capacity. The promotion of renewable energy integration, coupled with the optimized utilization of resources, underscores the study's commitment to sustainable practices.

For future works, a notable suggestion involves analyzing the integration of heavy-duty electric vehicles into the grid and exploring the feasibility of utilizing distributed solar generation to meet their charging requirements. This direction addresses the evolving landscape of electric transportation, focusing on the unique challenges posed by heavy-duty electric vehicles and proposing a sustainable approach through the utilization of solar energy.

Acknowledgment

The EV database used in the current research was created by way of a project sponsored in part by LADWP/DOE grant numbers 20699 and 20686, which were part of the LADWP Smart Grid Regional Demonstration Project awarded to the department of Mechanical and Aerospace Engineering and the Smart Grid Energy Research Center (SMERC) at UCLA. We are also grateful for the additional sponsorship received from the following grants 69763, 77739, and 45779 to the Department of Mechanical and Aerospace Engineering.

References

[1] R. Seguin, J. Woyak, D. Costyk, J. Hambrick, B. Mather, "High-Penetration PV Integration Handbook for Distribution Engineers," NREL - National Renewable Energy Laboratory, 1–109, 2016.

[2] B.A. Mather, B.L. Norris, J.H. Dise, L. Yu, D. Paradis, F. Katiraei, R. Seguin, D. Costyk, J. Woyak, J. Jung, K. Russell, R. Broadwater, B.A. Mather, B.L. Norris, J.H. Dise, L. Yu, D. Paradis, F. Katiraei, R. Seguin, D. Costyk, J. Jung, K. Russell, "NREL / SCE High Penetration PV Integration Project : FY13 Annual Report," (June 2014), 2014.

[3] A. Ahmadian, V. Ghodrati, R. Gadh, "Artificial deep neural network enables one-size-fits-all electric vehicle user behavior prediction framework," Applied Energy, **352**, 121884, 2023.

[4] B. Borlaug, M. Muratori, M. Gilleran, D. Woody, W. Muston, T. Canada, A. Ingram, H. Gresham, C. McQueen, "Heavy-duty truck electrification and the impacts of depot charging on electricity distribution systems," Nature Energy, **6**(6), 673–682, 2021.

[5] M. Muratori, B. Borlaug, Perspectives on Charging Medium-and Heavy-Duty Electric Vehicles, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2022.

[6] F.E.R. Commission, FERC Order No. 2222: Fact Sheet | Federal Energy Regulatory Commission, 2022.

[7] T. Bowen, C. Gokhale-Welch, "Behind-The-Meter Battery Energy Storage: Frequently Asked Questions What Is Behind-The-Meter Battery Energy Storage?," (2018), 1–10, 2021.

[8] O. Zinaman, T. Bowen, A. Aznar, "An overview of behind-the-meter solar-plus-storage regulatory design: Approaches and case studies to inform international applications (Report no. NREL/TP-7A40-75283)," NREL Report, (March), 1–76, 2020.

[9] U.S.D. of Energy, Department of Energy National Laboratories and Plants: Mobility Across the Complex, 2013.

[10] T.B. and C. Gokhale-Welch, Behind-The-Meter Battery Energy Storage: Frequently Asked Questions What Is Behind-The-Meter Battery Energy Storage?, 2021.

[11] M. Lavillotti, DER Energy Market Design: Dual Participation, New York Independent System Operator,.

[12] L. Bird, F. Flores, C. Volpi, K. Ardani, D. Manning, R. McAllister, "Review of Interconnection Practices and Costs in the Western States," National Renewable Energy Laboratory, (April), 2018.

[13] Z. Peterson, M. Coddington, F. Ding, B. Sigrin, D. Saleem, K. Horowitz, S.E. Baldwin, B. Lydic, "An overview of distributed energy resource (DER) interconnection: Current practices and emerging solutions," NREL Technical Report, (April 2019), 2019.

[14] R. Zahedi, A. Ahmadian, K. SedghiSigarchi, R. Gadh, "An Optimal Methodology for Mitigating the Impacts of EVs and Solar Systems on the Grid by Utilizing Existing Residential Battery Storage Capacity With No Further Grid Upgrades," in 2023 IEEE Transportation Electrification Conference & Expo (ITEC), IEEE: 1–5, 2023.

[15] R. Zahedi, G.B. Gharehpetian, H. Rastegar, "Environment Compatible Management Strategy of Distributed Generation Based on Neural Network with a Power Capacity Index," in Electrical Engineering (ICEE), Iranian Conference on, 1022–1026, 2018, doi:10.1109/ICEE.2018.8472698.

[16] S. Uimonen, M. Lehtonen, "Simulation of electric vehicle charging stations load profiles in office buildings based on occupancy data," Energies, **13**(21), 2020, doi:10.3390/en13215700.

[17] R. Zahedi, M. Moeini-Aghtaie, "Operational strategy optimization of a hybrid green power system based on fuzzy logic controller with considering for optimal sizing and analysis of different priorities for energy storage," Sustainable Energy, Grids and Networks, **32**, 100809, 2022, doi:10.1016/j.segan.2022.100809.

[18] J.A.M. Rupa, S. Ganesh, "Power flow analysis for radial distribution system using backward/forward sweep method," International Journal of Electrical, Computer, Electronics and Communication Engineering, **8**(10), 1540–1544, 2014.

[19] A. Savić, Ž. Durišić, "Optimal sizing and location of SVC devices for improvement of voltage profile in distribution network with dispersed photovoltaic and wind power plants," Applied Energy, **134**, 114–124, 2014, doi:10.1016/j.apenergy.2014.08.014.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 5, 1-7 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Transmission of the CAP Protocol through the ISDB-T Standard

Nelson Bolívar Benavides Cifuentes, Gonzalo Fernando Olmedo Cifuentes*

Universidad de las Fuerzas Armadas ESPE, Department of Electrical, Electroinics and Telecomunications, Sangolquí, 171103, Ecuador

ARTICLE INFO

Article history:
Received:11 April, 2023
Accepted:02 August, 2023
Online: 24 September, 2023

Keywords:
ISDB-T
CAP
Adalm Pluto
Transport Stream
DSM-CC
Emergency System

ABSTRACT

Early warning systems have had a significant impact on society by providing timely information to mitigate the effects of natural disasters. To enhance early warning capabilities, researchers are exploring the use of digital terrestrial television systems to broadcast alerts across large urban and rural areas. In this research project, the aim is to integrate the global early warning protocol CAP (Common Alerting Protocol) into the existing ISDB-T standard, alongside the standard's Emergency Warning Broadcasting System (EWBS). This integration will enable the creation of a hybrid system, allowing various global emergency alert devices that utilize the CAP protocol to be activated through the Digital Television signal. To achieve this, a CAP to EWBS translator was developed as part of the design proposal prior to transmission. Additionally, transmitters compliant with both the full-seg and one-seg ISDB-T standards were designed to support the CAP protocol. These transmitters utilize SDR (Software-Defined Radio) cards of the Adalm Pluto type. The CAP protocol, encoded in XML format, was transmitted through the ISDB-T transport stream using the DSM-CC data transmission protocol. By incorporating the CAP protocol into the ISDB-T standard and utilizing the DSM-CC data transmission protocol, this research project aims to enhance the early warning capabilities of digital terrestrial television systems.

1. Introduction

Traditionally, Early Warning Systems (EWS) have been implemented to alert the population about potential natural or man-made phenomena and mitigate their impacts [1-2]. To achieve this, various systems have adopted global protocols or developed their own protocols for emergency risk reduction [3-7].

In Ecuador, the ISDB-T International standard is being used for digital terrestrial television implementation. This standard, originating from Japan with Brazilian modifications, is widely adopted in South America [8-9]. The International ISDB-T standard includes an Early Warning Broadcasting System (EWBS) that utilizes area codes to deliver emergency information at the regional level, allowing for targeted alerts within specific areas [10-12].

However, relying solely on the EWBS system within the International ISDB-T standard poses limitations in terms of receiver compatibility and the replication of emergency alerts. As a result, there is a proposition to integrate global alert protocols for emergencies. One widely adopted international protocol is the Common Alerting Protocol (CAP), which interfaces with diverse technologies such as telephony, radio, fax, emails, and websites [13]. Hence, the objective of this article is to incorporate the CAP protocol into the International ISDB-T standard, thereby enabling hybrid digital terrestrial television standards to swiftly and effortlessly transmit emergency messages on a global scale.

The CAP protocol offers an open and non-proprietary digital message format for all types of alerts, utilizing XML for programming, which leads to reduced costs and operational complexity by providing a unified software interface for numerous existing warning systems [14].

The TV 3.0 project is a regional initiative led by Brazil that aims to develop and enhance the television transmission standard to provide viewers with a more advanced and enriched experience. It began in July 2020 and has been divided into several phases to achieve its objectives [15].

*Corresponding Author: Gonzalo Fernando Olmedo Cifuentes, Email: gfolmedo@espe.edu.ec

The first phase of the project involved inviting organizations from around the world to submit proposals for the six components of the TV 3.0 system: over-the-air physical layer, transport layer, video coding, audio coding, subtitles, and application coding [16]. A total of 36 responses were received from 21 different organizations. Similar proposals were merged to simplify testing and evaluations in the next phase [17-23].

The second phase took place during the years 2021 and 2022, focusing on comprehensive testing and evaluations of the proposed technologies and components for the TV 3.0 system. Currently, the third phase of the project is underway and is expected to last approximately two and a half years. A final decision on the over-the-air physical layer technology is anticipated in April 2024, with corresponding technical specifications being drafted from May to August 2024. Concurrently, research and development (R&D) activities are being conducted for the transport layer and application coding, including the development of adaptations, extensions, tools, and test suites. These activities will take place from March 2023 to August 2024. The official launch of the TV 3.0 system is expected to occur in 2025, offering viewers an enhanced television experience and a wide range of services and functionalities.

Within the TV 3.0 project, the implementation of an Advanced Emergency Alert System (EAS) has been identified as a crucial component. The EAS is designed to provide efficient and effective emergency alert notifications to the public through the TV 3.0 platform. It utilizes the Advanced Emergency Alerting (AEA) system based on ATSC 3.0, which is structured in an XML document called the Advanced Emergency Alerting Table (AEAT). This system enables broadcasters to transmit detailed emergency information, including text messages, images, and multimedia content, to inform and adequately prepare the public during emergency situations. Decision trees and protocols have been established to handle different types of alerts, including required alerts, optional alerts, and emergency action notifications, based on priority levels defined in the AEAT [22].

During the transition to digital television, there are still countries in the region that have not completed this process but have digital transmitters that can be utilized to send early emergency alerts through the broadcasting signal. This includes rural areas where television signals are available, as well as locations with Internet access. The research group has been working on projects related to the transmission and reception of emergency alerts using the EWBS system, which is part of the adopted ISDB-T standard in the country [24].

Additionally, the CAP protocol is a global protocol that is part of ATSC 3.0 and will be integrated into the TV 3.0 system. This protocol is used in other global communication mediums as well. Hence, there is a need to demonstrate that through the ISDB-T transport stream structure, it is possible to transmit any type of data, but this requires the implementation of specialized receivers. In this context, this article proposes a method to include the CAP protocol in the ISDB-T system, which is within the standard of the transport layer, and demonstrates its functionality in conjunction with the physical layer of the ISDB-T receiver according to the ISDB-T standard, which requires a special module. To achieve this, the design and implementation using software-defined radio are proposed.

This proposal and innovation aim to expand the capabilities of the ISDB-T system to enable the transmission of emergency alerts using the CAP protocol. By demonstrating the feasibility of this integration, it is expected to improve the efficiency and effectiveness of emergency alert systems in the context of digital television, contributing to the safety and well-being of the population.

This article is structured as follows: Section 2 outlines the methodology employed for information encryption and decryption. Section 3 presents the results obtained from the study, and Section 4 provides the concluding remarks.

2. Methodology

The aim of this study is to incorporate the CAP protocol into a Transport Stream (TS) flow and design a test scenario utilizing Software-Defined Radio (SDR) in both the transmission and reception stages.

To insert the CAP protocol in a TS stream, it is proposed to use the DSM-CC (Digital Storage Media Command and Control) standard. This standard allows the CAP Protocol to be transmitted cyclically using files, directories, services, and event streams through the data and object carousels [25-26].

For the test scenario in the transmission and reception stages, the cost-effective SDR Adalm Pluto will be utilized. This compact and easily transportable device serves various functions in the realms of education and research. The SDR Adalm Pluto operates in conjunction with the GNU Radio software, which is an open-source program compatible with multiple platforms. It allows for modular programming and simplifies the creation of pre-programmed Python blocks for signal transmission or reception [27].

2.1. TS Design

To use the DSM-CC in the International ISDB-T standard, the AIT table must be added to the PSI/SI tables; this table provides complete information about data transmission, the required activation status of the applications carried by it, etc. The data in the AIT table allows the transmitter to request that the receiver change the activation state of an application [28-32].

Within the PSI/SI Tables is the PMT table, which contains all the elements that make up the TS stream, such as audio, video, data, and the AIT table in the Elementary Stream field.

In each Elementary Stream field, some descriptors help to complement the description of each element of the TS stream, which is why the following descriptors are included in the AIT table: the Data Components descriptor and the Application signaling descriptor. These descriptors will indicate the type of application to be used within the TS flow. Its conformation is shown in Figure 1.

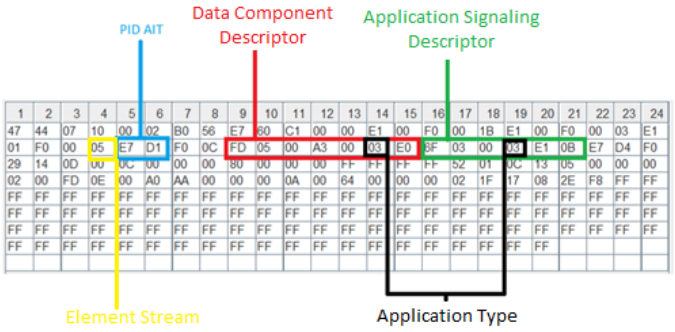


Figure 1: AIT Table Descriptors within the PMT Table

The descriptors for the DSM-CC section element are the TAG Association descriptor, the Identifier descriptor, and the Carousel Identifier descriptor. These descriptors mention the use of the data carousel within the TS stream. Its structure is presented in Figure 2.

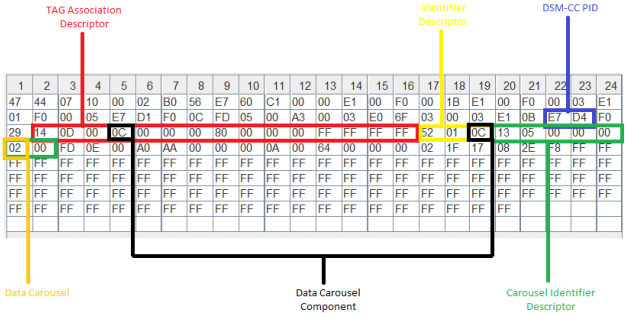


Figure 2: DSM-CC descriptors within the PMT table.

Next, the information that must go in the AIT table is presented, where the type of application is inserted, which must be the same as that shown in the PMT table; in turn, three descriptors are inserted, the first being the protocol descriptor transport indicating the use of data carousel. Second, the application descriptor suggests the priority of the data carousel in the TS flow. Finally, there is the application name descriptor, where the name of the CAP file is inserted in the ASCII code. These values can be seen in Figure 3.

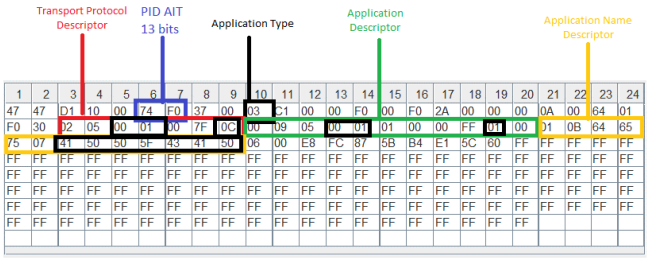


Figure 3: Components of the AIT table.

The TS flow and the DSM-CC section containing the multiplexed CAP protocol will be obtained with an audio and video signal.

2.2. Design of the Transmission and Reception System

With the generated TS stream, we build the transmission system in the GNU Radio software. The transmission system diagram with the GNU Radio software is presented in Figures 4, 5, and 6, and the description of each component is established in Table 1.

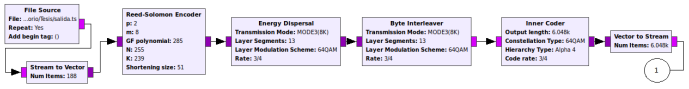


Figure 4: Transmission System in GNU Radio - Coding

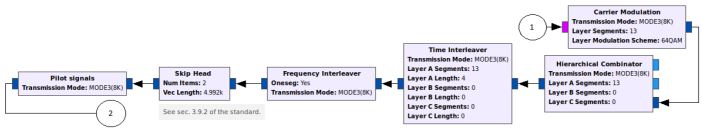


Figure 5: Transmission System in GNU Radio - Modulation

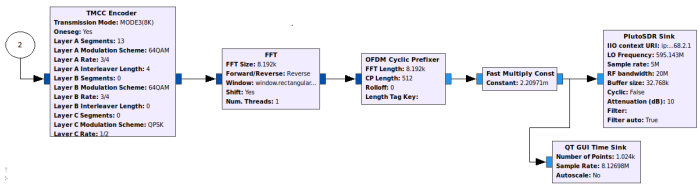


Figure 6: Transmission System in GNU Radio - Configuration

Table 1: Description of each Block of the Transmission System

Block	Description
FILE SOURCE	Read a stream from a file.
STREAM TO VECTOR	Converts an item stream to a GNU Radio block stream.
REED SOLOMON ENCODER	Create a Reed Solomon scrambler according to the standard.
ENERGY DISPERSAL	Implements an energy disperser compliant with the standard.
BYTE INTERLEAVER	Inserts a byte interleaver according to the standard.
INNER CODER	Create an internal encoder with puncturing.
VECTOR TO STREAM	Converts the GNURadio block stream to an item stream.
CARRIER MODULATION	Modulates the carrier with the parameters of Mode, number of segments, and modulation scheme.
HIERARCHICAL COMBINATOR	Divide the content to be transmitted into hierarchical layers.
TIME INTERLEAVER	Implementation of a Fornet interleaver as specified by the standard.
FREQUENCY INTERLEAVER	Performs frequency interleaving.

Block	Description
SKIP HEAD	Remove information considered junk from the header.
PILOT SIGNALS	Establishes pilot signals by the standard.
TMCC ENCODER	Block that encodes the TMCC carriers.
FFT	Implements the fast Fourier transform.
OFDM CYCLIC PREFIXER	Adds a cyclic prefix and performs pulse shaping on OFDM symbols.
FAST MULTIPLY CONST	Multiply the input by a constant to have the desired size.
PLUTO SDR SINK	Block that transmits the signal through the Adalm Pluto.
QT GUI TIME SINK	Graphically presents the transmission.

With these blocks, the SDR Adalm Pluto is connected, and for this software to recognize it, the IP address: 192.168.2.1 is put in the PLUTO SDR SINK block.

For the receiving system, a new program is created in GNU Radio. The reception system diagram is shown in Figures 7, 8, and 9, and the description of each component is established in Table 2.

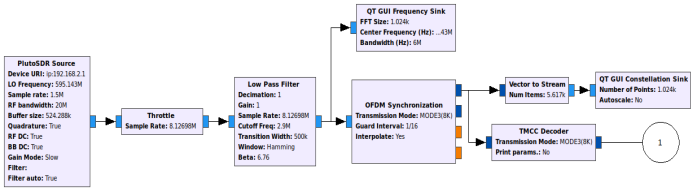


Figure 7: Reception System in GNU Radio – Configuration

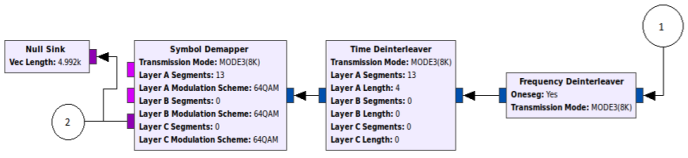


Figure 8: Reception System in GNU Radio - Demodulation

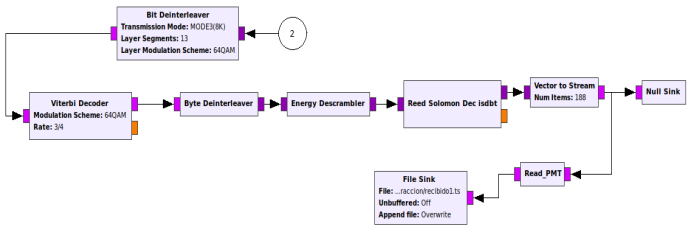


Figure 9: Reception System in GNU Radio - Decoding

Table 2: Description of each Block of the Reception System.

BLOCK	***DESCRIPTION***
PLUTO SDR SOURCE	Receive the signal from Adalm Pluto.
THROTTLE	Indicates the sample rate to use.
LOW PASS FILTER	Create a low pass filter to get the signal from the TS file.
QT GUI FREQUENCY SINK	Graphically indicates the shape of the signal.
OFDM SYNCHRONIZATION	Specifies OFDM information such as Mode, Guard Interval, and interpolation.
QT GUI CONSTEALLATION SINK	Graphically indicates the constellation of the system.
TMCC DECODER	Decodes TMCC carriers.
FREQUENCY DEINTERLEACER	Performs frequency interleaving.
TIME DEINTERLEAVER	Specifies a Forney interleaver as specified by the standard.
SYMBOL DEMAPPER	Removes the symbol map on each carrier and then generates a serial keyword.
BIT DEINTERLEAVER	Performs a bit of interleaving.
VITERBI DECODER	Implements a Viterbi decoder on the signal.
BYTE DEINTERLEAVER	Implements a Forney byte deinterleaver as specified by the standard.
ENERGY DESCRAMBLER	Specifies a power decoder as specified by the standard.
REED SOLOMON DEC ISDBT	Receives 204-byte blocks and performs a Reed Solomon decode.
VECTOR TO STREAM	Transforms the GNU Radio blocks to an elementary stream of 188 bytes for each packet.

BLOCK	***DESCRIPTION***
READ PMT	Block was created to verify the PMT table information mentioning its Elementary Stream types.
FILE SINK	Create a file with the information received.

With the received file, the information from the DSM-CC section is extracted as follows:

1. Look up the PID from the PMT Table and then identify the PID from the DSM-CC section.
2. Compare the PID frame by frame and verify that it matches the PID of the DSM-CC section. When a frame is found with the DSM-CC PID, its content is analyzed.
3. Identify the Table id to indicate if it is a data or control block.
4. Extract the content stored in a TS file as a data block.
5. Being a control block, the type of block is identified.
6. Finally, the number of data and control blocks found is obtained.

Subsequently, the data carousel is extracted, which allows removing the headers of the DBB sections that contain the data. To perform this function, the following steps are followed:

1. The number of modules that were generated in the extraction of the DSM-CC is verified.
2. Each module is analyzed to check if the file still contains information.
3. The message_length segment located in the byte following the table id is analyzed from each frame. This segment will indicate the size of the CAP file.
4. This process is repeated until the beginning and end of the message are stored.

Then, the object carousel is extracted; in this section, the type of file inside the TS is verified, this; case, it will only be a file type file where the CAP file will be obtained in XML format. The steps are the following:

1. The file size generated in the previous function is obtained.
2. The type of file is obtained.
3. Its header is parsed to extract its content.
4. At the end, you will have the extracted content.

3. Results

To verify the operation of the CAP protocol transmission, the TS flow is created with the information from the PAT, PMT, NIT, SDT, AIT, and DSM-CC tables.

Once all these elements have been verified, the signal is transmitted and received with the help of the GNU Radio software and SDR Adalm Pluto, the scenario mentioned in Figure 10.

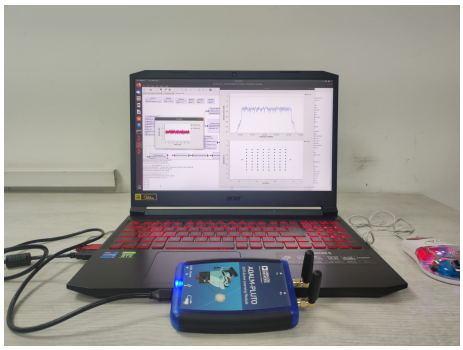


Figure 10: Scenario with GNU Radio and Adalm Pluto.

To check the operation of the communications system, a block called Read PMT was created at the reception; this block will read the information from the PMT and its four elementary Streams: the audio, the video, the AIT table, and the DSM-CC section. This content is indicated in Figure 11.

```
--------------------------------------------------
PMT
--------------------------------------------------
PID: 1031 y  in hex: 0x407
Section length: 86 y  in hex: 0x56
Section syntax error: 1 entonces True
CRC-32: [ 23  8  46 248]
Program number: 59232 y en hex: 0xe760
Version Number: 0 y en hex: 0x0
Current next indicator:  1 entonces True
PCR_PID: 256 y  in hex: 0x100
Program information length: 0 y  in hex: 0x0
--------------------------------------------------
---------- Information 1 ----------
Stream Type identifier: 27  in hex: 0x1b : Video ITU Recommendation H.264 e ISO/IEC 14496-10
Elementary Stream PID: 256  in hex 0x100
ES information length: 0  in hex: 0x0
--------------------------------------------------
---------- Information 2 ----------
Stream Type identifier: 3  in hex: 0x3 : Audio  ISO/EC 11172-3
Elementary Stream PID: 257  in hex 0x101
ES information length: 0  in hex: 0x0
--------------------------------------------------
---------- Information 3 ----------
Stream Type identifier: 5  in hex: 0x5 : Seccion
Elementary Stream PID: 2001  in hex 0x7d1
ES information length: 12  in hex: 0xc
--------------------------------------------------
---------- Information 4 ----------
Stream Type identifier: 11  in hex: 0xb :  ISO/IEC 13818-6 (tipo B)
Elementary Stream PID: 2004  in hex 0x7d4
ES information length: 41  in hex: 0x29
--------------------------------------------------
Stream type identifier: [27, 3, 5, 11]
Elementary Stream PID: [256, 257, 2001, 2004]
ES information length: [0, 0, 12, 41]
```

Figure 11: Content of the PMT Table at the reception

After some time, receiving and transmitting are paused, and the received.ts file is obtained. The steps for removing the DSM-CC section using a Python file are followed to extract the information from this received file, getting Figure 12.

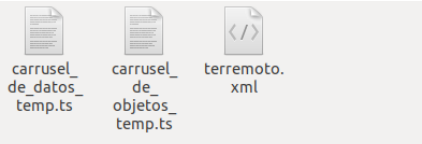


Figure 12: File Extracted from TS Stream.

In the end, the terremoto.xml file is successfully extracted. The cap-validator software checks the CAP file, resulting in a valid file, as shown in Figure 13 [33].

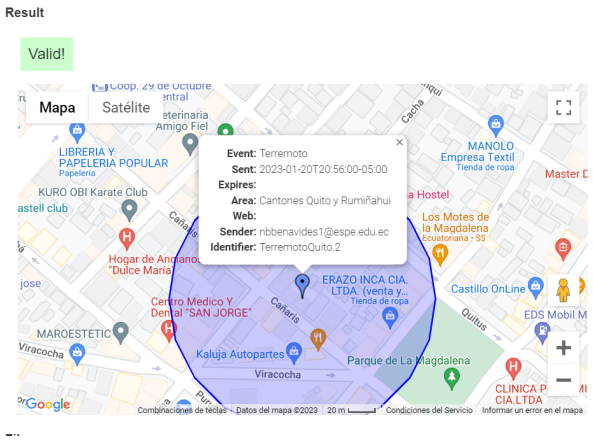


Figure 13. Result of cap-validator software.

The resulting process and the used diagram are shown in the QR code below in Figure 14.

Figure 14: QR code of CAP protocol transmission and extraction display.

4. Conclusions

The present study successfully demonstrated the transmission and reception of CAP protocol data in XML format through the DSM-CC data structure within an ISDB-T transport stream. The utilization of the CAP protocol resulted in the generation of minimal bytes in the DSM-CC, showcasing its efficiency in terms of bandwidth utilization when compared to transmitting an interactive application under the same transmission mode.

By employing GNU Radio and the SDR Adalm Pluto, a robust communication system for ISDB-T with full-seg content was developed. The system exhibited flawless reproduction of audio and video reception, while ensuring the complete download of the CAP file. Furthermore, the implementation of a one-seg reception system using the SDR Adalm Pluto and the DEKTEC DTU 215 modulating card yielded successful results, thanks to the minimal bandwidth requirements of the transmitted CAP file.

To ensure the quality of the CAP file, both during its construction and upon reception in both one-seg and full-seg scenarios, the cap-validator software was employed. This validation process confirmed the accuracy and integrity of the CAP information. Overall, the communications system demonstrated optimal performance and is well-suited to serve as a valuable asset in supporting the EWBS emergency system. By integrating the CAP system into the ISDB-T transport layer platform, emergency authorities and broadcasters can send timely and accurate emergency alerts, ensuring the safety and well-being of the population. The CAP structure, which is both part of ATSC 3.0 and will be part of TV 3.0, aims to provide a powerful tool for public authorities to disseminate news, weather updates, and critical information during emergencies, enhancing the overall emergency response capabilities within the broadcasting industry. Through the proposed approach in this article, the integration of CAP can even be achieved during the transition to digital television using the ISDB-T system, allowing for readiness in any migration to an advanced digital television system in the near future.

References

[1] F. Guzzetti, S. Luigi Gariano, S. Peruccacci, M. Brunetti, I. Marchesini, M. Rossi, M. Melillo, Geographical landslide early warning systems, Earth-Science Reviews, **200**, 2020. ISSN 0012-8252, https://doi.org/10.1016/j.earscirev.2019.102973.

[2] A. J. L. García and R. C. Rojas, "Abstract Extended of Novel Method for broadcasting Multiples Emergency Alert and Alarm Using ISDB-T Digital Terrestrial Television," 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Jeju, Korea (South), 2019, 1-3, doi: 10.1109/BMSB47279.2019.8971946.

[3] G. Olmedo, F. Acosta, D. Villamarín, F. Santander, R. Achig, V. Morocho (2021). Prototype of a Centralized Alert and Emergency System for Digital Terrestrial Television in Ecuador, Advances in Intelligent Systems and Computing, 1326. Springer, Cham. https://doi.org/10.1007/978-3-030-68080-0_14.

[4] S. NURBADRI, K. ANWAR, D. ARSENO, Study On Early Warning Systems (EWS) for Indonesia Digital Terrestrial Television. JMECS (Journal of Measurements, Electronics, Communications, and Systems), [S.l.], **6**(1), 11-19, june 2020. ISSN 2477-7986

[5] Li, Z., Fang, L., Sun, X. et al. 5G IoT-based geohazard monitoring and early warning system and its application. J Wireless Com Network **2021**, 160, 2021. https://doi.org/10.1186/s13638-021-02033-y.

[6] G. Olmedo, F. Acosta, R. Haro, D. Villamarín, N. Benavides. (2019). Broadcast Testing of Emergency Alert System for Digital Terrestrial Television EWBS in Ecuador, Communications in Computer and Information Science, 1004. Springer, Cham. https://doi.org/10.1007/978-3-030-23862-9_13.

[7]. A. J. L. García and R. C. Rojas, "Abstract Extended of Novel Method for broadcasting Multiples Emergency Alert and Alarm Using ISDB-T Digital Terrestrial Television," 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Jeju, Korea (South), 2019, 1-3, doi: 10.1109/BMSB47279.2019.8971946.

[8] M. Takada and M. Saito, "Transmission System for ISDB-T," in Proceedings of the IEEE, **94**(1), 251-256, Jan. 2006, doi: 10.1109/JPROC.2005.859692.

[9] L. C. d. P. Costa, C. S. Kurashima, M. G. De Biase, R. H. Alonso and M. K. Zuffo, "A technical analysis of digital television broadcasting in brazil," 2013 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), London, UK, 2013, 1-5, doi: 10.1109/BMSB.2013.6621795.

[10] ISDB-T HARMONIZATION DOCUMENT PART 3: Emergency Warning Broadcast System (EWBS) (November/2015) by ISDB-T International Forum.

[11] ARIB STD-B24:2008, Data coding and transmission specification for digital broadcasting.

[12] ARIB TR-B14:2006, Operational guidelines for digital terrestrial television broadcasting.

[13] P. K. Dalela, A. Saldhi, P. Bhave and V. Tyagi, "Common Alerting Protocol Compliant Emergency Warning And Alert System for Legacy Broadcasting Networks," 2020 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA), Victoria, BC, Canada, 2020, 1-5, doi: 10.1109/CogSIMA49017.2020.9216105.

[14] ITU-T, ITU-T Rec. X.1303 bis (03/2014) Common alerting protocol (CAP 1.2), 2014.

[15] FORUM ISDB-T, "TV3.0 Project", 2020, https://forumsbtvd.org.br/tv3_0/.

[16] FORUM ISDB-T, "Call for Porposals: TV3.0 Project", 17 July 2020, https://forumsbtvd.org.br/wp-content/uploads/2020/07/SBTVDTV-3-0-CfP.pdf.

[17] FORUM ISDB-T, "Testing and Evaluation Report: TV 3.0 Project - Over-the-air Physical Layer Laboratory Tests", 03 December 2021, https://forumsbtvd.org.br/wp-content/uploads/2021/12/SBTVD-TV_3_0-PL-Lab-Report.pdf.

[18] FORUM ISDB-T, "Testing and Evaluation Report: TV 3.0 Project - Over-the-air Physical Layer Field Tests", 03 December 2021, https://forumsbtvd.org.br/wp-content/uploads/2021/12/SBTVD-TV_3_0-PL-Field-Report.pdf.

[19] FORUM ISDB-T, "Testing and Evaluation Report: TV 3.0 Project - Transport Layer", 03 December 2021, https://forumsbtvd.org.br/wp-content/uploads/2021/12/SBTVD-TV_3_0-TL-Report.pdf.
[20] FORUM ISDB-T, "Testing and Evaluation Report: TV 3.0 Project - Video Coding", 03 December 2021, https://forumsbtvd.org.br/wp-content/uploads/2021/12/SBTVD-TV_3_0-VC-Report.pdf.
[21] FORUM ISDB-T, "Testing and Evaluation Report: TV 3.0 Project - Audio Coding", 03 December 2021, https://forumsbtvd.org.br/wp-content/uploads/2021/12/SBTVD-TV_3_0-AC-Report.pdf.
[22] FORUM ISDB-T, "Testing and Evaluation Report: TV 3.0 Project - Captions", 03 December 2021, https://forumsbtvd.org.br/wp-content/uploads/2021/12/SBTVD-TV_3_0-CC-Report.pdf.
[23] FORUM ISDB-T, "Testing and Evaluation Report: TV 3.0 Project - Application Coding", 03 December 2021, https://forumsbtvd.org.br/wp-content/uploads/2021/12/SBTVD-TV_3_0-AP-Report.pdf.
[24] G. Olmedo and Roman Lara-Cueva, "Evaluation of Emergency Warning Broadcast System (EWBS) for Digital Terrestrial Television in Ecuador", IEEE Global Communications Newsletter, December 2021, https://gcn.comsoc.org/evaluation-emergency-warning-broadcast-system-ewbs-digital-terrestrial-television-ecuador.
[25] Hongguang Zhang, Tianpu Jiang, Zhiqi Gu and Shibao Zheng, "Design and implementation of broadcast file system based on DSM-CC data carousel protocol," in IEEE Transactions on Consumer Electronics, **50**(3), 929-933, Aug. 2004, doi: 10.1109/TCE.2004.1341702.
[26] G. Olmedo, A. Nunez and D. Villamarín, "Design, implementation and evaluation of data carrousel extractor algorithm on MPEG-2 TS for digital terrestrial television," 2016 XLII Latin American Computing Conference (CLEI), Valparaiso, Chile, 2016, 1-8, doi: 10.1109/CLEI.2016.7833325.
[27] GNU RADIO, GNU Radio The Free Software & Open Software Radio Ecosystem, 2021.
[28] RTS/JTC-DVB-371, TS 102 809 - V1.3.1 - Digital Video Broadcasting (DVB); Signalling and carriage of interactive applications and services in Hybrid broadcast/broadband environments, 2017.
[29] Associação Brasileira de Normas Técnicas, ABNT NBR 15602-3: Codificação de vídeo, áudio e multiplexação Parte 3 Sistemas de multiplexação de sinais, 2007.
[30] Associação Brasileira de Normas Técnicas, ABNT NBR 15603-1: Multiplexación y servicios de información (SI) Parte 1 SI del sistema de radiodifusión, 2007.
[31] Associação Brasileira de Normas Técnicas, ABNT NBR 15603-2: Multiplexación y servicios de información (SI) Parte 2 Estructura de datos y definiciones de la información básica de SI, 2007.
[32] Associação Brasileira de Normas Técnicas, ABNT NBR 15603-3: Multiplexación y servicios de información (SI) Parte 3Sintaxis y Definiciones de Información extendida del SI, 2007.
[33] Google, CAP-VALIDATOR, 2014.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 5, 8-16 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

xASTESJ

ISSN: 2415-6698

Augmented Reality Based Visual Programming of Robot Training for Educational Demonstration Site

Lucksawan Yutthanakorn, Siam Charoenseang*

Institute of Field Robotics, King Mongkut's University of Technology Thonburi, Bangkok, 10140, Thailand

ARTICLE INFO

Article history:
Received:27 July, 2023
Accepted:02 September, 2023
Online: 24 September, 2023

Keywords:
Augmented Reality
Educational Demonstration Site
Smart Factory
Robot Training System
Visual Block-based Programming

ABSTRACT

The human resource development of robotics and automation in the smart factory is an important factor in "Thailand 4.0" roadmap, which is following the industry 4.0 model. To pursue this goal of Thailand 4.0 roadmap of labor development, the effective and intuitive training system must be easy to understand. This study proposes the implementation of augmented reality (AR) technology for training purposes due to its ability to visualize real-time invisible data, such as device status. This involves the development of the AR-based visual programming interface with an educational demonstration site (demo site). The AR-based training system is started with animation content explaining the smart factory concept, followed by hands-on learning using Microsoft HoloLens 2 and IoT hardware devices in demo site. The demo site using an MQTT protocol, simulates an automated packing line in the smart factory. The hardware status is published in real-time to the MQTT broker. This approach enables users to comprehend the interconnected relationship between data and hardware functionality and allowing them to create their own programs to control the IoT hardware in the demo site. With this training, 22 targeted users have successfully grasped the smart factory concept and its correlation with hardware functionality. The block-based visual programming employed in the system enables easy comprehension of robot commands. Moreover, the AR application provides a smooth display at 48-60 frames per second. The proposed system's usability and value for specific tasks received high scores, ranging from 4 to 5 points, confirming its effectiveness for the targeted users. The proposed AR-based training system has proved that it has benefit for human resource development in the robotics and automation sector following the Thailand 4.0 roadmap. The proposed system empowered users to understand the smart factory concept and its practical implementation, leading to create new ideas for integrating AR and smart factory concepts into their manufacturing in the future.

1. Introduction

The industry 4.0 is the era of using the cyber-physical system [1]-[2] to manage data in manufacturing, which could be called as the smart factory. By transforming production into a digital process and connecting factories to the internet, industry 4.0 can enhance manufacturing performance. Germany is the first nation that mentions about the Industrie 4.0. The German government announces as the national strategic initiative [3] and establishes the platform for industry development to 4.0 and the criteria for evaluating the potential of the factory's deve lopment. The United State of America (USA) has an important strategy called the Smart Manufacturing. The strategy aims to manage the investments in manufacturing. The National Institute of Standards and Technology (NIST) of USA supports about implementing the robotics in manufacturing and creates the tests for evaluating potential of the automation in the production line. The quality control is the key to modern factory [4]. In Southeast Asia, Singapore prepares the Smart Industry Readiness Index (SIRI) for measuring industrial development readiness. The method emphasizes three dimensions are the process, the technology, and the organization. SIRI shares the framework for developing the industry to industry 4.0.

In Thailand, there is a model for developing the industry called Thailand 4.0.The model aims to help small and medium-sized enterprises (SMEs) become high-potential enterprises. Thailand 4.0 aims to implement the robotics and automation in the production line and labor skill improvement [5]. The Thailand 4.0 strategy identifies five new S-Curve industries: robotics, aviation and logistics, digital, biofuels and biochemicals, and medical hub.

*Corresponding Author: S. Charoenseang, KMUTT, Bangkok, 10140, Thailand, siam.cha@kmutt.ac.th

[6]. Labors need to improve their skills to support these industries. 70% of factories in Thailand are in 2.0-3.0 stage of industry [7]. The labors and entrepreneurs should prepare for the digital disruption and industry 4.0. These five new targeted industries need high-potential workers and the experts to drive them forward.

Although training in the industry is popular, there are few courses that offer hands-on workshops with hardware or demonstration sites. However, the industrial training system can benefit from the implementation of augmented reality (AR) technology. An AR-based application provides the visualization of data from real IoT equipment installed in an educational demonstration site (demo site). The demo site simulates the working of an automated packing line, demonstrating how the smart components work such as robot arm for pick-and-place jobs or sensors for detecting palettes. Furthermore, the AR-based application of the proposed system provides a visual block-based programming panel that supports beginner user in creating their own programs. The proposed solution, which combines the AR technology and the demo site, aims to increase worker's understands of the fundamentals of the smart factory concept and improve their robotics and automation skills.

2. Related Works

2.1. Comparison of AR, VR, and MR concepts

In the present, the augmented reality, virtual reality, and mixed reality are implemented in real life with various purposes, such as medical training application, assembly assistance task, augmented reality with data visualization in production line [8] and using these technologies for training system.

The experts explain the concept of augmented reality (AR) as a merging of 3D graphics with the real world [9]. There are three parts of AR-based system which are user, digital world, and real world. User can control the device in real world by interacting with augmented graphics appearing in the head-mount display device such as Microsoft HoloLens.

Virtual reality (VR) is created by all 2D or 3D graphic environment. User does not see any parts of real world. 2D or 3D graphic and virtual environments need a specific device to present the contents. User can see the virtual contents by using headset with head tracker and hand controller to interact with these graphics.

There is no specific clear definition for mixed reality (MR) that can explain the differences between MR and AR. However, the continuum of the popular source defines MR as the technology between the fully real world and fully virtual world [10]. In the other words, AR is a type of MR. Both AR and MR have augmented objects attached to the real environment, but MR may have more complex features such as attaching graphic textures with real characters from the real world, which is called Augmented Virtually – AV. The summary characteristics of AR, VR, and MR are shown in Table 1.

The proposed system aims to display status data in real time while hardware in the demo site is working. The user should be able to see the variation of the data while also seeing the real environment. Therefore, augmented reality technology would be more appropriate for this system than virtual reality.

Table 1: AR, VR, MR Characteristics Comparison

Topic	AR	VR	MR
Can see real environment through the device.	✓	×	Only specific area.
Environment [11]	Real world	Virtual world	Mixed
Can interact with virtual object	✓	✓	✓
Can interact with physical object [12]	✓	×	Only specific object.

2.2. Augmented Reality in Training and Industry

AR is frequently used in daily life, such as in the camera filters of social media applications, image tracking of 2D markers for modern-day advertising, and training systems in the medical field and industry.

Research on augmented reality technology for training has found that AR-based training systems are most commonly used in industry (35%), vocational training (26%), and medical (13%) applications [13]. Another research in training topics using AR or VR technology showed the successful implementation about AR-based training system in maintenance and assembly task [14]. The traditional way to learn the welding process is to practice with real equipment. However, virtual and augmented reality can now be used to create simulation training systems that are more flexible and cost-effective than the traditional methods. A study that compared two groups of students who used VR and AR for training and practical application found that the group that used the simulation training system had lower operating stress and higher levels of focus than the other group [15].

Furthermore, AR technology is not only for training system, but this technology can also be used for data visualization in CNC machines. The data, which can be inspected in real time by integrating AR technology with digital twins, is helpful for making the best decisions in the future [16].

2.3. Thailand 4.0, Thailand Professional Qualification Institute, and the survey about the smart factory training topics in Thailand.

Thailand is now a developing country. A roadmap to drive Thailand to be a developed country in 2032 is called Thailand 4.0 [17]-[18]. One of the agenda in the roadmap is about labor improvement.

There are 5 new s-curves industries including robotics for industry, aviation and logistics, biofuels and biochemicals, digital, and medical hub. To develop and re-skill labors in industry field requires a necessary skill about robotics and automation for digital transformation and industry 4.0 revolution.

The Thailand Professional Qualification Institute – TPQI creates career-standard certifications for qualified individuals. The

robotics and automation standard is under the robotics cluster in Mechatronics standard including 19 professional qualifications, 66 credits. This is a standard for 5 career levels, which are for operator, technician, engineer or staff, manager, system design engineer, and system integrator [19].

A review of training sessions showed that the curriculum emphasizes automation in industry 3.0, but there is not enough information about smart factories. Most training courses did not offer hands-on hardware practice. There are 3 of the 16 topics that contribute to the demo site for students, as shown in Table 2.

Table 2: Survey of Training Topics

Organization	Area of content	Providing the demo site for hands-on practice
Thai-German Dual Education and E-Learning Development Institute	Wireless connection, IoT, Programming	×
	Database	×
	Remote control, Sensor and actuator, Simulation, Programming	×
	Product management, Vision system, Robot and automation, Sensor and actuator, Programming	×
Technology Promotion Association (Thailand-Japan)	IoT, Sensor and actuator, Programming	×
Thailand Productivity Institute	AR/VR, IoT, Barcode scanner and actuator	×
	Data, Cyber security	×
	Cloud computing	×
	Robot and automation	×
	Robot and automation, Programming	×
Thai-German Institute (TGI)	Database	×
	Database, sensor and actuator	×
	Database, Wireless connection	✓
	AI	✓
	Wireless connection, IoT, Programming	✓
	Wireless connection, Sensor and actuator	×

Developing human resources for interpreters and engineers about the smart factory contents is beneficial for the company. This is the opportunity to develop an appropriate training system for this group using new technology and augmented reality to create a new training experience in industry. The benefits of AR technology, such as data visualization and real-world viewing, can help users easily understand the relationship between device operation and data. Additionally, user's understanding of the smart factory concept can be improved by watching the video animations provided in this proposed system.

3. Proposed System

The Thailand 4.0 roadmap requires human development to support the robotics and automation industry which is one of five new S-curves industries. To achieve this, reskilling and upskilling efforts in robotics, particularly in smart factory operations, are also required.

This proposed system is an AR-based training system for smart factory concept. It implements augmented reality (AR) technology to create an effective and intuitive training system. The system scenario while using the AR-based application of this proposed system is shown in Figure 1.

The contents of this system include three animation videos that cover the basics of smart factory operations and robot arm programming. The AR application includes content about the automation devices used in smart factories. The demonstration site (demo site) of smart factory simulates how the automation devices in a smart factory line work and are controllable. The AR contents will be displayed on the Microsoft HoloLens 2 headset.

3.1. System Scenario

The proposed system provides an AR-based application that displays the augmented UI panels, AR user interface, on the Microsoft HoloLens 2 headset. Users can view these panels above the demo site, and they are divided into two parts: the left panel and the right panel.

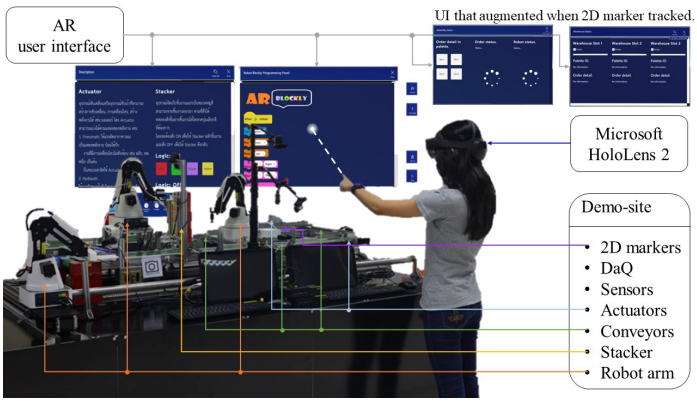


Figure 1: System Scenario

The left panel is the knowledge panel describing the device description, which includes four devices' contents explaining about the sensors used in the demo site and other popular sensors such as proximity sensors. Secondly, the actuator content provides details about various types of actuators used in the demo site. The third content is a content about conveyor which displays the type of conveyors used in the demonstration site. The final content is the robot arm content, which describes different types of robots and how they work with the demo site. User can switch the contents by clicking the button of each section and control the actuator, the conveyor, and the robot arm by interacting with the

buttons that display on each page. The real-time status of each sensor is displayed in the sensor content.

The right panel is for the block-based programming area. Each block is prepared for a specific command, such as controlling robots and conveyors, or if-else logic. To generate a command block, the user needs to click on the desired block displayed on the left side of the panel. The generated block will then appear on the right corner. User needs to drag each command block and snap them together to create an action command. The play button displayed next to this panel is for sending command to the demo site. After the commands are sent, the demo site will execute them step-by-step.

The module's status panel appeared after the 2D marker is tracked. The status includes the module's operational status. For the assembly module, the status panel provides information about operating states such as waiting, working, and so on. The status on the warehouse's panel consists of slot availability. The status of each module is displayed to users in real time.

3.2. System Diagram

Based on Figure 2: System Diagram, the status messages (msg) of each device at the physical level can be transferred to the local broker using the Message Queuing Telemetry Transport (MQTT) protocol. The Node-RED API receives the status and transfers it to the application. When the 2D marker is tracked, the user interface (UI) displays an augmented status panel over the 2D marker.

Once the user finishes programming, the command (cmd) is published from the application. The Node-RED API decodes the command at the API level and distributes it to the devices accordingly. The control of the device is based on the topic that was set earlier.

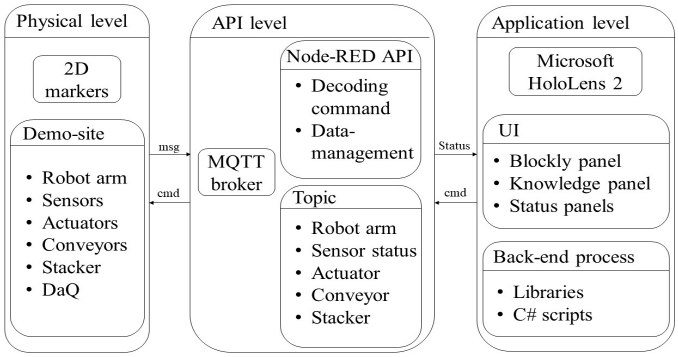


Figure 2: System Diagram

The Physical Level

The physical level consists of five components: MQTT broker, devices, IoT devices, Node-RED execution, and a 2D marker for tracking. The devices and IoT devices implemented in the demo site include robot arms, sensors, actuators, conveyors, stacker, and data acquisition system (DaQ). These physical devices in the real world can be controlled by the user through the application. 2D markers are attached in the demonstration site for providing module information to user when the markers are tracked.

The API Level

There are various libraries used for developing the API level (Application Program Interface), back-end management, and data publication through the broker. The application subscribes to the MQTT topics to receive and send the payload, and a broker publishes a command to that topic for the device to execute at the physical level.

The Application Level

This proposed system was developed using the Unity game engine to create the user interface for the application designed for the Microsoft HoloLens 2. Users can control the demo site by interacting with the user interfaces that appear in the Microsoft HoloLens 2. The user interfaces include the main programming panel (blockly), which provides a block-based programming area, and the knowledge panel, which displays information about the devices.

3.3. Application State Diagram

In Figure 3, the application starts in the initial state and then moves to the connecting state. At this state, the application, the local broker, and the demo site establish a connection. If the connection fails, the Wi-Fi connection in the setting window on the Microsoft HoloLens 2 needs to be checked. Once the connection is successful, the user can interact with the user interfaces displayed on the device to create commands that control the demo site. The application includes a decoding API that converts long commands into step-by-step tasks for the demo site to execute. Finally, the demo site completes the tasks according to the received command.

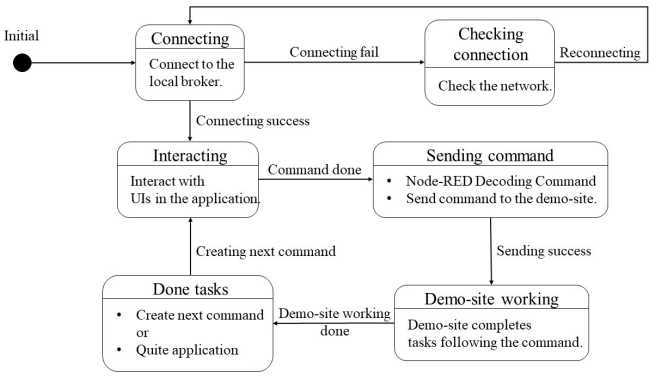


Figure 3: Application State Diagram

3.4. Data Flow

Data flow diagram of the proposed system is shown in Figure 4. There are two parts in the data flow. The first part starts with number 1 and is about the user interface and tracking system. The second part starts with number 2, and includes the MQTT connection and data from the demo site.

Starting at the time before using this system, the application and the demo site need to be connected to the local broker (process 2.0). The connection status shown in the graphic that is augmented on the Microsoft HoloLens 2 display.

The user interface shown on the device is managed by the UI control unit. This unit manages the position of the panels, the position of block command in blockly panel. The snapping function is a part of this unit and is the function for managing the blockly interface. The augmented status panel of the 2D markers developed by using the Vuforia Engine library to define the

markers. The UI control unit provides the position of these augmented graphics.

After user publishes their command, the command management system, which includes Node-RED decoding, receives the command and puts it in order. The first command is sent to the demo site and the next command is sent to the demo site after the first command is finished. While the demo site is working, the status data will be updated in real time and shown in the user interface.

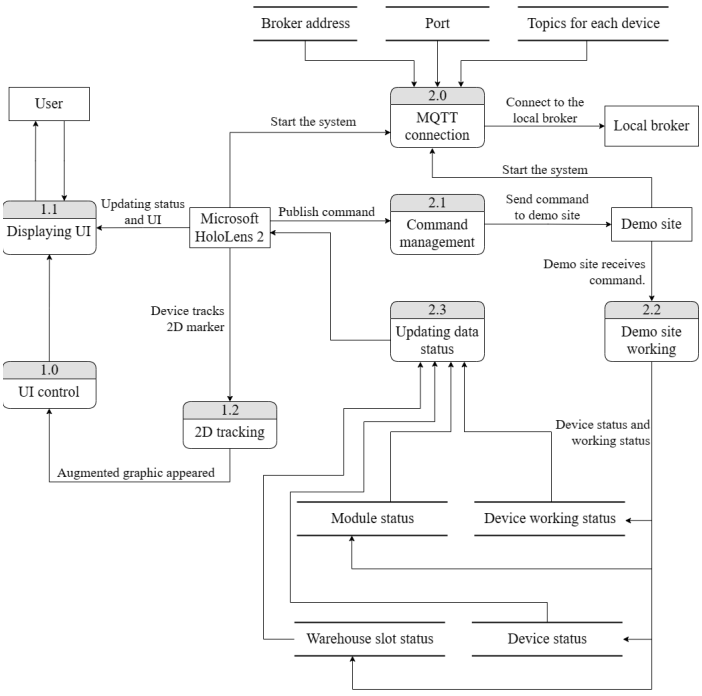


Figure 4: Data Flow Diagram

3.5. Data Synchronization

The device's status data is published to the MQTT local broker and updated in real time. The information such as sensor status (detected or no detected) or module's working status (working, waiting, done) is exchanged under the specific topic. The command order is sent after the previous order is completed. Therefore, the device's working status data and module's status data must be synchronized between the demo site, API, and the AR application.

3.6. Platforms (Hardware and Software)

Microsoft HoloLens 2

Figure 5 shows the Microsoft HoloLens 2, a head-mounted display for augmented reality applications. The see-through holographic display has a 3:2 aspect ratio and a 2K resolution. It is powered by a Qualcomm Snapdragon 850 computing platform with 4 GB of RAM and 64 GB of storage. It supports WIFI and Bluetooth 5 connections. A single battery charge can last for two to three hours. The device has various sensors including a gyroscope for head tracking, two infrared cameras for eye autofocusing, a 1-megapixel time-of-light depth sensor, light cameras, an accelerometer, a magnetometer, and an 8-megapixel camera for capturing 1080p30 video. It also has built-in spatial sound speakers [20].

Figure 5: Microsoft HoloLens 2 [21]

Hand gestures can be used to control the Microsoft HoloLens 2. The device can track the index finger and extend a pointer from it towards the augmented object displayed on the screen. To point, the user lifts the left or right hand up and open the palm as shown in Figure 6 (left) and Figure 7 (left). The dash line graphic will appear from the index finger to the target augmented object. The index finger can then be attached to the thumb to drag the graphic to a desired location, or released to click or choose, as shown in Figure 6 (right). Another method for clicking is to push the index finger through the interactive button. The device can detect which graphic is selected, as shown in Figure 7 (right).

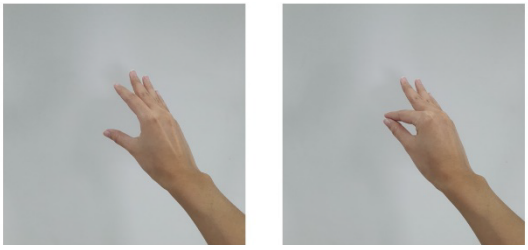


Figure 6: Pointing gesture (left), dragging gesture (right)

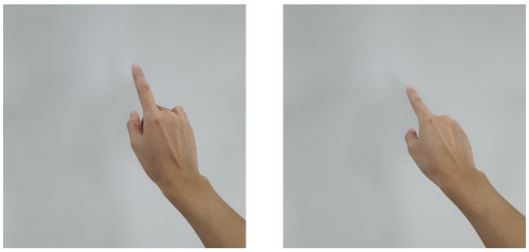


Figure 7: Pointing gesture (left), clicking gesture (right)

Microsoft has recommended size for the holographic buttons depends on the distance between user and the target holographic. The shorter distance (45 cm.) uses direct hand interaction which user can use the fingertip to interact the holographic directly. However, this proposed system chooses the hand-ray interaction instead since the ray that augmented on the fingertip is easy to point to the target holographic for the beginner. The target size should be 3.5 x 3.5 cm. for hand ray interaction. The designed distance is around 0.8-1 m. Then, the minimum designed target size of the button is 3.2 x 3.2 cm. Hence, user can interact with the button comfortably.

2D Tracking

The library for 2D tracking system is Vuforia Engine 10.8 which supports the Unity game engine to track the 2D markers using the computer vision based image recognition [22]. This library was developed for AR application [23]. Developers can upload their 2D markers to the Vuforia's database, with a minimum width of 320 pixels and a maximum size of 2.25 MB for

each file [24]. The Vuforia Engine library supports many devices, including the Microsoft HoloLens 2. Once the marker is tracked, the augmented object will appear either above the marker or in a specific position.

Demonstration Site

The demo site prepared for this proposed system is a collaborative project with the educational smart factory platform, which is a research project of the human-computer interface (HCI) laboratory of the Institute of Field Robotics (FIBO), King Mongkut's University of Technology Thonburi (KMUTT), Thailand.

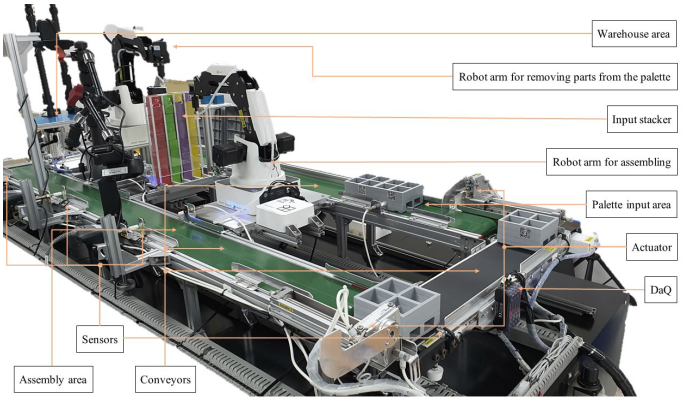


Figure 8: Devices in the Demonstration Site

This demo site simulates the assembly line and operates in separated modules, as shown in Figure 8. The proposed system focuses on the assembly module, the DaQ module consisting of sensors and actuators, and the conveyor module. The hardware includes robot arms, conveyors, sensors, actuators, and input stacker. The communication between the modules in the demo site utilizes the MQTT protocol.

3.7. AR-based Training System

The main contribution of this proposed system is to develop an AR-based application for smart factory concept training system. The system consists of two parts: an AR application including the API, and the contents of smart factory concept.

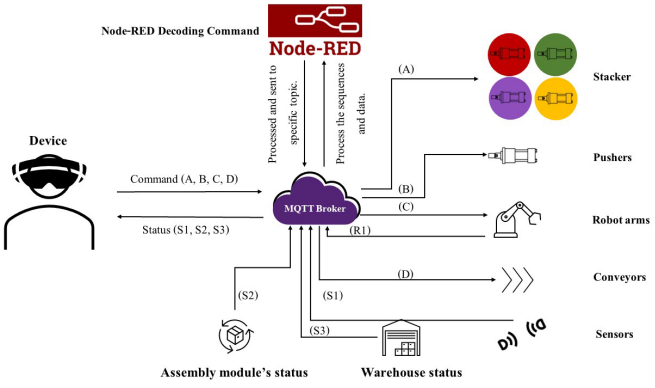


Figure 9: MQTT Data Flow Diagram

Communication between the AR application and devices.

The MQTT protocol is utilized for communication between the AR application and the hardware of the demo site. MQTT broker is configured as the local broker. The scenario of the communication is shown in Figure 9. Each device has its own topics for receiving commands (S1, S2, S3) and updating their operational status (A, B, C, D).

The node-red decoding commands manage the received commands and send them in the correct order to devices, such as stackers, pushers, robot arms, conveyors, and sensors. All statuses from the ended devices will be published to the broker and displayed on the augmented status panels in the Microsoft HoloLens 2.

Node-RED Decoding Command

Node-RED is an online service block-based programming tool [25]. It can create an API for devices and online service using MQTT protocol.

The main API for managing and sending commands to devices in the correct order is the node-red decoding command. The flow chart diagram is shown in Figure 10. The purple block stands for the communication using MQTT protocol.

Starting with the initial state, the sensor and actuator in the palette module will be started. The command (cmd) from the application is sent to MQTT broker in form of string. After the node-red decoding command gets this payload, it will decode command string to command list. The first command set on the index 0 (cmd[0]) of the list will be sent to the device.

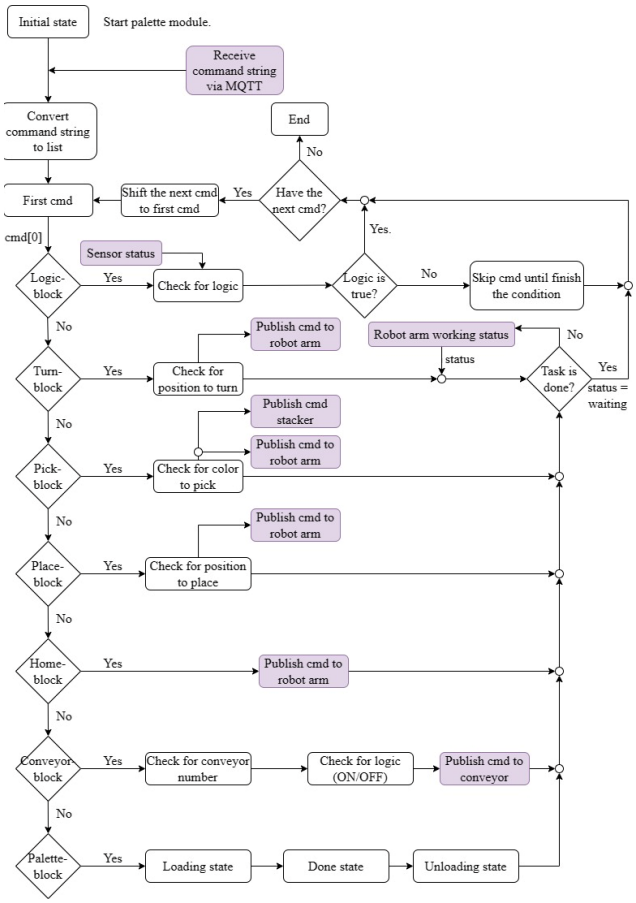


Figure 10: Node-red Decoding Command Flow Chart

The commands such as turning, picking, placing, and homing are related to the robot arm's operations. The working status of the

robot needs to be recognized in real time to prevent the mis-order tasks. The conveyor block controls whether the specific conveyor is on or off.

The logic block will check the sensor status and the logic that the user has previously identified such as the status of object detection. If the condition is true, the demo site will proceed to the next task under this condition scope. If the logic is false, the Node-RED decoding command will skip the tasks under this condition scope and proceed to the next tasks outside of this condition scope.

The user interface in this application consists of two panels as shown in Figure 11. The left panel displays the knowledge contents, while the right panel is the block-based programming area. Both panels are overlaid on the demo site, allowing the user to view them and the environment simultaneously.

Figure 11: User interface Displayed over the Demo Site (User's View).

Contents

Three video animations and four device contents are provided for the content about the smart factory concept. The videos cover key areas such as the relationship and differences between industry 4.0 and smart factories, the benefits of upgrading traditional manufacturing to a smart one, the cyber-physical system with the Internet of Things (IoT), and the Industrial Internet of Things (IIoT). They also introduce the basic phase of the programming structure for controlling the robot arm and provide the basic instructions for the robot arm's three kinds of movements, which are linear, joint, and circular. Additionally, the video animation also presents the professional standards for robotics and automation careers in Thailand, as defined by the Thailand Professional Qualification Institute (TPQI).

The AR application features four device contents that introduces the devices implemented in the demo site. Furthermore, the information about each device helps the users understand the other categories of devices that are frequently used in the smart manufacturing. The device contents are the sensor content, the actuator content, the conveyor content, and the robot arm content.

The real time sensor status is displayed in the sensor content panel. The actuator content panel provides examples of actuator types and easy control buttons that users can use hand gestures to interact with them. In general, the actuators are controlled by electricity or the pressure, such as pneumatic actuators. The buttons can control the input stacker and pusher actuators to push and pull the objects. The elements in the conveyor content panel and the robot arm content panel are similar to those in the actuator content panel. There are the knowledge contents displaying text and the buttons that can control the devices.

Examination and Feedback Forms

This application was evaluated using the pre-test, post-test, system satisfaction feedback forms, and content satisfaction feedback form.

The closed-ended questions in system satisfaction feedback form and the content satisfaction feedback form use the Likert's scale [26] to measure user's attitudes. The scale starts from 1 to 5 points where the minimum point stands for strongly disagree / very poor / definitely not, and the maximum point stands for strongly agree / excellent / definitely.

The ideas to develop the factory are collected from the open-ended questions. These questions include how to implement the knowledge gained from the training system in the user's factory, ideas of implementing the AR technology to their factory, and the suggestions for improving the system.

4. Experimental Results

4.1. Population and Requirement

The proposed system is developed for workers who need to reskill or upskill their performance about robotics and automation in smart factory. The requirements for volunteers include: 1. Working in the automotive or electrical industry. 2. Being a key person in developing their factory into a smart factory. 3. Being an engineer, manager, or staff who is interested in robotics in smart factories. There are 22 volunteers who evaluated the proposed system.

4.2. Tools

Pre-test and Post-test

The purpose of the pre-test is to assess the user's knowledge background. Questions in pre-test are about contents in the smart factory and fundamental of each device used in the demo site.

The post-test's questions are the same as the ones in the pre-test. User takes the post-test after watching all video contents and the AR-based training system. The expectation of these tests is that user should get more points after attending the training system. Points gained after using the AR-based training system demonstrate that user has more knowledge about the smart factory concept than they did previously.

System Satisfaction Feedback Form

This form includes general questions about user's background, usability of this training system, and suggestion for improving the system. The usability questions cover the assessment of values for specific task of the proposed system. The writing parts in this form is used to collect user's ideas about the factory improvement based on experience after attending this training system.

Content Satisfaction Feedback Form

This form is used to collect the feedback about the contents. The content satisfaction also includes questions for evaluating the values for specific task.

4.3. Tasks

Task 1: User Assessment

There are 2 examinations for user assessment. User needed to follow these steps.

1. Taking the pre-test to assess the user's knowledge background.
2. After users have completed all the contents and used the application, they must take the post-test to evaluate their knowledge progress.

Task 2: Learning

There are two types of smart factory contents that users need to learn. The first type is three videos about the smart factory concept. The second type is contents about the devices that are mostly used in smart manufacturing. The device contents are provided in the AR application.

Task 3: Using the Application

1. This task is about how to use the application in the Microsoft HoloLens 2 by watching a tutorial video of the application. The video provided the suggested hand gestures to interact with the Microsoft HoloLens 2, the block's name of the Blockly, and how to use the Blockly by dragging and snapping the blocks together.
2. User started by learning about the device contents (sensor, actuator, conveyor, robot arm) that appear in the application. They practiced their hand gestures for interacting by pushing interactive buttons in the knowledge panel.
3. User created commands to control the demo site using Blockly panel and completed the given task (Algorithm 1).
4. After user completed using the application, user needed to take the system satisfaction feedback form and content satisfaction feedback form.

During operation, the user's view in the Microsoft HoloLens 2 was live streamed to the developer in real time. If user had the problems with the system or the interaction, developer would suggest how to solve them.

User took an hour to hour and half per one test following the steps 1-9. The time they took is up to how fast they can adjust their hand gestures to interact with the user interface, and how to solve the given programming task by themselves.

The risk of using this training system may cause the motion sickness for some people. To relieve the symptom, user needs to take the Microsoft HoloLens 2 off, sit on the chair and get eyes rest for 10-15 minutes. No one in this study had a motion sickness symptom while using this training system.

Algorithm 1: Making the Condition.

```
Start;
Instruction 1;
 if condition then
   | Instruction 2;
   | Instruction 3;
 else
   | Instruction 4;
Instruction 5;
 end
```

4.4. Results

The frame rate of the proposed system is in the range of 48-60 fps, which can display the augmented graphics clearly and user can interact with those graphics smoothly.

The formula (1) is to calculate the learning rate of all users measured by the percent change of post-test scores and pre-test scores. While "Post" stands for post-test score, and "Pre" stands for pre-test score.

$$\%Change = \frac{(Post-Pre)}{Pre} \times 100 \quad (1)$$

Table 3: User's Changed Learning Rate

Group	Experience	% changes (%)	Amount of users
1	Had experience about smart factory and had to implement it in their factory before.	23.42	9
2	Had knowledge about the smart factory concept but did not implement it before.	36.37	4
3	Had no knowledge and experience about the smart factory concept before.	75.25	9

The Table 3 showed that the % changes of all users were increased as 23.42% for 9 users, 36.37% for 4 users, and 75.25% for 9 users. The result showed that this proposed system can deliver the knowledge of the smart factory concept to users. The differences of the % changes are because user had some experience about smart factory before using this proposed system. Users in group 1 had less % changes than the other groups but it did not mean that they obtained few knowledge. The users in group 1 have more ideas of implementing new technology such as AR in their manufacturing processes because they can apply their priori knowledge and experience with this proposed AR-based training system.

The evaluations of the usability and values for specific task are based on 5-point Likert's scale as shown in Table 4. The usability score of the proposed system obtained 4.44 from 5 points, while score of the values for specific task of the system is about 4.23 from 5 points, and the values for specific task related to the contents obtained 4.30 from 5 points. Moreover, all users had more ideas of implementation new technology such as AR and the smart factory concept to their works.

Table 4: 5-point Likert's Scale.

Point	Meaning
1	Strongly disagree
2	Disagree
3	Neither agree nor disagree
4	Agree
5	Strongly agree

5. Conclusions and Discussions

Thailand has the direction to increase the manufacturing performance using the implementation of the smart factory. In the smart industry, skills related to robotics and automation are required. Worker skill is the most important factor in the factory

development. The AR-based training system for industrial training with the industry 4.0 contents aimed to develop worker's knowledge and skills. This proposed system covered the foundation contents about the smart factory and devices used in the smart production line. Moreover, the block-based programming panel is provided in the AR application to support users to practice their programming skills. The Microsoft HoloLens 2 device was selected to work with this system. The graphic contents are displayed on the Microsoft HoloLens 2 screen over the real environment. Worker upskilling can be evaluated with the scores of the pre-test and post-test. The points gained after completing the training from the proposed system indicate that the user's knowledge of smart factories and robotics and automation has increased.

The smart factory contents in this proposed system enhanced the user's knowledges. The proposed system can display with 48-60 fps. along with 1440 x 936 pixels per eye. Furthermore, the evaluation scores of the values for specific task and usability were in the range of 4-5 points, indicating that users were satisfied with this proposed system.

The experimental results discovered that the user's learning rate also depends on their priori knowledge and experiences. However, this proposed AR-based training system can delivery knowledge and experience about the smart factory concept and robot visual programming through augmented reality technology. This helps users to be able to utilize the obtained experiential learning for improving their manufacturing process.

Currently, the contents in this proposed system follows the standards in robotics and automation in TPQI standards. In the future, the blockly panel could be improved to increase the UI performance in order to solve the display delay problem. The various contents about the smart factory and digital disruption can be added to the system. This proposed system can be developed for various contents for training in some specific production lines. Moreover, the display of the device's status in the proposed system can be implemented for the data visualization in the smart production line.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We would like to acknowledge the funding from PMU-B's AI for All 1-2 funds, the KMUTT's fundamental fund for financial supports, and the Human-Computer Interface Lab at the Institute of Field Robotics for providing the infrastructure and hardware implemented in this research.

References

[1] The National Electronics and Computer Technology Center (NECTEC), 2019, Cyber-Physical System, https://www.nectec.or.th/research/research-project/nectec-cps.html.

[2] Z. Suleiman, S. Shaikholla, D. Dikhanbayeva, E. Shehab, A. Turkyilmaz, "Industry 4.0: Clustering of concepts and characteristics," Cogent Engineering, **9**(1), 2022, doi:10.1080/23311916.2022.2034264.

[3] Germany Industrie 4.0, https://ati.ec.europa.eu/sites/default/files/2020-06/DTM_Industrie%204.0_DE.pdf.

[4] NIST, Smart Manufacturing | NIST, https://bit.ly/2IJXL4o.

[5] The Excise Department, Thailand 4.0, https://www.excise.go.th/cs/groups/public/documents/document/dwnt/mjgy/~edisp/uatucm282681.pdf.

[6] Business Opportunities, https://www.eeco.or.th/en/business-opportunities.

[7] The Federation of Thai Industries, 2020, Thailand Competetiveness 2016: Tha IMD Perspective, 6, http://www.nfcrbr.or.th/site/attachments/article/81/White%20paper.pdf .

[8] D. Mourtzis, V. Siatras, V. Zogopoulos, "Augmented reality visualization of production scheduling and monitoring," Procedia CIRP, **88**, 151–156, 2020, doi:10.1016/j.procir.2020.05.027.

[9] M. Speicher, B.D. Hall, M. Nebeling, "What is Mixed Reality?," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA: 1–15, 2019, doi:10.1145/3290605.3300767.

[10] Paul Milgram, Fumio Kishino, "A Taxonomy of Mixed Reality Virtual Displays", in IEICE Transaction on Information and Systems, Figure 1, 3.

[11] Reid Manning, 2022, Augmented, Virtual, and Mixed Reality Training Landscape [Online], Available: https://www.spheregen.com/augmented-virtual-and-mixed-reality-employee-training.

[12] S. Ke, F. Xiang, Z. Zhang, Y. Zuo, "A enhanced interaction framework based on VR, AR and MR in digital twin," Procedia CIRP, **83**, 753–758, 2019, doi:10.1016/j.procir.2019.04.103.

[13] F.-K. Chiang, X. Shang, L. Qiao, "Augmented reality in vocational training: A systematic review of research and applications," Computers in Human Behavior, **129**, 107125, 2022, doi:10.1016/j.chb.2021.107125.

[14] N. Gavish, T. Gutiérrez, S. Webel, J. Rodríguez, M. Peveri, U. Bockholt, F. Tecchia, "Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks," Interactive Learning Environments, **23**(6), 778–798, 2015, doi:10.1080/10494820.2013.815221.

[15] F. Torres-Guerrero, L. Neira-Tovar, L. Torres-Treviño, An Introductive Training for Welding Workshop: A Biometric Evaluation Using Virtual Reality Scenes to Improve Practice, 319–331, 2019, doi:10.1007/978-3-030-18715-6_27.

[16] Z. Zhu, C. Liu, X. Xu, "Visualisation of the Digital Twin data in manufacturing by using Augmented Reality," Procedia CIRP, 81, 898–903, 2019, doi:10.1016/j.procir.2019.03.223.

[17] Educational Management, 2559, Thailand 4.0 Blueprint, https://waa.inter.nstda.or.th/stks/pub/2017/20171114-draeqa-blueprint.pdf.

[18] Office of The Official Information Commission, Thailand 4.0, http://www.oic.go.th/FILEWEB/CABINFOCENTER3/DRAWER049/GENERAL/DATA0000/00000702.PDF.

[19] Thailand Professional Qualification Database System, Robotics and Automation, https://tpqi-net.tpqi.go.th/home/occ/industrialInfo/RAC.

[20] Microsoft, HoloLens 2, https://www.microsoft.com/en-us/HoloLens/hardware#OneGDCWeb-Banner-ffqoo6k.

[21] Mikko Ruotsalainen, Lari Komulainen, Alex Tejada, "Virtual Lenses to Support Study Work at Customer Site", School of Engineering Science, 2021.

[22] Vuforia, Getting Started with Vuforia Engine in Unity, https://library.vuforia.com/getting-started/getting-started-vuforia-engine-unity.

[23] Vuforia, Image Targets, https://library.vuforia.com/objects/image-targets.

[24] D. Amin, S. Govilkar, "Comparative Study of Augmented Reality Sdk's," International Journal on Computational Science & Applications, **5**(1), 11–26, 2015, doi:10.5121/ijcsa.2015.5102.

[25] Node-RED, https://nodered.org.

[26] I. Elaine Allen and Christopher A. Seaman. (2022). Likert Scales and Data Analyses, https://www.bayviewanalytics.com/reports/asq/likert-scales-and-data-analyses.pdf.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 9, No. 2, 09-16 (2024)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Double-Enhanced Convolutional Neural Network for Multi-Stage Classification of Alzheimer's Disease

Pui Ching Wong*, Shahrum Shah Abdullah, Mohd Ibrahim Shapiai

Department of Electronic System Engineering, Malaysia–Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM), 54100, Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received: 13 January, 2024
Revised: 20 February, 2024
Accepted: 21 February, 2024
Online:23 March, 2024

Keywords:
Double-enhanced CNN
Multi-stage classification
Attention module
Generative adversarial network

ABSTRACT

Being known as an irreversible neurodegenerative disease which has no cure to date, detection and classification of Alzheimer's disease (AD) in its early stages is significant so that the deterioration process can be slowed down. Generally, AD can be classified into three major stages, ranging from the "normal control" stage with no symptoms shown, the "mild cognitive impairment (MCI)" stage with minor symptoms, and the AD stage which depicts major and serious symptoms. Due to its generative features, MCI patients tend to easily progress to the AD stage if appropriate diagnosis and prevention measures are not taken. However, it is difficult to accurately identify and diagnose the MCI stage due to its mild and insignificant symptoms that often lead to misdiagnosis. In other words, the classification of multiple stages of AD has been a challenge for medical professionals. Thus, deep learning models like convolutional neural networks (CNN) have been popularly utilized to overcome this challenge. Nevertheless, they are still limited by the issue of limited medical images and their weak feature representation ability. In this study, a double-enhanced CNN model is proposed by incorporating an attention module and a generative adversarial network (GAN) to classify magnetic resonance imaging (MRI) brain images into 3 classes of AD. MRI images are obtained from the Open Access Series of Imaging Studies (OASIS) database and four experiments are done in this study to observe the classification performance of the enhanced model. From the results obtained, it can be observed that the enhanced CNN model with GAN and attention module has achieved the best performance of 99% as compared to the other models. Hence, this study has shown that the double-enhanced CNN model has effectively boosted the performance of the deep learning model and overcame the challenge in the multi-stage classification of AD.

1. Introduction

This paper is an extension work originally presented at the 2023 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET) [1]. It has been extended and improved in terms of its model for the multi-stage classification of Alzheimer's disease (AD). Following the trend of improved performance obtained from single-enhanced convolutional neural network (CNN) models, this study proposes a double-enhanced CNN model by incorporating an attention module together with a generative adversarial network (GAN) for further improvement over existing methods.

As AD has been known as a progressive brain disorder that is still incurable to date, its early detection and diagnosis have been identified as a significant step by medical professionals [2]. Appropriate treatments can help to alleviate the deterioration process of AD even though it is impossible to fully cure an Alzheimer's patient. World Health Organization (WHO) [3] has reported that AD is the seventh leading cause of death globally and there are currently more than 55 million affected individuals worldwide. Thus, numerous research has been done in this field throughout the years to alleviate the condition [4]. Accurate classification and diagnosis of the disease are important as they can help doctors to effectively provide suitable treatments for slowing down the deterioration process so that patients will be able to continue their daily life as usual [2]. However, it is not easy to classify and diagnose the stages of AD accurately due to the complexity of medical images and the ambiguity among the stages [5].

Generally, there are three major stages of Alzheimer's, which range from the normal control (NC) stage with no symptoms, mild

*Corresponding Author: Pui Ching Wong, puiching1997@graduate.utm.my

cognitive impairment (MCI) with slight or mild symptoms, and the AD stage with serious and major symptoms. The challenge in classifying between these multiple stages is the difficulty in identifying the MCI stage which consists of faint symptoms that are hardly noticed [6]. However, there is a need to accurately identify and classify this prodromal stage of AD as the symptoms might worsen over time and progress to the next serious stage of AD [7]. Existing studies have shown that multi-stage classification is more difficult to detect than binary classification which only comprises two distinct stages, NC and AD. Thus, deep learning classification models have been utilized to efficiently aid medical practitioners in the classification and diagnosis of medical images [8, 9].

In [10], the authors used a type of CNN model known as the Siamese CNN to classify multiple dementia stages. Similarly, authors in [11] and [12] also applied different types of CNN models, such as VGG-16, AlexNet, ResNet-18, and GoogleNet for the classification of MRI scans into their respective stages. Aside from the CNN model which is well-known for its great ability to learn from raw images and classify them accurately [13], other types of deep learning models have been utilized as well. For instance, in [14] and [15], the authors applied recurrent neural networks (RNN) for the classification task due to the model's ability to solve the issue of incomplete feature extraction. On the other hand, authors in [16] used a long short-term memory (LSTM) model for multi-stage classification, while in [17], the authors utilized a GAN for image enhancement and classification tasks. All these studies that utilized various types of deep learning networks have proven that these models can effectively classify and diagnose AD stages accurately.

Nevertheless, deep learning models generally require large amounts of data for training and learning image features [18]. Hence, the complexity and limited availability of medical images have appeared to be a challenge for AD classification recently. Magnetic resonance imaging (MRI) brain images of AD patients, which are commonly used to visualize the inside of the human brain and observe any abnormalities, have a complex structure that requires much experience, time, and professionalism for analysis [5]. Besides, they are also available in limited amounts due to the privacy protection of patients' data, leading to insufficient data for research purposes. Thus, classification models have been enhanced with data augmentation models recently to overcome this limitation. For instance, CNN models are being enhanced by incorporating a GAN model for dataset expansion before the classification task. Authors in [17], [19], and [20] have shown examples of state-of-the-art studies that applied single-enhanced CNN models using GAN for their research to overcome this challenge.

However, another issue in CNN models has arisen, which affects the accuracy of AD classification tasks. The lack of adaptive channel weighting in CNN models has been identified recently and this leads to the inability of the model to effectively learn significant features from MRI brain images to accurately classify AD [21]. Thus, to solve this issue, the CNN model needs to be enhanced using an attention module to improve its feature representation ability. The attention module is a type of enhancement module that can identify regions of interest (ROIs) in an image and apply attention to them [22]. Using the information provided by the attention module, the CNN model will be able to focus on learning the features significant to each stage of Alzheimer's, and then accurately classify them. In [23] and [24], the authors presented studies that applied a single enhanced CNN model using an attention module in their works and achieved improved results. Thus, this study aims to propose a double-enhanced CNN classification model to overcome the mentioned challenges.

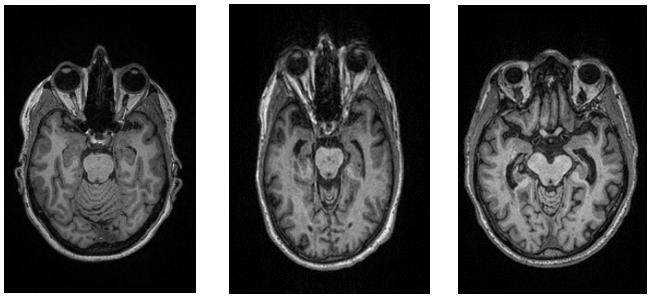
2. Methodology

2.1. Dataset

In this study, MRI data required for training the deep learning model to perform AD classification are obtained from the Open Access Series of Imaging Studies (OASIS) database (https://www.oasis-brains.org/)[25]. The OASIS database is a publicly open-access database on medical imaging which is commonly used for research purposes. Since a multi-class classification task of Alzheimer's disease (AD) is carried out in this study, MRI data from three different classes are collected for experiment purposes. For each class, namely the normal control (NC), mild cognitive impairment (MCI), and Alzheimer's disease (AD) classes respectively, 100 subjects each are collected to maintain a balanced dataset. Table 1 shows the demographic and clinical characteristics of the MRI data collected from the OASIS database. As seen from the table, the subjects are chosen based on a clinical dementia rating (CDR) score, where certain ranges of CDR represent different classes of AD. Moreover, all the MRI images collected are standardized by using only T1-weighted MRI sequences scanned with a 3 Tesla scanner. Figure 1 shows examples of MRI data collected from the OASIS database.

Table 1: Clinical and Demographic Characteristics of MRI Data

Classes	**AD**	**MCI**	**NC**
Number of subjects	100	100	100
CDR score	CDR > 2	0.5 < CDR < 1	CDR = 0
Gender	Males and females		
Age	65 years old < Age < 74 years old		
MRI sequence	T1-weighted		
Scanner type	3 Tesla scanner		



(a) NC class (b) MCI class (c) AD class

Figure 1: Example of MRI data collected from the OASIS database

2.2. Proposed Methodology

The proposed methodology for this study is an enhanced CNN model with an attention module and GAN model. It is proposed to enhance and improve the performance of the CNN model in the multi-class classification task of diagnosing AD. Figure 2 shows the overall flowchart of the proposed methodology. Firstly, the GAN model will be trained using the MRI dataset collected to expand the dataset. Then, the expanded dataset will be inputted to the CNN model with an attention module incorporated for training and validation. The attention module will be responsible for identifying the significant features and regions of interest (ROIs) that correspond to each stage of AD. With this information, the CNN model will undergo training to learn and classify MRI images into the respective 3 stages of AD. After training, the trained CNN model will be validated with an unseen set of MRI data to validate its performance and generalizability. Since there are generally 2 phases in this proposed methodology, the MRI dataset collected is distributed into 2 sets, namely training and validation, as shown in Table 2. From the 300 subjects collected from the OASIS database, a total of 3000 MRI slice images are included as the dataset for this study after filtering and choosing only the slices with significant information. Then, the dataset is split into two subsets of training and validation with a ratio of 8:2.

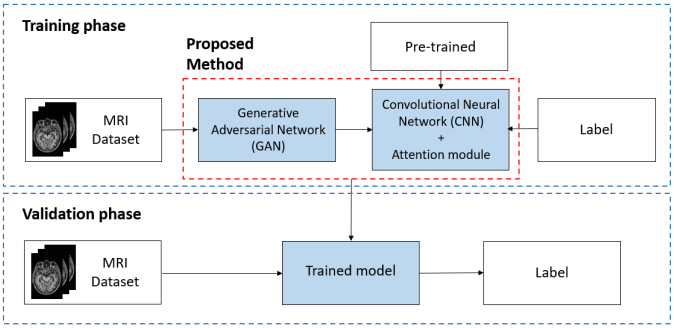


Figure 2: Overall flowchart of proposed method

Table 2: Distribution of MRI Data for Training and Validation

Classes	AD	MCI	NC	Total
Training	800	800	800	2400
Validation	200	200	200	600

2.3. Convolutional Neural Network (CNN)

In this study, a pre-trained convolutional neural network (CNN) is used to learn and classify MRI images into 3 classes, namely NC, MCI, and AD. CNN is a type of deep neural network popularly used for image analysis and classification tasks due to its powerful ability in feature extraction and learning discriminative representations from raw data [26]. Its architecture is illustrated and shown in Figure 3.

However, current trends show that transfer learning is often being used together with CNN models recently for better training experience and performance [27]. Transfer learning can aid CNN models by providing better start value and asymptote, as there is no need to train a model from scratch. This is because the CNN models have been previously trained using ImageNet, a large dataset with millions of images [28], and thus, pre-trained models have been equipped with the optimum parameters for the best training results. With just some fine-tuning, the CNN model can be easily reused for other tasks, such as for AD diagnosis in this case.

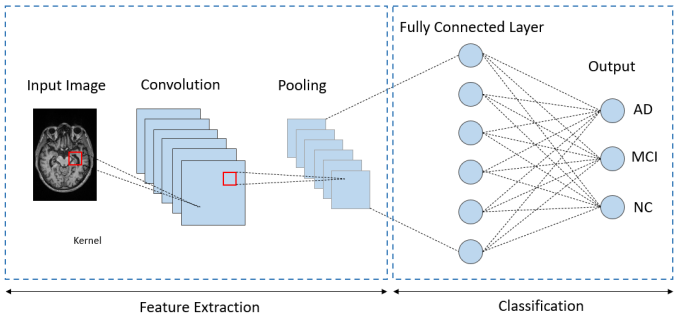


Figure 3: CNN architecture [26]

Since there are numerous types of pre-trained CNN models available, this study has chosen to utilize a ResNet-18 model for the multi-stage classification task of AD. Due to its network architecture that uses skip connections and residual blocks, it can support many convolutional layers, enable smoother gradient flow, and solve the issue of vanishing gradients [29]. Figure 4 shows its network architecture.

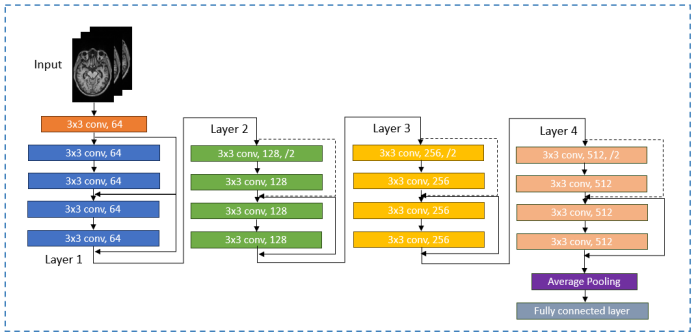


Figure 4: ResNet-18 network architecture [29]

2.4. Attention Module

Adapted from [1], the attention module is a popularly used enhancement module recently in deep learning applications due to its ability to aid in a network's feature representation power. It has been successfully applied and used in various fields such as computer vision and natural language processing (NLP) [30]. The main characteristic of the attention module is that it can help to strengthen the feature extraction ability of a CNN model, and thus enhance the model's analysis and classification performance. Since it is a computer-aided mechanism inspired by the human beings' visualization system [31], the attention module can easily recognize which regions are significant and carry important information in an image, just like the ability of the human eye. Then, it will help the deep learning model to concentrate on learning the important information only by applying focus on these regions of interest (ROIs), such as assigning a higher degree of importance to them. Thus, during training, the CNN model will be

able to learn better as they can focus on the important regions only instead of learning from the whole image which contains abundant information.

In this study, a type of attention module known as self-attention is incorporated into the CNN model for the multi-stage classification of AD. Among the various types of attention modules available, self-attention is chosen due to its characteristics that can learn global or long-range dependencies. It can help in addressing the problem of computational inefficiency brought on by CNN's limited ability that only perform local operations. Thus, it is incorporated into the CNN model to enhance its feature-capturing ability [24]. Figure 5 shows the structure of the self-attention module.

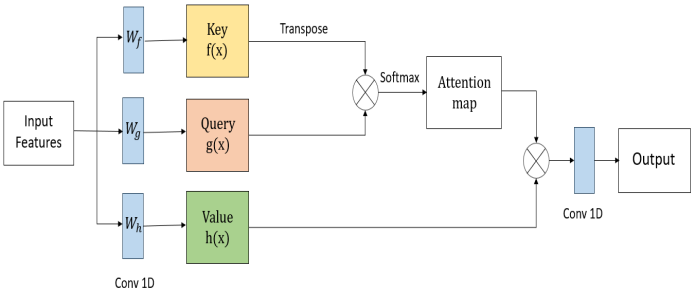


Figure 5: Self-attention module [24]

To help the CNN model capture global information efficiently, the self-attention module maps a key, query, and value to the input. The key and the value are the extracted features from the input MRI images, while the query determines the significant features to be learnt by CNN. As shown in Figure 5, the key, query, and value are transformed into vectors by using a 1x1x1 convolution filter labelled with Wf, Wg, and Wh respectively. After the transformations, they are represented as shown in (1), (2), and (3):

$$Key = f(x) = W_f x \quad (1)$$

$$Query = g(x) = W_g x \quad (2)$$

$$Value = h(x) = W_h x \quad (3)$$

As shown from (1), (2), and (3), the $x \in R^{C\,x\,N}$ represents the features from the original feature map, where C is the number of channels and N is the number of locations of features respectively. Then, the self-attention map $a_{i,j}$ is calculated as below:

$$a_{i,j} = \frac{\exp\left(f(x_i)^T g(x_j)\right)}{\sum_{i=1}^{N} \exp\left(f(x_i)^T g(x_j)\right)} \quad (4)$$

, where $a_{i,j}$ denotes the degree of attention of the model on the i-th region during the synthesis of the j-th region. Following this, the output of the self-attention module is represented by O_j, where $O = (O_1, O_2, \ldots O_j, O_N) \in R^{C\,x\,N}$, and

$$O_j = W_v \left(\sum_{i=1}^{N} a_{i,j} h(x_i) \right) \quad (5)$$

A 1x1x1 convolutional filter is again used to reduce the number of channels of the final output for standardization and memory efficiency as shown in Figure 5.

2.5. Generative Adversarial Network (GAN)

As shown in the overall methodology flowchart in Figure 2, MRI data will first be inputted into a Generative Adversarial Network (GAN) model before they are used to train the CNN model for multi-class classification of AD. GAN is a type of unsupervised deep neural network popularly used for data augmentation and expansion tasks due to its progressive nature. Introduced in [32], the author had proven that it is able to generate realistic and high-quality samples by applying the knowledge from learning the probability distribution of input variables. Thus, it has been increasingly used in recent studies that faced the problem of insufficient data, such as in this research domain where MRI images of Alzheimer's patients are always in limited quantity [17, 19, 33]. Figure 6 shows the structure of a GAN model.

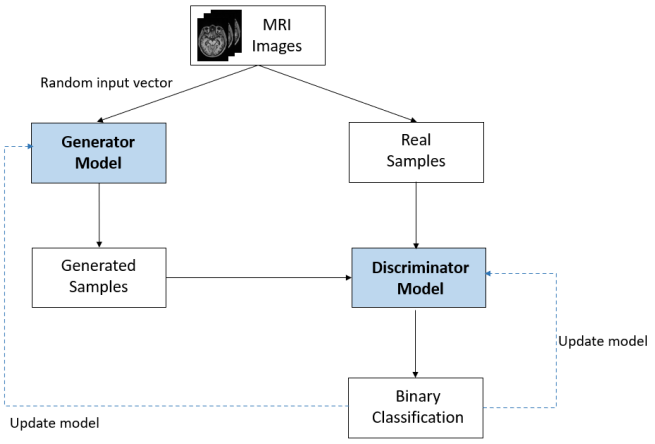


Figure 6: GAN model structure

From Figure 6, it can be seen that there are generally 2 network models in a GAN, which are the generator and discriminator models respectively. The generator model typically generates new samples based on the received random input vector and then passes them to the discriminator model for the following step. On the other hand, the discriminator model receives the real MRI images from the input as well as the generated samples from the generator model. It then compares them and distinguishes the generated samples from the real ones. Both models will continuously update and improve their models to achieve a point of equilibrium where the discriminator can no longer distinguish generated samples from the real data. Equation (6) shows the loss function of the GAN where both the generator and discriminator models will continuously learn until they achieve the goal of maximizing the number of real-looking sample images.

$$\min_G \max_D V(D, G) = \mathbb{E}x{\sim}p_{\text{data}}(x) \log D(x) + \mathbb{E}z{\sim}p_z(z) \log (1 - D(G(z))) \quad (6)$$

2.6. Experiment Setup

In this study, four experiments are carried out to verify and observe the difference in the models' performances when classifying MRI images into multiple stages of AD. Table 3 shows the model used for each experiment and the differences between them. Experiment I is done using only the CNN model without applying any enhancements. Experiment II utilizes an enhanced

CNN with an attention module, while a GAN-enhanced CNN model is utilized for Experiment III. Lastly, the proposed enhanced model in this study, which is the CNN enhanced with attention module and GAN is applied in Experiment IV.

Table 3: Four experiments done in this study

Experiment	CNN model	Enhanced with attention module	Enhanced with GAN model
I	/	X	X
II	/	/	X
III	/	X	/
IV	/	/	/

In Experiments I to IV, each of the models is developed using PyTorch, the Python-based deep learning framework and trained using the graphics processing unit (GPU). The hyperparameters used for the models' training are tabulated in Table 4. To avoid bias, the same amount of MRI data as shown in Table 2 is applied for models' training and validation. However, for Experiments III and IV which involve the enhancement using the GAN model, the amount of data used for the multi-stage classification task is later increased due to the addition of generated MRI data samples by GAN, which will be further explained in the results and discussion section below.

Table 4: Training hyperparameters used in Experiments I to IV

Hyperparameters	GAN model	CNN model
MRI image size	64 x 64 x 3	224 x 224 x 3
Epoch	100	100
Batch size	64	8
Learning rate	0.0002	0.0001
Optimizer	Adam optimizer	Adam optimizer
Loss function	Binary cross entropy	Categorical cross-entropy

2.7. Performance Evaluation

After training the models in Experiments I to IV, their performances are measured and evaluated for comparison purposes. It is required to observe the differences in performances of the various combinations of models used in the four experiments so that the best model can be identified. It is also to validate whether the proposed model can perform as desired to achieve the objective of this study. Thus, a few performance metrics are used to evaluate and measure the performance of the 4 different models in the multi-stage classification task of AD. Equations (7) to (10) show the four types of evaluation metrics used in this study together with their method of calculations. The accuracy, precision, recall, and F1-score of the four models used in Experiments I to IV are calculated and tabulated in the results section for performance evaluation. Based on Equations (7) to (10), it can be seen that the metrics are calculated by using the true and false positives and negatives obtained from the classification tasks.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F1 - score = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \quad (10)$$

3. Result and Discussion

After carrying out four experiments by using four different combinations of models to verify the proposed methodology, all the results are tabulated as shown in Table 5. The table shows the performance of the models in terms of performance metrics such as accuracy, precision, recall, and f1-score. The different models used in the four experiments are listed below:

a) Experiment I: CNN model only
b) Experiment II: Enhanced CNN model with attention module only
c) Experiment III: Enhanced CNN model with GAN only
d) Experiment IV: Enhanced CNN model with attention module and GAN model

For each experiment, the CNN model is trained to classify MRI images into different stages of AD using the same number of data as shown in Table 2 to avoid bias. However, for experiments III and IV which involve the usage of GAN, the MRI dataset is expanded before being inputted into the classification model. Figure 7 shows the results from training the GAN model using MRI images obtained from the OASIS database. As shown in the figure, the grid image on the left illustrates the real MRI images from the MRI dataset, while the grid image on the right shows the fake MRI images that are generated by the GAN model. Thus, from the figure, it can be seen that the fake images look quite similar to the real images with most of the image features accurately generated by the GAN after learning them during model training. For Experiments III and IV, the generated MRI images output by the GAN model are added to the original MRI dataset before the CNN models are trained for multi-stage classification tasks. In both experiments, 160 generated MRI sample images were added to the original dataset, accounting for a total of 960 MRI images used for the CNN classification task.

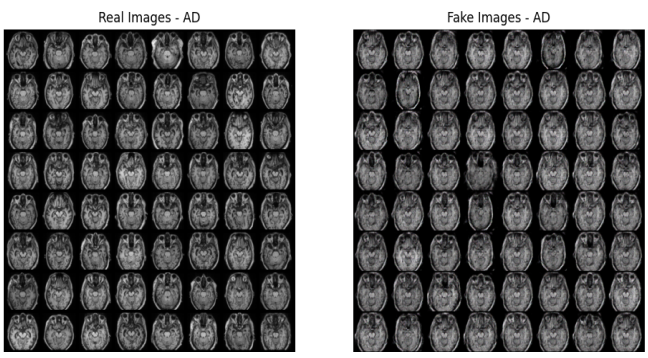


Figure 7: Comparison between real and generated images by GAN

Table 5 shows the complete summary of the results obtained from the multi-stage classification task of AD in terms of accuracy, precision, recall, and f1-score for Experiments I to IV. On the other hand, Table 6 shows a comparison between the performances of the four experiments, focusing on the accuracy metrics. Hence, it can be seen that in general, the results obtained from Experiment I such as the accuracy, precision, recall, and f1-score have the lowest values among the 4 experiments. This is because a basic CNN model without any enhancement is used for classification in Experiment I. Experiments II and III apply a single enhancement to the CNN models, and a slight increase in the performance metrics' results is seen. For instance, the training accuracy has improved from 95% in Experiment I to 98% in Experiment II, while the f1-score has improved from 90% in Experiment I to 96% in Experiment III.

Following this, a more remarkable improvement in the results can be observed from Experiment IV which utilized a double-enhanced CNN model, as proposed in this study. From Table 5, the enhanced CNN model with attention module and GAN model has managed to output the highest accuracy at 99% during training and 92% during validation. These can also be validated from the accuracy and loss graphs that are shown in Figures 8 and 9 respectively. From Figure 8, it can be seen that the accuracy achieved by the proposed model matches the results tabulated in Tables 5 and 6, while in Figure 9, the model loss during training precisely decreases until approximately zero as the number of training epochs increases. However, a slight limitation can be observed from the graphs, which is the presence of some overfitting in the model training, as the validation performance is generally lower than the training results.

Nevertheless, great improvements in the other performance metrics such as precision, recall, and f1-score in Experiment IV can be clearly observed in Table 5 as well. The results suggest that the proposed double-enhanced CNN model achieved the highest performance as compared to the other 3 models in the other experiments. Thus, from the tables and graphs shown, it can be concluded that the classification model's performance significantly improved when a double-enhanced CNN model with both attention module and GAN is utilized, as proposed.

Table 5: Results of performance metrics obtained from Experiments I to IV

	Accuracy	Precision	Recall	F1-Score
Experiment I				
Training	0.9580	0.9017	0.9020	0.9018
Validation	0.8786	0.8384	0.8369	0.8357
Experiment II				
Training	0.9817	0.9587	0.9587	0.9587
Validation	0.8917	0.8752	0.8746	0.8726
Experiment III				
Training	0.9743	0.9604	0.9604	0.9604
Validation	0.8983	0.8796	0.8803	0.8788
Experiment IV				
Training	0.9906	0.9636	0.9636	0.9636
Validation	0.9200	0.8838	0.8828	0.8809

Table 6: Comparison of accuracies obtained from different models

Experiment	Model	Accuracy
Experiment I	CNN only	0.9580
Experiment II	CNN + Attention	0.9817
Experiment III	CNN + GAN	0.9743
Experiment IV	CNN + Attention + GAN	0.9906

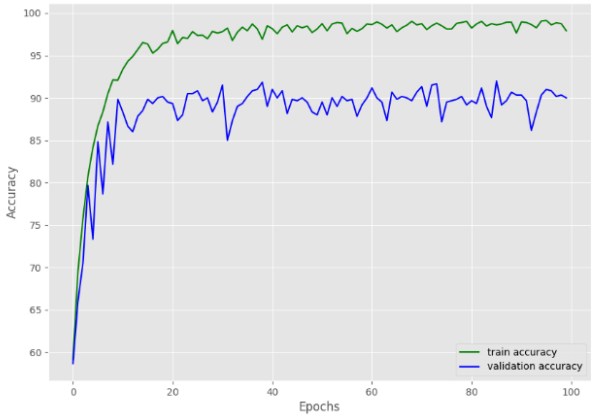


Figure 8: Accuracy graph of the proposed model obtained from Experiment IV

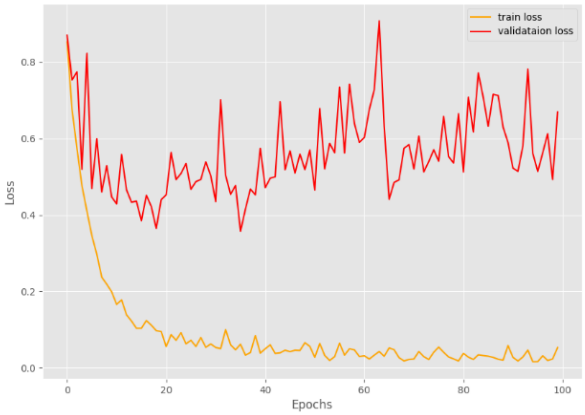


Figure 9: Loss graph of the proposed model obtained from Experiment IV

4. Conclusion

In a nutshell, the results achieved in this study as discussed in the previous section have validated and proved that the proposed objective has been achieved successfully. A double-enhanced CNN model using an attention module and a GAN model has resulted in an improved multi-stage classification performance of Alzheimer's disease, where an accuracy as high as 99% had been achieved. As compared with unenhanced or only single-enhanced CNN models, the classification performance of this double-enhanced model is clearly improved and is the highest among the others. Thus, the significance of the double-enhanced CNN model is proved as it can appropriately solve the challenge in the multi-stage classification of AD, where the MCI stage is difficult to identify, and the issue of limited medical images available. Hence, the findings obtained in this study can significantly contribute to

this research domain by verifying the importance of a double-enhanced classification model for the multi-stage classification of AD. Lastly, future research directions may involve improving the stability of the GAN model for better samples, enhancing the generalization ability of the CNN model to overcome overfitting issues, and trying to incorporate other types of enhancement modules for better classification performance.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

The authors would like to express their gratitude towards Universiti Teknologi Malaysia (UTM) for their financial support for this work under the Fundamental Research Grant (Q.K130000.3843.22H18).

References

[1] W. P. Ching, S. S. Abdullah, M. I. Shapiai, "Enhancing Multi-Stage Classification of Alzheimer's Disease with Attention Mechanism", 2023 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), 230-235, 2023, doi: 10.1109/IICAIET59451.2023.10291792.

[2] M. Liu, F. Li, H. Yan, K. Wang, Y. Ma, L. Shen, M. Xu, "A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease", Neuroimage, **208**, 116459, 2019, doi: 10.1016/j.neuroimage.2019.116459.

[3] W. H. Organization. Dementia. https://www.who.int/news-room/fact-sheets/detail/dementia. (2023).

[4] J. E. Galvin, "Prevention of Alzheimer's Disease: Lessons Learned and Applied", J Am Geriatr Soc, **65**(10), 2128-2133, doi: 10.1111/jgs.14997.

[5] C. S. Lee, P. G. Nagy, S. J. Weaver, D. E. Newman-Toker, "Cognitive and system factors contributing to diagnostic errors in radiology", AJR Am J Roentgenol, **201**(3), 611-617, doi: 10.2214/ajr.12.10375.

[6] A. Association, "2019 Alzheimer's disease facts and figures", Alzheimer's & dementia, **15**(3), 321-387.

[7] Y. Zhang, Q. Teng, Y. Liu, Y. Liu, X. He, "Diagnosis of Alzheimer's disease based on regional attention with sMRI gray matter slices", Journal of Neuroscience Methods, **365**, 109376, doi: 10.1016/j.jneumeth.2021.109376.

[8] J. Liu, M. Li, Y. Luo, S. Yang, W. Li, Y. Bi, "Alzheimer's disease detection using depthwise separable convolutional neural networks", Computer Methods and Programs in Biomedicine, **203**, 106032, doi: 10.1016/j.cmpb.2021.106032.

[9] M. A. Ebrahimighahnavieh, S. Luo, R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review", Computer Methods and Programs in Biomedicine, **187**, 105242, doi: 10.1016/j.cmpb.2019.105242.

[10] A. Mehmood, M. Maqsood, M. Bashir, Y. Shuyuan, "A deep siamese convolution neural network for multi-class classification of alzheimer disease", Brain Sciences, doi: 10.3390/brainsci10020084.

[11] B. C. Simon, D. Baskar, V. S. Jayanthi, "Alzheimer's Disease Classification Using Deep Convolutional Neural Network", 2019 9th International Conference on Advances in Computing and Communication (ICACC), doi: 10.1109/ICACC48162.2019.8986170.

[12] R. Jain, N. Jain, A. Aggarwal, D. J. Hemanth, "Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images", Cognitive Systems Research, doi: 10.1016/j.cogsys.2018.12.015.

[13] D. Lin, A. V. Vasilakos, Y. Tang, Y. Yao, "Neural networks for computer-aided diagnosis in medicine: A review", Neurocomputing, **216**, 700-708, doi: 10.1016/j.neucom.2016.08.039.

[14] R. Cui, M. Liu, A. D. N. Initiative, "RNN-based longitudinal analysis for diagnosis of Alzheimer's disease", Computerized Medical Imaging and Graphics, doi: 10.1016/j.compmedimag.2019.01.005.

[15] J. Koo, J. H. Lee, J. Pyo, Y. Jo, K. Lee, "Exploiting multi-modal features from pre-trained networks for Alzheimer's dementia recognition", arXiv preprint arXiv:2009.04070, doi: 10.21437/Interspeech.2020-3153.

[16] A. Aqeel, A. Hassan, M. A. Khan, S. Rehman, U. Tariq, S. Kadry, A. Majumdar, O. Thinnukool, "A Long Short-Term Memory Biomarker-Based Prediction Framework for Alzheimer's Disease", Sensors, doi: 10.3390/s22041475.

[17] X. Zhou, S. Qiu, P. S. Joshi, C. Xue, R. J. Killiany, A. Z. Mian, S. P. Chin, R. Au, V. B. Kolachalama, "Enhancing magnetic resonance imaging-driven Alzheimer's disease classification performance using generative adversarial learning", Alzheimer's Research & Therapy, doi: 10.1186/s13195-021-00797-5.

[18] J. Islam, Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks", Brain Informatics, doi: 10.1186/s40708-018-0080-3.

[19] J. N. Cabreza, G. A. Solano, S. A. Ojeda, V. Munar, "Anomaly Detection for Alzheimer's Disease in Brain MRIs via Unsupervised Generative Adversarial Learning", 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), doi: 10.1109/ICAIIC54071.2022.9722678.

[20] E. Jung, M. Luna, S. H. Park, "Conditional Generative Adversarial Network for Predicting 3D Medical Images Affected by Alzheimer's Diseases", Predictive Intelligence in Medicine: Third International Workshop, PRIME 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings, Lima, Peru, doi: 10.1007/978-3-030-59354-4_8.

[21] R. Kadri, M. Tmar, B. Bouaziz, F. Gargouri, Alzheimer's Disease Detection Using Deep ECA-ResNet101 Network with DCGAN. 376-385, doi: 10.1007/978-3-030-96305-7_35.

[22] J. Zhang, B. Zheng, A. Gao, X. Feng, D. Liang, X. Long, "A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification", Magnetic Resonance Imaging, doi: 10.1016/j.mri.2021.02.001.

[23] S. H. Wang, Q. Zhou, M. Yang, Y. D. Zhang, "ADVIAN: Alzheimer's Disease VGG-Inspired Attention Network Based on Convolutional Block Attention Module and Multiple Way Data Augmentation", Frontiers in Aging Neuroscience, **13**, doi: 10.3389/fnagi.2021.687456.

[24] X. Zhang, L. Han, W. Zhu, L. Sun, D. Zhang, "An explainable 3D residual self-attention deep neural network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI", IEEE Journal Of Biomedical And Health Informatics, **26**(11), 5289-5297, doi: 10.1109/JBHI.2021.3066832.

[25] J. L. Pamela, L. S. B. Tammie, C. M. John, K. Sarah, H. Russ, X. Chengjie, G. Elizabeth, H. Jason, M. Krista, G. V. Andrei, E. R. Marcus, C. Carlos, M. Daniel, "OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease", medRxiv, 2019.2012.2013.19014902, doi: 10.1101/2019.12.13.19014902.

[26] Phung, Rhee, "A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets", Applied Sciences, **9**, 4500, doi: 10.3390/app9214500.

[27] N. Daldal, Z. Cömert, K. Polat, "Automatic determination of digital modulation types with different noises using Convolutional Neural Network based on time–frequency information", Applied Soft Computing, **86**, 105834, doi: 10.1016/j.asoc.2019.105834.

[28] J. Deng, W. Dong, R. Socher, L. J. Li, L. Kai, F.F. Li, "ImageNet: A large-scale hierarchical image database", 2009 IEEE Conference on Computer Vision and Pattern Recognition, doi: 10.1109/CVPR.2009.5206848.

[29] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition", Proceedings of the IEEE conference on computer vision and pattern recognition, doi: 10.48550/arXiv.1512.03385.

[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, "Attention is all you need", Advances in neural

information processing systems, **30**, doi: 10.48550/arXiv.1706.03762

[31] T. Yang, C. Tong, "Real-time detection network for tiny traffic sign using multi-scale attention module", Science China Technological Sciences, **65**(2), 396-406, doi: 10.1007/s11431-021-1950-9.

[32] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative adversarial networks", Communications of the ACM, **63**(11), 139-144, doi: 10.48550/arXiv.1406.2661.

[33] T. Bai, M. Du, L. Zhang, L. Ren, L. Ruan, Y. Yang, G. Qian, Z. Meng, L. Zhao, M. J. Deen, "A novel Alzheimer's disease detection approach using GAN-based brain slice image enhancement", Neurocomputing, **492**, 353-369, doi: 10.1016/j.neucom.2022.04.012.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 9, No. 1, 9-14 (2024)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Verify of Left and Right Differences in Sleep Index using Ring-type Sensor

Yutaka Yoshida, Emi Yuda*

Tohoku University, Graduate School of Information Sciences, Sendai, 980-8579, Japan

ARTICLE INFO

Article history:
Received: 20 October, 2023
Revised: 19 December, 2023
Accepted: 19 December, 2023
Online: 20 January, 2024

Keywords:
Ring type wearable sensor
Sleep index
Left and right difference

ABSTRACT

In this study, two healthy women in their 40s wore Oura Rings on both fingers and verified left and right differences in sleep index. Bed time, sleep time, rem sleep, sleep score, and heart rate had small left and right differences in both subjects, and a high correlation of 0.8 or more was observed(subjectA:P<0.001,subjectB:P<0.01). Subject B showed a significant difference in the mean values of sleep latency, awakening time and sleep efficiency, and one finger was overestimated(P<0.05). Although the detection accuracy of sleep stages differs depending on the stage, it was shown that the influence of left and right differences is particularly small for bed time, sleep time, rem sleep, sleep score and heart rate.

1. Introduction

This paper is an extension of the 2022 World Automation Congress presentation [1]. Recent ring-shaped sensors can non-invasively measure heart rate, body surface temperature, and body acceleration. Indexes related to sleep time and sleep quality can also be calculated and visualized using a smartphone application. Among ring-shaped sensors, the Oura Ring has received particular attention in recent years and has been verified the accuracy of detecting HRV (heart rate variability) index and sleep stages during sleep using a combination of Electrocardiogram (ECG) and polysomnography (PSG) [2-5]. In addition, studies have been reported on accuracy comparisons with other wearable devices [6-9] and long-term tracking of physiological characteristics related to menstrual cycles and sleep [10].

In a study on accuracy evaluation regarding the wearing position of wearable sensors, it has become clear that the index obtained from wristwatch-type sensors varies depending on the wearing position [11], making it somewhat difficult to evaluate a stable index. However, there are no study reports that examine the left and right differences in the index when the Oura Ring is worn on both fingers.

Therefore, in this study, we wore Oura Ring to both fingers to record biological signals during sleep and verified left and right differences in sleep index calculated by the application.

2. Method

2.1. Experimental protocol

Subjects were two healthy women (subjectA:42y.o., subjectB:41 y.o.). The gist of the study was explained to the two subjects, and consent was obtained. The studies involving human participants were reviewed and approved by Center for Data-driven Science and Artificial Intelligence Tohoku University Institutional Review Board (No.2022-6, Approved 25 August, 2022).

Before going to bed, the subjects wore two ring type wearable sensor on both index fingers to record their sleep index for the night. The measurement period was from January 25, 2022 to October 24, 2022, with 172 days of data for Subject A and 10 days for Subject B. The measurements were not performed consecutively every day, but the subjects were allowed to decide for themselves, and the experiment was performed without straining while taking into account the prevention of skin irritation.

2.2. Ring type wearable sensor

The ring-type wearable sensor used Oura Ring Generation 3(Oura, Inc., USA). Oura Ring has 15 sensors inside a titanium ring and can measure biological information such as heart rate, skin temperature, physical activity, and sleep index.

Figure 1: Oura Ring Generation 3

*Corresponding Author: Emi Yuda, yuda@ieee.org

Pulse wave measurement uses an LED light sensor. Infrared and red LED run in dark places because they are more affected by ambient light, and green LED runs in bright places because they are less affected by ambient light. These sensors measure the amount of blood flow (change in blood vessel volume) that

Figure 2: State of charging

changes with the pulse of the heart. Relative temperature is measured with an NTC (negative temperature coefficient) thermistor. Physical activity and body motion are measured with the 3-axis accelerometer. The measured biological information is transferred to the Oura application via Bluetooth and output to a CSV file.

The size is 6 to 13 (ring width 7.8mm), and the weight is 4 to 6g. Battery life is 5-7 days, charging time is 20-80 minutes, and memory capacity is 16MB. The outside temperature range for use is -10℃ to 52℃.

2.3. Sleep index and Statistical analysis

The sleep index is shown in Table 1. Heart rate is the average heart rate for 5 minutes, and respiratory rate is measured only once during one sleep. Mean values, standard deviations, and correlation coefficients of sleep indices obtained from both fingers were calculated. Mean values were compared by Paired t-test. The significance level was set at 5%, and significant at $P < 0.05$.

3. Results

Table 1 shows the statistical results of the sleep index of the two subjects. For subject A, there was no significant difference between the left and right mean values, but significant correlations were observed in all indices ($P<0.001$). However, the strength of the correlation varies by index. In particular, bed time, sleep time, shallow non-rem, rem sleep, sleep score, heart rate and respiration rate showed a high correlation of 0.8 or more. On the other hand, subject B showed a significant difference in the mean values of sleep latency, awakening time and sleep efficiency, and one finger was overestimated ($P<0.05$). Bed time, sleep time, rem sleep, awakening time, sleep efficiency, sleep score and heart rate showed a significant correlation of 0.8 or more ($P<0.01$). Bed time, sleep time, rem sleep, sleep score, and heart rate were highly correlated with small left and right differences in both subjects.

Figure 3 and figure 4 are scatter plots of the left and right finger sleep indices for subject A and subject B, respectively. The vertical axis is the right finger and the horizontal axis is the left finger. In subject A, the distribution of sleep latency was such that the longer the right finger, the shorter the left finger, or the longer the left finger, the shorter the right finger. Awakening time and sleep efficiency had a square distribution. In subject B, the sleep latency was longer with the right finger than with the left. There is a high correlation of awakening time and sleep efficiency, but awakening time was longer with the right finger, and the sleep efficiency was greater with the left finger, resulting in a bias.

4. Discussion

There are many reports on the accuracy evaluation of HRV and sleep stages measured by Oura Ring during night sleep.

Reference [2] reported detection accuracy of 65%, 51%, and 61% for shallow sleep, deep sleep, and rem sleep, respectively. A study using machine learning combined accelerometers, body temperature, and HRV to detect four stages of light NREM sleep, deep NREM sleep, REM sleep, and wake with 79% accuracy[3].Studies comparing HRV parameters during nocturnal sleep using ECG and Oura Ring have shown very high heart rate correlations during the night, indicating that HR can be measured with high accuracy [4,5]. A study comparing TST (total sleep time) between Samsung Gear Sport watches and Oura Ring showed a correlation of 0.86, and the average difference was within an acceptable range [7]. Based on the results of these previous studies, the cause of the difference between the left and right is thought to be the movement of the fingers and the detection accuracy of the device's algorithm. Since heart rate is detected as an average heart rate over a 5-minute period, there may be some fluctuations from beat to beat, but when looking at the baseline (average value) of heart rate, the difference between the left and right becomes smaller.

Table 1: Sleep index

Bedtime (min)	Awakening time (min)
Sleep time (min)	Sleep efficiency (%)
Sleep latency (min)	Sleep score
Shallow non-rem (min)	Heart rate (bpm)
Deep non-rem (min)	Respiration rate (frequency/min)
Rem sleep (min)	Relative body temperature (℃)

Table 2: Statistical results of the sleep index.

Sleep index	Subject A				Subject B			
	R-L	Mean±S.D.	Correlaion		R-L	Mean±S.D.	Correlaion	
			r	P-value			r	P-value
Bed time (min)	R	403±89	0.843	<0.001	R	461±107	0.832	0.003
	L	404±82			L	437±97		
Sleep time (min)	R	339±68	0.917	<0.001	R	375±68	0.815	0.004
	L	339±67			L	373±68		
Sleep latency (min)	R	11±9	0.195	<0.001	R	14±10*	0.529	<0.001
	L	11±9			L	7±4		
Shallow non-rem (min)	R	194±48	0.812	<0.001	R	218±63	0.577	0.081
	L	197±45			L	211±47		
Deep non-rem (min)	R	40±21	0.590	<0.001	R	33±17	0.324	0.362
	L	38±20			L	42±21		
Rem sleep (min)	R	105±33	0.845	<0.001	R	123±28	0.800	0.006
	L	103±33			L	119±32		
Awakening time (min)	R	64±36	0.500	<0.001	R	87±49*	0.837	0.003
	L	65±34			L	64±50		
Sleep efficiency (%)	R	85±6	0.368	<0.001	R	82±6*	0.824	0.003
	L	84±7			L	87±8		
Sleep score	R	66±9	0.848	<0.001	R	70±8	0.861	<0.001
	L	66±10			L	72±8		
Heart rate (bpm)	R	63.38±3.63	0.994	<0.001	R	64.97±4.50	0.870	0.001
	L	63.39±3.67			L	63.63±4.05		
Respiration rate (frequency/min)	R	15.87±0.45	0.828	<0.001	R	15.31±0.78	0.224	0.534
	L	15.85±0.43			L	15.66±0.40		
Relative body temperature (℃)	R	-0.03±0.26	0.735	<0.001	R	0.05±0.25	0.612	0.06
	L	-0.03±0.27			L	-0.03±0.27		

*P<0.05(right finger vs. left finger)

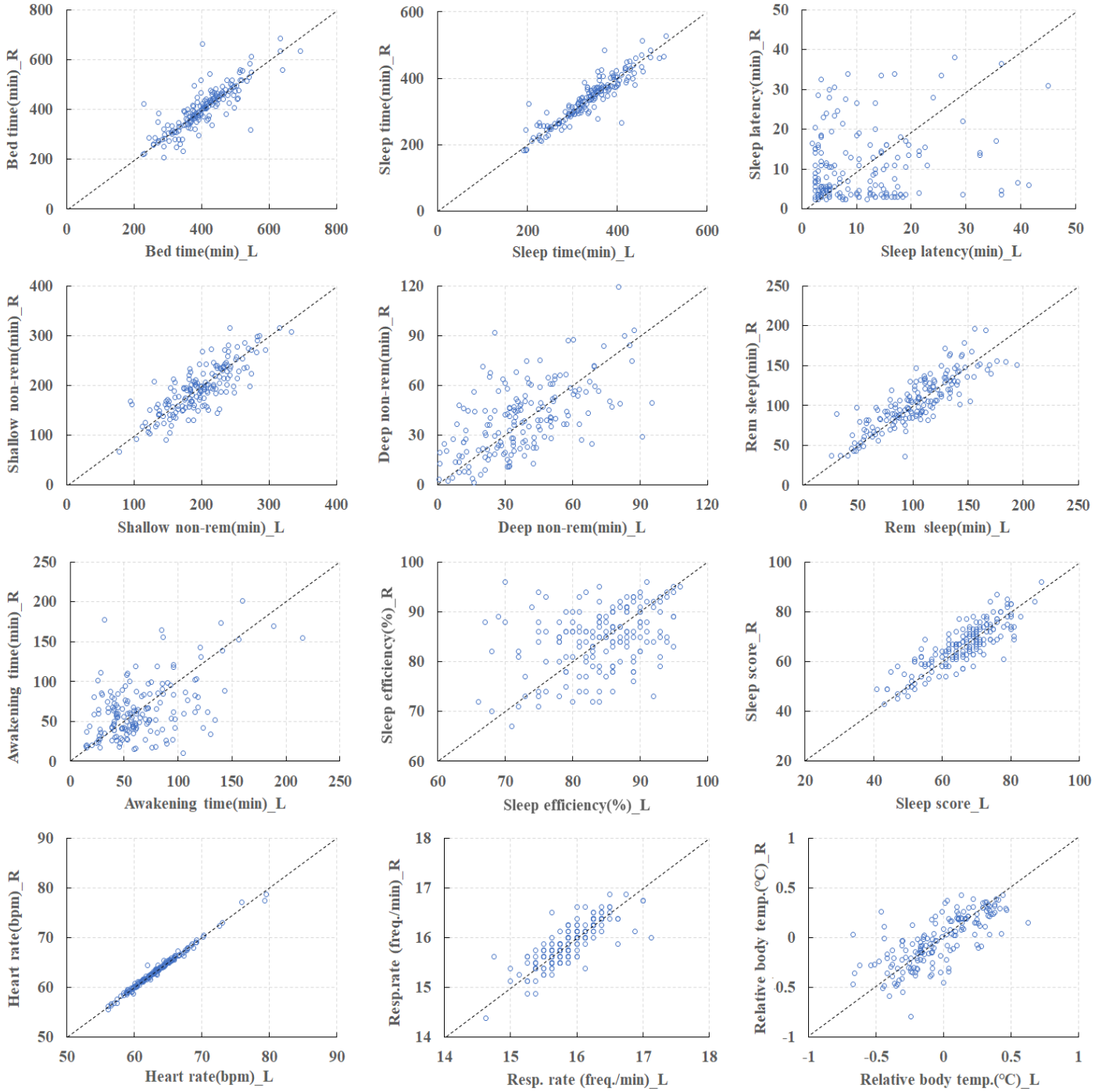


Figure 3: Scatter plots of the left and right finger sleep index for subject A.

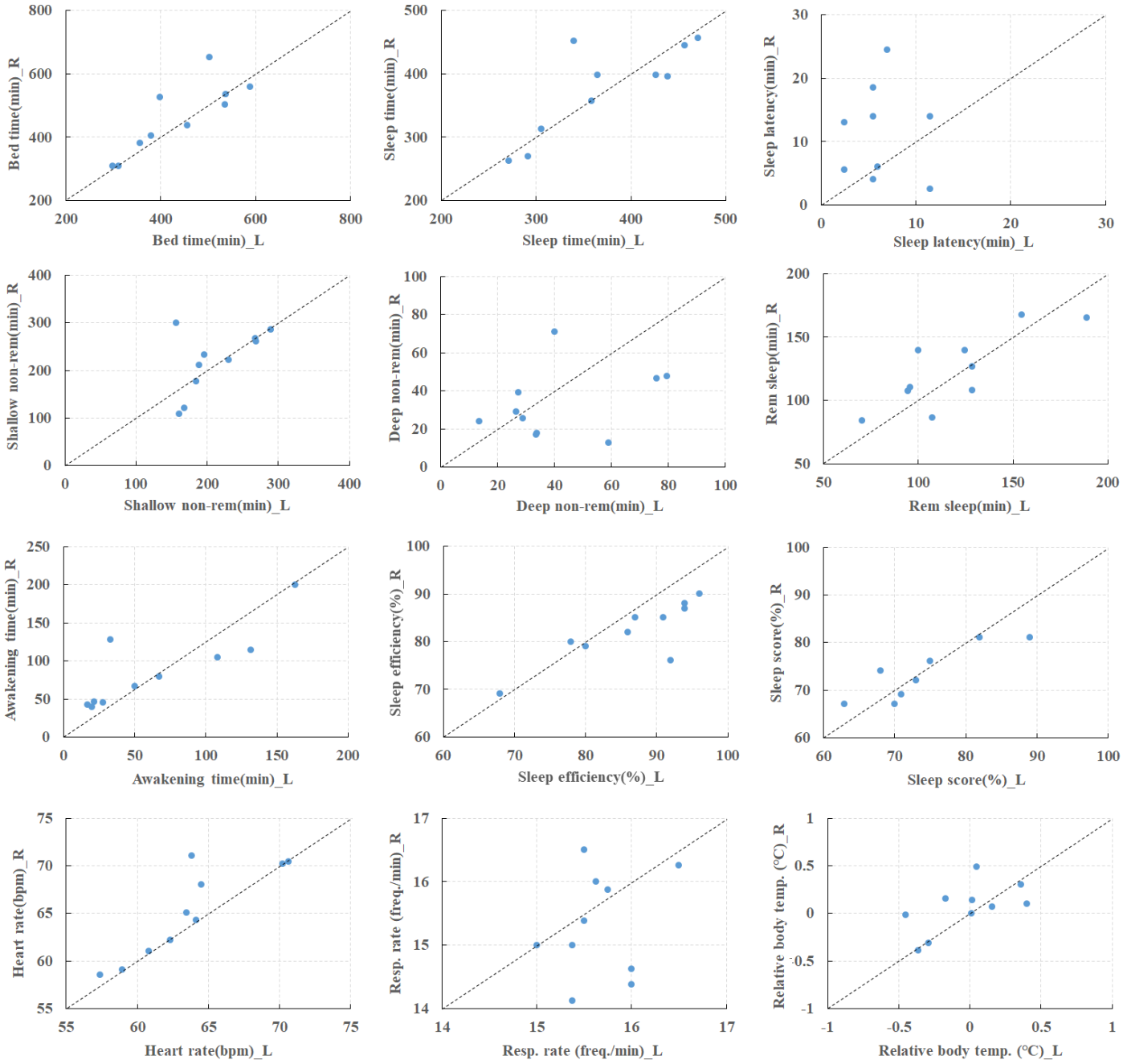


Figure 4: Scatter plots of the left and right finger sleep index for subject B.

In rem sleep, sympathetic nerve activity predominates, and the heart rate increases, whereas in non-rem sleep, parasympathetic nerve activity predominates and the heart rate decreases [12]. Therefore, the stages of rem sleep and non-rem sleep are easily detected, and the left and right differences in rem sleep are particularly small. The accuracy of detecting the shallow non-rem and deep non-rem stages is likely to decrease, and large differences between the left and right sides may occur. On the other hand, heart rate fluctuations are stable during non-rem sleep, but it is difficult to detect the difference between shallow non-rem and deep non-rem. The left and right difference in non-rem sleep may also be due to the algorithm.

Awakening time is the time for turning over in bed, arousal during sleep, and although the Oura Ring's detection algorithm is unknown, it is thought that the acceleration sensor is involved in determining these times. Sleep efficiency is calculated as the ratio of total sleep time to bed time [13-15], and the longer the Awakening time, the more the sleep time is lost, and the lower the sleep efficiency. Therefore, bias occurs due to the difference in movement between the left and right fingers. In subject B, the

right finger had a longer awakening time, so it is thought that sleep efficiency was lower on the right and higher on the left.

Heart rate measured by the Oura Ring during night sleep is measured with high accuracy, and furthermore, it is suggested that wearing it on the left and right fingers does not affect the measurement accuracy. Although the detection accuracy of sleep stages differs depending on the stage, it was shown that the influence of left and right differences is particularly small for bed time, sleep time, rem sleep, sleep score and heart rate.

5. Conclusion

The left and right differences in sleep index obtained from the Oura Ring were verified. There were small left and right differences in bed time, sleep time, rem sleep, sleep score, and heart rate. It is suggested that indices can objectively and accurately evaluate sleep quality and fatigue recovery for health promotion. Other indicators are thought to differ between left and right depending on the individual, and it is necessary to take these differences into account when evaluating sleep. In addition, the subjects' health conditions, skin conditions such as wet skin or dry skin, and body positions during sleep were not recorded in this experiment, so indicators with large variations could not be categorized and analyzed. Considering these things, we will examine the effects of age and gender differences by increasing the number of subjects.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was supported by a Grant-in-Aid for Strategic intellectual property/standardization, commercialization promotion support project for the Beyond 5G era from Ministry of Internal Affairs and Communications, Japan and the New Energy and Industrial Technology Development Organization (NEDO) Intensive Support for Young Promising Researchers Foundation.

References

[1] Y.Yoshida, H.Sakamoto, E.Yuda, " Precision of human workout-time detection using Random Forests and Wearable Sensor Data," 2022 World Automation Congress Proceedings, 429-432, 2022, doi:10.23919/WAC 55640.2022.9934354.

[2] M.D. Zambotti, L.Rosas, I.M.Colrain, F.C. Baker, "The Sleep of the Ring: Comparison of the ŌURA Sleep Tracker Against Polysomnography," Behavioral Sleep Medicine, **17**(2), 124-136, 2019, doi:10.1080/15402002.2017.1300587.

[3] M.Altini, H.Kinnunen, "The Promise of Sleep: A Multi-Sensor Approach for Accurate Sleep Stage Detection Using the Oura Ring,"Sensors (Basel), 21(13), 4302, 2021, doi: 10.3390/s21134302.

[4] R.Cao, I.Azimi, F.Sarhaddi, H.N.Vilen, A.Axelin, P.Liljeberg, A.M. Rahmani,"Accuracy Assessment of Oura Ring Nocturnal Heart Rate and Heart Rate Variability in Comparison With Electrocardiography in Time and Frequency Domains: Comprehensive Analysis," Journal of Medical Internet Research, **24**(1), e27487, 2022, doi:10.2196/27487.

[5] H.Kinnunen, A.Rantanen, T.Kenttä, H.Koskimäki,"Feasible assessment of recovery and cardiovascular health: accuracy of nocturnal HR and HRV assessed via ring PPG in comparison to medical grade ECG,Physiological Measurement", **41**(4), 04NT01, doi: 10.1088/1361-6579/ab840a.

[6] D.J.Miller, C.Sargent, G.D.Roach, "A Validation of Six Wearable Devices for Estimating Sleep, Heart Rate and Heart Rate Variability in Healthy Adults," Sensors (Basel), **22**(16), 6317, 2022, doi:10.3390/s22166317.

[7] M.A.Mehrabadi, I.Azimi, F.Sarhaddi, A.Axelin, H.N.Vilén, S.Myllyntausta, S.Stenholm , N.Dutt, P.Liljeberg , A.M.Rahmani, "Sleep Tracking of a Commercially Available Smart Ring and Smartwatch Against Medical-Grade Actigraphy in Everyday Settings: Instrument Validation Study, " JMIR Mhealth Uhealth, **8**(10), e20465, 2020, doi:10.2196/20465.

[8] E.D.Chinoy, J.A.Cuellar, J.T.Jameson, R.R.Markwald,"Performance of Four Commercial Wearable Sleep-Tracking Devices Tested Under Unrestricted Conditions at Home in Healthy Young Adults," Nature and Science of Sleep, **14**, 493-516, 2022, doi:10.2147/NSS.S348795.

[9] L.Nauha, V.Farrahi, H.Jurvelin, T.Jämsä, M.Niemelä, M.Kangas, R.Korpelainen, "Comparison and agreement between device-estimated and self-reported sleep periods in adults," Annals of Medicine, **55**(1),2191001,2023,doi:10.1080/07853890.2023.2191001.

[10] E.Alzueta, M.D.Zambotti, H.Javitz, T.Dulai, B.Albinni, K.C.Simon, N.Sattari, J.Zhang, A.Shuster, S.C.Mednick, F.C.Baker,"Tracking Sleep, Temperature, Heart Rate, and Daily Symptoms Across the Menstrual Cycle with the Oura Ring in Healthy Women," International Journal of Women's Health, **14**, 491-503,2022, doi:10.2147/IJWH.S341917.

[11] E.Yuda, K.Yamamoto, Y.Yoshida, J.Hayano, "Differences in pulse rate variability with measurement site," Journal of Physiological Anthropology, **39**(1), 4, 2020, doi:10.1186/s40101-020-0214-1.

[12] M.Yoshimoto, I.Yoshida, K.Miki,"Functional role of diverse changes in sympathetic nerve activity in regulating arterial pressure during REM sleep," Sleep, **34**(8),1093-1101,2011,doi:10.5665/SLEEP.1168.

[13] B.V.Vaughn, S.R.Quint, J.A.Messenheimer, K.R.Robertson,"Heart period variability in sleep,"Electroencephalogr Clin Neurophysiol, **94**(3), 155-162,1995, doi:10.1016/0013-4694(94)00270-u

[14] P.Busek, J.Vanková, J.Opavský, J.Salinger, S.Nevsímalová,"Spectral analysis of the heart rate variability in sleep," Physiological Research, **54**(4), 369-376,2005.

[15] D.L.Reed, W.P.Sacco,"Measuring Sleep Efficiency: What Should the Denominator Be? ," Journal of Clinical Sleep Medicine, **12**(2), 263–266, 2016, doi:10.5664/jcsm.5498.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 4, 11-17 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Design and Manufacturing of Soft Grippers for Robotics by Injection Molding Technology

Helmy Dewanto Bryantono[1], Melsiani Rosdiani Fillipin Saduk[1,2], Jiaqi Hong[1], Meng-Hsun Tsai[1], Shi-Chang Tseng[1,*]

[1]*Mechanical Engineering Department, National Yunlin University of Science and Technology, Douliu, 640301, Taiwan*

[2]*Mechanical Engineering Department, State Polytechnic of Kupang, Kupang, 85361, Indonesia*

ARTICLE INFO

Article History:
Received: 25 March, 2023
Accepted: 22 June, 2023
Online: 21 July, 2023

Keywords:
Deformation
Injection molding
Soft gripper
Soft Robot

ABSTRACT

Soft robots have softer, more flexible, and more compliant characteristics than traditional rigid robots. These qualities encourage more secure relationships between people and machines. Nevertheless, traditional robots continue to rule the commercial sector. Due to the high cost of gripper production, soft robots are very far from being commercially feasible. This research focuses on fabricating a soft robotic gripper with the potential for mass production using injection molding technology. The material used for manufacture is Thermoplastic Elastomer (TPE). This study gives an injection molding optimization strategy by using Moldex 3D simulation to minimize production time for soft grippers. Furthermore, using an Ansys workbench, this study simulated soft gripper deflections with variable pressures by finite element analysis and then compared it with the actual experiment. The simulation results of TPE warpage volume shrinkage are 11.969% and 11.96% in the molding experiment. Thus, the shrinkage and warpage for the simulation and actual experiment are similar. According to the simulation outcome, the success of TPE hollow injection molding facilitates soft gripper creation. The maximum pressure used in the FEM simulation of the bending experiment was achieved at the pressure of 50 kPa with 152.02 mm of deformation and compared to the experimental data, 145,03 mm. This error is less than 5%. Finally, a better soft gripper design was achieved by Ansys simulation, and the soft gripper appears to be ready for mass-produced by TPE injection molding.

1. Introduction

Robots have typically been made of stiff materials in order to generate accurate, predictable movements with minor deformation. As a result, rigid robots can accomplish activities with great precision and accuracy in controlled environments. However, nature regularly utilizes compliant and flexible structures in unstructured conditions to generate movement and manipulation. Therefore, because of such organic designs, soft robotics is becoming increasingly popular [1–6].

Soft grippers are needed for soft robots to generate motion and force. Soft grippers can develop the flexibility and adaptation necessary to construct extraordinarily adaptive robotics for soft interactions in the absence of an inflexible skeleton in soft robots. Because of its versatility in filling the gap left by typical rigid robots, the soft gripper is gaining popularity. Therefore, one area where soft robotics can have a substantial impact is the development of soft grippers. Soft grippers are one kind of soft robotics that uses softness to provide highly compliant and adjustable grasping capability [7].

The utilization of rigid grippers designed to handle delicate objects is frequently problematic or limited. As one potential solution, a soft device that can passively adjust to the object and its surroundings could replace the gripper's role. As a result, numerous soft mechanical architectures have been proposed, and extensive research on soft robotics has been conducted [8–10]. Soft robotics, unlike rigid robotics, can adapt to changing environmental conditions and interact with people more securely and flexibly. As a result, soft robot design and fabrication have significantly progressed over the last ten years, and new actuator design and fabrication processes have been created and improved [11,12].

The development of novel materials and soft components is advancing in order to create lighter, simpler general grippers. Compliance has always been considered important in grasping. Shocks induced by contact with a firm gripper and a rigid object, if not adequately regulated, can injure the product or force it outside of the targeted route. Incorporating sensitive materials into robotic end effector grabbing portions is a basic but only partially successful alternative (for example, in the shape of basic rubber pads). Interaction between bodies, on the other hand, impedes their mobility [13–15]. Soft grippers use the responsiveness and elasticity of materials to build highly flexible robotic arms that allow for secure contact between devices and the surrounding. Therefore, when making soft grippers, it is vital to consider the materials and production processes [16–18].

Soft grippers, in contrast to standard robots, are often made from materials with Young's moduli equivalent to soft

*Corresponding Author: Shi-Chang Tseng, Douliu, *640301*, Taiwan, tsengsc@yuntech.edu.tw

www.astesj.com
https://dx.doi.org/10.25046/aj080402

biological materials such as muscles, tendons, and skin, typically about 1e-7GPa-1GPa. Compatibility and "softness" are produced using morphology and material characteristics in naturally pliable materials. However, soft materials have limitations regarding potential manufacture, non-linear response, modeling complexity, self-repair, and fatigue performance. Therefore, elastomers and rubbers, which can sustain reversibly (>100%) due to tremendous loads, are the most frequently utilized soft grip materials [16,19]. Thermoplastic elastomers have emerged as the most popular option for industrial applications in this study because of their low toxicity, robustness, and mechanical dampening characteristics and their simplicity of manufacture [20].

However, due to the high cost of gripper production, soft robots are very far from commercially feasible. When it comes to mass production, forming is the most cost-efficient and has the lowest unit cost of the three basic types of manufacturing: forming, machining (subtractive), and additive manufacturing [21].

In addition, the intricacy of the fabrication procedure may result in a more extended fabrication period that may last from hours to days, and the high expense of typical gripper production methods all contribute to this. Therefore, an efficient fabrication process is required to ensure acceptable mechanical performance while reducing fabrication time and complexity.

Thus, this research focuses on fabricating soft robotic grippers that have the potential for mass production. Consequently, this research concentrates on the material used for soft grippers and a unique process for manufacturing soft grippers with high throughput and mass production potential. For this reason, TPE hollow injection molding was chosen because it is a high-speed, automatic technique that can produce parts with a wide range of sizes and extremely complicated geometries in large quantities.

2. Theoretical Background

There are numerous mathematical models for viscosity accessible in Moldex3D. TPE is used in all injection molding simulations in this study. Moldex3D's suggested viscosity model for this material is the Modified Cross Model 3, often called the Cross-WLF model. In this model, the melt viscosity is dependent on the shear rate $\dot{\gamma}$, temperature T, and pressure p, according to:

$$\eta(\dot{\gamma},T,p)=\frac{\eta_0(T,p)}{1+\left(\frac{\eta_0(T,p)\dot{\gamma}}{\tau}\right)^{1-n}} \quad (1)$$

where η_0 (zero-shear rate viscosity), n (power-law index with a value between 0 and 1), τ^* (relaxation stress). The viscosity versus shear rate obtained from the Cross Model using the TPE material is depicted in Figure 1 for temperatures of 190°C, 215°C, and 240°C, respectively.

Table 1: Cross-WLF model data for TPE melt viscosity [22]

n	0.264907	-
Taus	153557	dyne/cm^2
B	4.30664e-06	g/(cm.sec)
Tb	10819	K
D	0	cm^2/dyne

The shear-dependent parameters are identical to those in the Modified Cross Model. Correlating low-temperature viscosity with the Cross-WLF model is often more accurate. The Cross-WLF model outperforms the Cross-Exp model at temperatures below Tg+100°C.

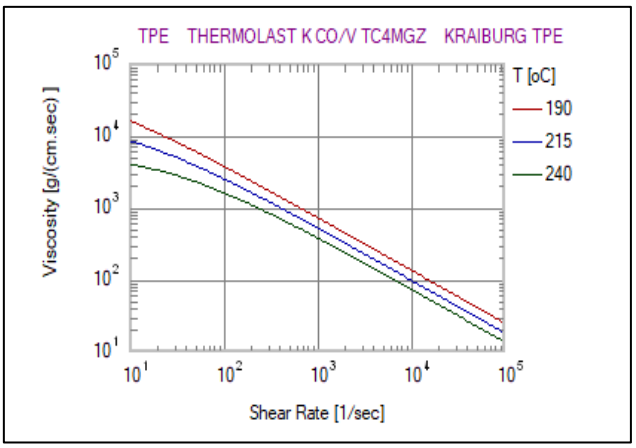


Figure 1: Viscosity versus shear rate for TPE

Modified Tait Model 2, used in this work and offered by Moldex 3D, can be used to explain the PVT model.

$$v(T,p)=v_0(T)\left[1-Cln\left(1+\frac{p}{B(T)}\right)\right]+v_t(T,p) \quad (2)$$

$v(T,p)$ is the specific volume (contrarywise proportional to density) at Temperature, T. Pressure, P, $v_0(T)$ is a specific volume at zero gauge pressure, C is a constant of 0.0894 and $B(T)$ accounts for the material pressure sensitivity, as shown in Figure 1. For the upper-temperature region $(T > T_t)$ and the lower-temperature region $(T \leq T_t)$ respectively. Figure 2 illustrates the specific volume dependency for TPE on Temperature and pressure.

Table 2: PVT data for TPE [22]

b1L [0.8415, 1....	1.03922	cc/g
b2L [0.0005017...	0.000713614	cc/(g.K)
b3L [5.13e+08, ...	9.5127e+08	dyne/cm^2
b4L [1e-06, 0.0...	0.00474979	1/K
b1S [0.8263, 1....	1.0367	cc/g
b2S [0.000352, ...	0.000548239	cc/(g.K)
b3S [5.2e+08, ...	1.01167e+09	dyne/cm^2
b4S [0.002665, ...	0.00302857	1/K
b5 [338.1, 493.1]	448.973	K
b6 [1.9e-09, 04...	2.75172e-08	cm^2.K/dyne
b7 [0, 0.0711]	0	cc/g
b8 [0, 0.41]	0	1/k

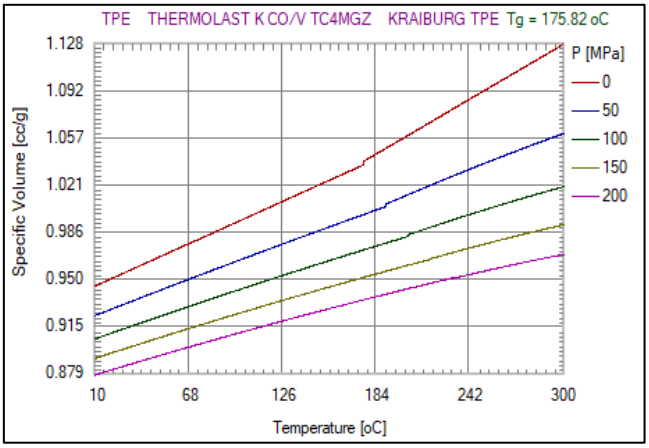


Figure 2: PVT diagram for TPE

Because these are equilibrium states, the PVT characteristics of polymers in the molten state may be appropriately described by the equation of state (EoS).

The generalized Newtonian fluid (GNF) theory predicts that the polymer melt will behave [23,24]. Because the GHS flow model depicts the flow within two plates that are close together, the width of the gap is considered to be substantially slighter than the other flow parameters. This notion concludes that local geometry has the greatest influence on flow at a single spot. Because velocity in the gap-wise orientation is ignored and pressure is simply an expression of planar coordinates, the lubrication approximation may be used. The following equations describe the melt flow of the Hele-Shaw polymer:

Continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \quad (3)$$

Where ρ (density) and t (Time), and **v** (velocity vector). Equation (3 is a standard continuity equation.

Momentum equation:

$$\frac{\partial}{\partial t}(\rho \mathbf{v}) + \nabla \cdot (\rho \mathbf{v}\mathbf{v}) = -\nabla \mathrm{p} + \rho g + \nabla . \tau \quad (4)$$

Where g (gravity) and τ (shear stress).

Energy equation:

$$\frac{\partial}{\partial t}\left(\rho C_p T\right) + \nabla .\left(\rho C_p \mathbf{v} T - k \nabla \mathrm{T}\right) = \eta \dot{\gamma}^2 \quad (5)$$

where T (temperature field), C_p (heat capacity), k (thermal conductivity coefficient), η (viscosity of the fluid), and $\dot{\gamma}$ (shear rate).

The function used in the elasticity problem is the structure's total potential energy. In the first step, assume an estimated form for the solution, and the answer should be acceptable and meet the crucial internal compatibility and boundary restrictions. The optimal constant values are then obtained using the stationary potential energy theorem. The theory states that among all acceptable arrangements of a conservative system, the ones that satisfy the equations of equilibrium render the potential energy constant when allowable displacement fluctuations are taken into consideration., i.e., equilibrium configurations are defined when:

$$\frac{\partial \Pi_p}{\partial a_i} = 0 \quad (6)$$

where Π_p (potential) and a_i (ith dof).

$$\Pi_p = U + \Omega \quad (7)$$

where U (strain energy) and Ω (potential of the load system). In a stressed body, strain energy is provided by

$$U = \frac{1}{2}\int \{\varepsilon\}^T [E]\{\varepsilon\} dv \quad (8)$$

where $[E]$ (constant elastic matrix). It expresses the connection between stress $\{\sigma\}$ and strain $\{\varepsilon\}$.

The potential of the load system is

$$\Omega = -\{u^e\}^T\{F^e\} \quad (9)$$

The total potential of the system is,

$$\Pi_p = U + \Omega = \frac{1}{2}\int \{\varepsilon\}^T [E]\{\varepsilon\} dv - \{u^e\}^T\{F^e\} \quad (10)$$

A shape function, $[N]$, is used to relate the vector of displacements in the element, $\{u\}$, to the vector of nodal displacements, $\{u^e\}$.

After that, the strain vector may be calculated by partial differentiation of the displacement vector, giving

$$\{\varepsilon\} = \left\{\frac{du}{dx}\right\} = \left[\frac{dN}{dx}\right]\{u^e\} = [B]\{u^e\} \quad (11)$$

where $[B]$ is the strain-displacement matrix. Substituting (10) into (9),

$$\Pi_p = \frac{\{u^e\}^T}{2}\int [B]^T[E][B] dv \{u^e\} - \{u^e\}^T\{F^e\} \quad (12)$$

Applying the stationary potential energy theorem,

$$\frac{\partial \Pi_p}{\partial \{u^e\}} = 0 = \int [B]^T[E][B] dv \{u^e\} - \{F^e\} \quad (13)$$

$$\{F^e\} = \int [B]^T[E][B] dv \{u^e\} \quad (14)$$

And as $\{F^e\} = [K^e]\{u^e\}$:

$$[K^e] = \int [B]^T[E][B] dv \quad (15)$$

The soft gripper body in this research is made of a thermoplastic elastomer (TPE), a hyper elastic material. Because the material feature of a hyper elastic material is non-linear under external forces, the strain energy density function is extensively employed to describe the mechanical features of TPE material. In this paper, the Yeoh model proposes a non-linear association between material stress and strain. Yeoh thought that the strain tensor invariants I_2 had no effect on the strain energy and might be ignored entirely. Here is the simplified strain energy density function:

$$W = \sum_i^N C_{i0}(\overline{I_1} - 3)^i + \sum_{k=1}^N \frac{1}{d_k}(J-1)^{2k} \quad (16)$$

where N, C_{i0}, and d_k are material constants discovered during the material experiment and $J = 1$ for incompressible materials. The typical two-parameter form is

$$W = C_{10}(I_1 - 3)^1 + C_{20}(I_1 - 3)^2 \quad (17)$$

The 2nd Yeoh Parameter hyper-elastic material model was used to model the TPE material and fit its average stress-strain data.

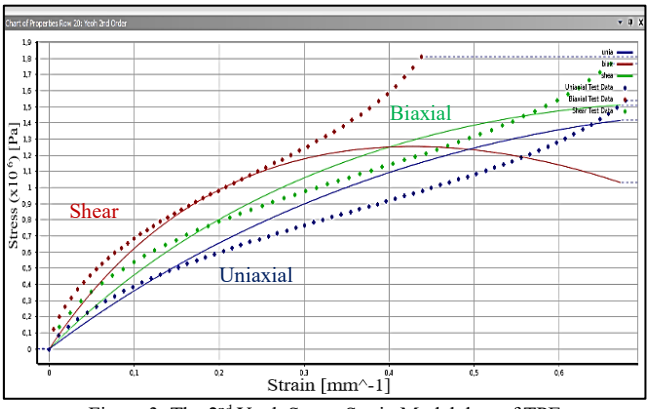


Figure 3: The 2nd Yeoh Stress-Strain Model data of TPE

Table 3: TPE Hyper-Elastic Material Model Constants

Model for TPE Hyper-Elastic Materials	Constant Material	Value (Pa)
The 2nd Yeoh Parameter	C10	6.6275+05
	C20	-62451
	Incompressibility Parameter D1	0
	Incompressibility Parameter D2	0

3. Design and Experimental of TPE Soft Gripper

3.1 Soft Gripper Design

A soft gripper form was created in this study, as illustrated in Figure 4. The TPE soft grippers' air chamber is enclosed within the gripper body. Once air pressure is applied and motion is formed, this air chamber enlarges to accomplish bending deformation. The dimension of the TPE soft gripper is 114 x 19 x 16.5 mm [1]. This design has been proven can accomplish pleasant bending deformation by previous studies [25].

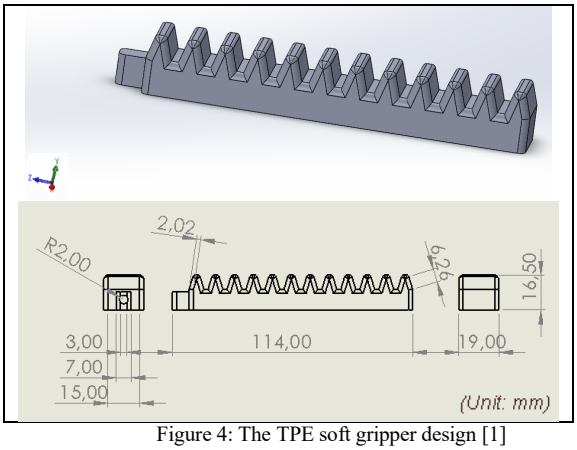


Figure 4: The TPE soft gripper design [1]

3.2 Soft Gripper Mold Design

The soft gripper's mold design was achieved after designing a soft gripper body, as shown in Figure 5.

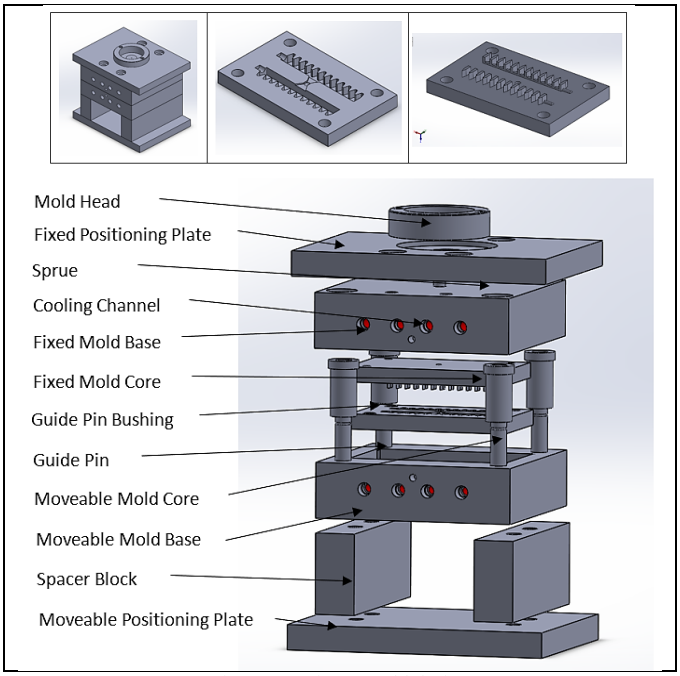
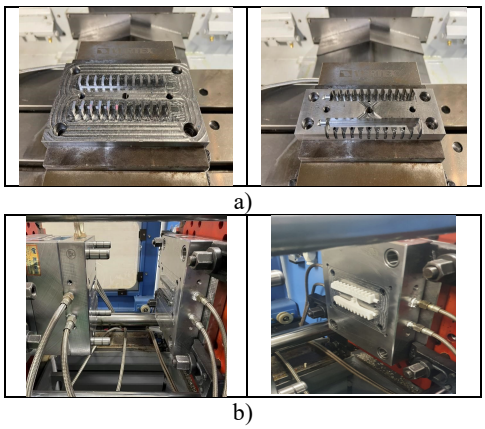


Figure 5: Gripper mold design

3.3 Mold Manufacturing

A High-Speed/Closed-Loop Hybrid Injection Molding Machine produces the soft gripper (AF Series). The soft gripper is meant to be flexible in this experiment, and TPE was employed. Figure 6 depicts the soft gripper mold used to create the soft gripper.



a)

b)

Figure 6: TPE Hollow Gripper a) Mold [1] and b) Injection Molding Machine

3.4 The Simulation of Soft Gripper

Simulations carried out in ANSYS Workbench using a Static Structural Analysis. The finite element diagram in Figure 7 illustrates and describes the processes necessary to model a soft gripper.

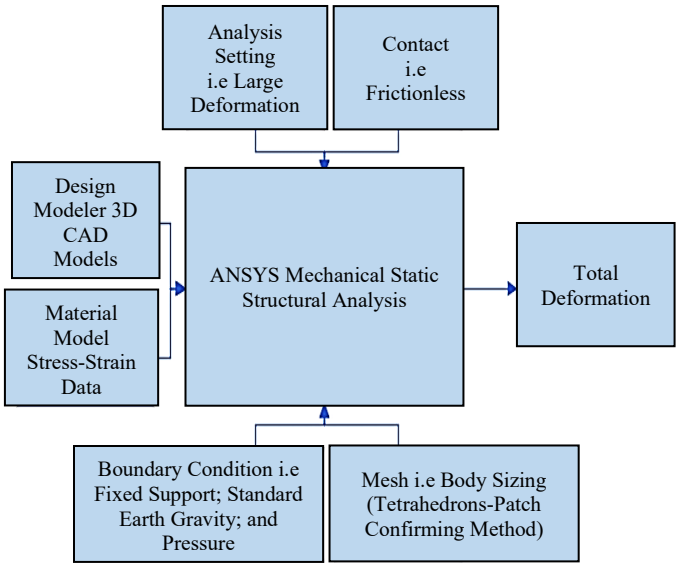


Figure 7: Diagram of finite element simulation

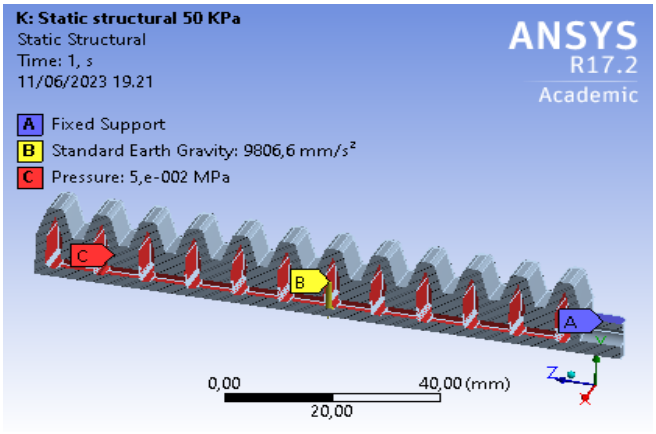


Figure 8: Simulation boundary condition diagram

TPE soft gripper uses tetrahedrons meshed, with the element size adjusted to 1 mm. The whole deformation is caused by pressure given to the TPE soft gripper, exhibiting the gripper is stretched at various phases of the pressure applied. Figure 8 displays a boundary condition simulation of the soft gripper.

Aside from the numerical simulations used (Ansys software), further unique and powerful numerical approaches for strain-stress analysis of isotropic or anisotropic media have recently been developed. The "Gaussian Quadrature" and "Bezier" techniques, among others, demonstrated more stability and accuracy than previous numerical approaches. Alternatively, these approaches may forecast soft gripper deflections with variable pressure [26,27].

3.5 Grasping Experiment

The grasping experiment is applied in this research to examine the condition of the gripping force. Figure 9 depicts the fixture design, which includes the length, pitch between the chambers, and air pressure. As a result, the plastic ball object is grasped well without any defect on the object, as shown in Figure 15.

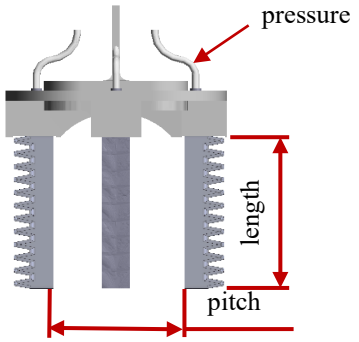


Figure 9: The fixture design: length, pitch, and the air pressure

4. Result and Discussion

4.1 Injection Molding Simulation

Moldex 3D has been used to simulate Injection Molding tests of TPE soft gripper. First, as shown in Table 4, the results were achieved by simulating the injection molding process. Following that, the filling phase simulation is seen in Figure 10.

Table 4: Soft Gripper Injection Molding Simulation

Parameter	Value
Melt Temperature (°C)	210
Cooling Channel Temperature (°C)	20
Mold Temperature (°C)	50
Maximum Injection Pressure (MPa)	30
Maximum Packing Pressure (MPa)	20
Cooling Flow Rate (cm^3/sec)	120
Filling Time (sec)	2
Cooling Time (sec)	30
Holding Time (sec)	2
Mold-Open Time (sec)	5
Eject Temperature (°C)	70

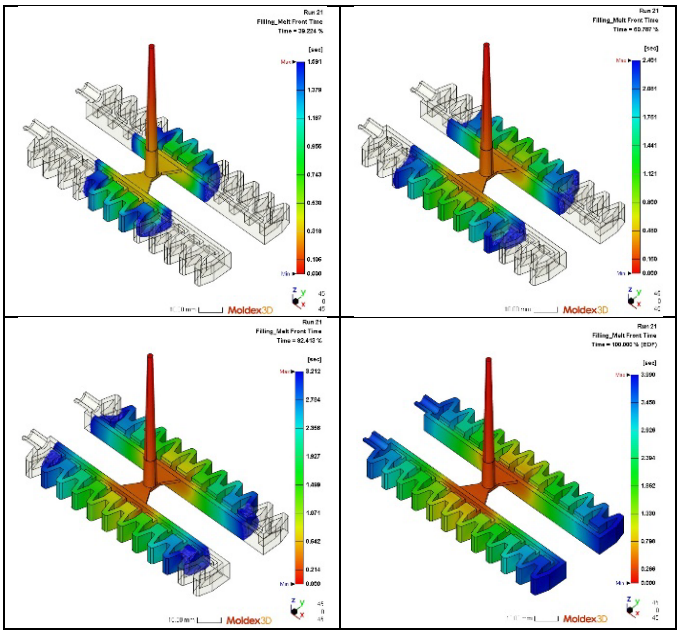


Figure 10: Filling phase simulation of the TPE 40%, 60%, 80%, and 100%

4.2 Soft Gripper Manufacturing

This manufacturing of the soft gripper is displayed in Figure 11. To prevent the TPE material from absorbing moisture and causing pores in the finished product. The material must be dried before being injected into the molding machine. After determining the injection volume, the finished product has shrinkage and slight overflow. Therefore, the mold temperature has to adjust to reduce the mold temperature and increase the packing pressure to prevent the overflow and shrinkage of the finished product. The finished product is displayed below in Figure 11.

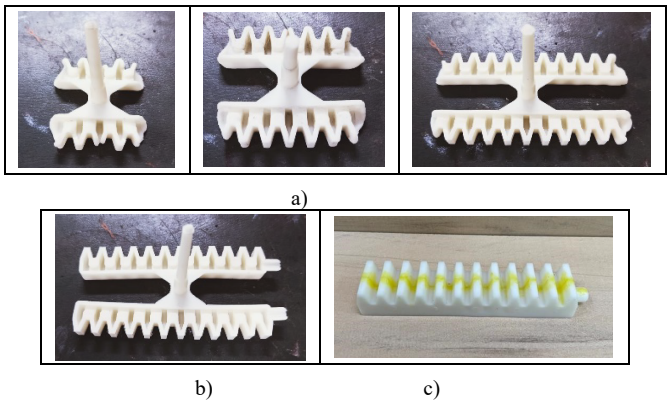


Figure 11: a) Filling experiment of the soft gripper 40%, 60%, and 80%, b) Gripper ejected from the mold after 100% filling, and c) Final gripper after glued

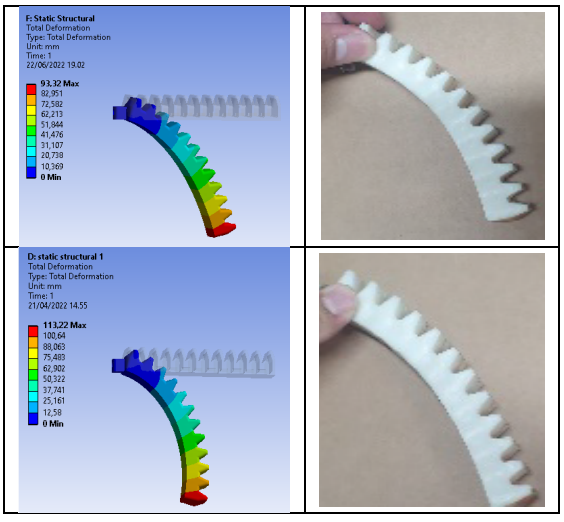
Compared to traditional production, the production time might range from hours to a day. However, most soft gripper manufacturers use 3D printing, which takes 35 minutes [28]. Therefore, the production time of one soft gripper is less than 5 minutes in this work, and we obtained an optimization technique for lowering time and expenses in creating soft grippers.

4.3 Warpage and Shrinkage Comparison

By the result of the total warpage displacement of Moldex 3D simulation and the experiment of the injection molding machine, the final of the simulation and experiment of the soft gripper. According to the result of Moldex 3D, warpage occurs, and the total warpage volumetric shrinkage is 11.969% based on the simulation. The original design size of the soft gripper was 114x19x16.5 mm. However, due to the warpage, the dimension of the soft gripper changed to 103x18.64x16.43 mm. Thus, the warpage volume shrinkage of the actual experiment is 11.96%. Remarkably similar for both simulation and experiment results.

4.4 Bending Comparison

The TPE soft gripper's bending experiment performs well in the bending pressure experiment, both positive and negative pressure. Furthermore, this design works as expected, considering the bending restriction from the previous research is eliminated as the gripper has no edge touching for the negative bend.



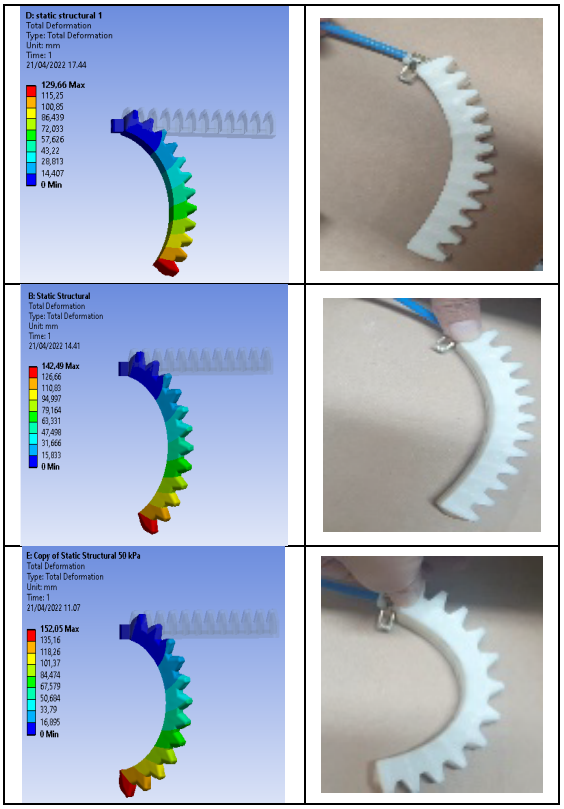


Figure 12: Actual positive bending deformation of soft gripper

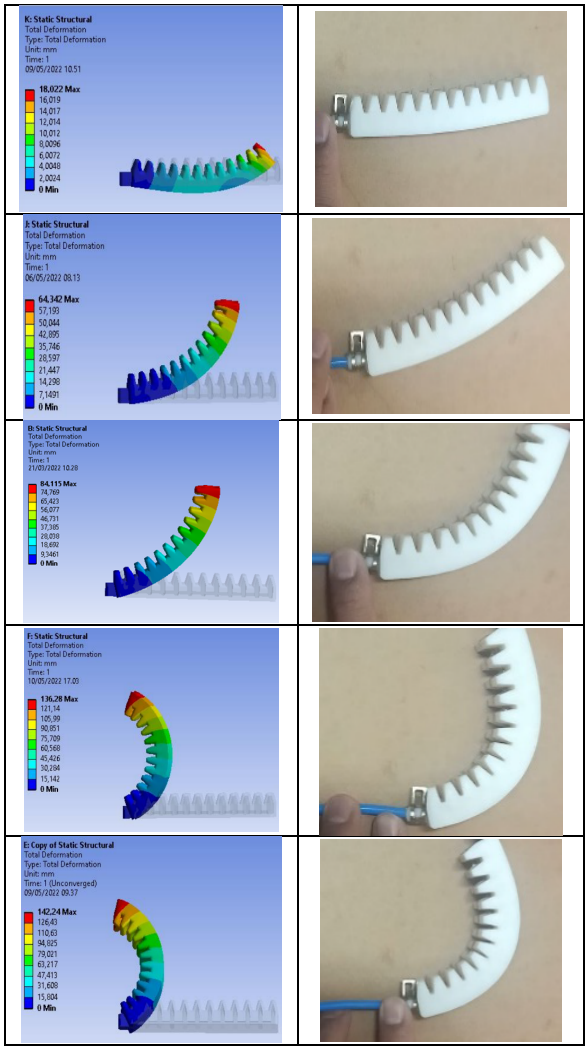


Figure 13: Actual negative bending deformation of soft gripper

However, this product's weakest sections can also be identified as the areas where fastening processes are used. Thus, processes involving high pressure or heavy loads increase the product's propensity to failure. Under excessive pressure, it will lead to failure and may cause the material to burst, rip, or deform permanently. Due to the fact that this soft gripper design must be glued after being manufactured, significant air pressure may cause separation between the two sections of the gripper finger as well. Furthermore, the chemicals employed in bonding can potentially harm the environment. Therefore, in this experiment, the range of air pressure applied only between -50 to 50 KPa, as shown in Figures 12 and 13.

Based on the comparison between simulation and actual experiments of bending deformation of the soft gripper with several pressure differences, Figure 14 shows the influence of the pressure.

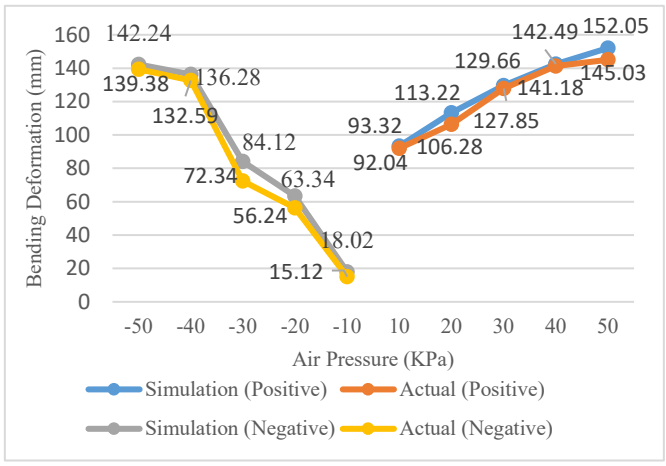


Figure 14: TPE bending deformation comparison influenced by the pressure

4.5 *Grasping Experiment*

Three soft grippers have been constructed using the fixture design in Figure 9 and utilized to pick up a softball as the goal is to examine the grasping force, as shown in Figure 15. The air pressure applied to the gripper is 40 KPa in order to get enough bending deformation to grasp the object. A 3D print weight object was also attached to the ball to examine the bending strength to encounter the additional weight beside the ball.

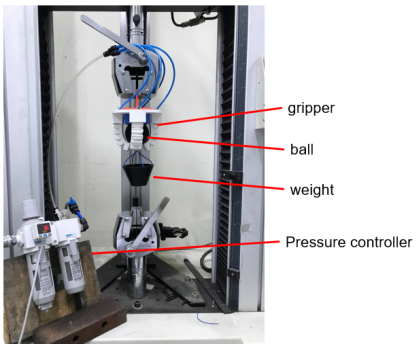


Figure 15: Grasping experiments of the soft gripper [1]

5. Conclusion

The numerical analysis of injection molding was done by Moldex 3D. The soft gripper deformation was successfully predicted by Ansys software, then proved by actual experiment. In both positive and negative pressure conditions, higher air pressure causes more widespread deformation of the soft gripper. Injection molding and FEM simulations indicated that creating a soft gripper is straightforward. Furthermore, the

modeling and real outcomes are fairly similar. In Moldex 3D simulation, the value of shrinkage is 11.969%, compared to the experiment of the soft gripper is 11.96%. Thus, the percentage warpage volumetric shrinkage for the simulation and actual experiment is similar. However, for future research, the directions may be to discover other convenient designs or material combinations to fulfill manufacture on a mass scale following the needs of Industry 4.0. It is recommended to use an injection molding machine as the manufacturing method since it is a popular technique for producing plastic components on a massive scale.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to thank the Ministry of Science and Technology (MOST) in Taiwan for the financial support of MOST-107-2221-E-224-045.

References

[1] M.R.F. Saduk, H.D. Bryantono, S.C. Tseng, "Engineering Design of Soft Gripper and Manufacturing by TPE Hollow Injection Molding," in Proceedings - 2022 IET International Conference on Engineering Technologies and Applications, IET-ICETA 2022, Institute of Electrical and Electronics Engineers Inc., 2022, doi:10.1109/IET-ICETA56553.2022.9971524.

[2] M.R. Fillipin Saduk, H. Jiaqi, S.C. Tseng, "Modeling and Experiments of a Soft Gripper for Robotics Arms Applications," in Proceedings of the 2022 8th International Conference on Applied System Innovation, ICASI 2022, Institute of Electrical and Electronics Engineers Inc.: 179–182, 2022, doi:10.1109/ICASI55125.2022.9774440.

[3] C. Majidi, Soft Robotics: A Perspective - Current Trends and Prospects for the Future, Soft Robotics, **1**(1), 5–11, 2014, doi:10.1089/soro.2013.0001.

[4] D. Rus, M.T. Tolley, Design, fabrication and control of soft robots, Nature, **521**(7553), 467–475, 2015, doi:10.1038/nature14543.

[5] D. Trivedi, C.D. Rahn, W.M. Kier, I.D. Walker, "Soft robotics: Biological inspiration, state of the art, and future research," Applied Bionics and Biomechanics, **5**(3), 99–117, 2008, doi:10.1080/11762320802557865.

[6] S. Kim, C. Laschi, B. Trimmer, Soft robotics: A bioinspired evolution in robotics, Trends in Biotechnology, **31**(5), 287–294, 2013, doi:10.1016/j.tibtech.2013.03.002.

[7] J. Shintake, V. Cacucciolo, D. Floreano, H. Shea, Soft Robotic Grippers, Advanced Materials, **30**(29), 2018, doi:10.1002/adma.201707035.

[8] P. Bharath Vamsi, V. Ragavendra Rao, "Design and fabrication of soft gripper using 3D printer," in IOP Conference Series: Materials Science and Engineering, Institute of Physics Publishing, 2018, doi:10.1088/1757-899X/402/1/012026.

[9] T. Hassan, M. Manti, G. Passetti, N. d'Elia, M. Cianchetti, C. Laschi, "Design and development of a bio-inspired, under-actuated soft gripper," in 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE: 3619–3622, 2015.

[10] R. Maruyama, T. Watanabe, M. Uchida, "Delicate grasping by robotic gripper with incompressible fluid-based deformable fingertips," in 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE: 5469–5474, 2013.

[11] Y. Yang, Y. Wu, C. Li, X. Yang, W. Chen, "Flexible Actuators for Soft Robotics," Advanced Intelligent Systems, **2**(1), 1900077, 2020, doi:10.1002/aisy.201900077.

[12] J. Pinskier, D. Howard, "From Bioinspiration to Computer Generation: Developments in Autonomous Soft Robot Design," Advanced Intelligent Systems, **4**(1), 2100086, 2022, doi:10.1002/aisy.202100086.

[13] M. Kaur, W.S. Kim, "Toward a Smart Compliant Robotic Gripper Equipped with 3D-Designed Cellular Fingers," Advanced Intelligent Systems, **1**(3), 1900019, 2019, doi:10.1002/aisy.201900019.

[14] F. Schmitt, O. Piccin, L. Barbé, B. Bayle, Soft robots manufacturing: A review, Frontiers Robotics AI, **5**(JUN), 2018, doi:10.3389/frobt.2018.00084.

[15] Z. Wang, Y. Torigoe, S. Hirai, "A Prestressed Soft Gripper: Design, Modeling, Fabrication, and Tests for Food Handling," IEEE Robotics and Automation Letters, **2**(4), 1909–1916, 2017, doi:10.1109/LRA.2017.2714141.

[16] J. Hughes, U. Culha, F. Giardina, F. Guenther, A. Rosendo, F. Iida, Soft manipulators and grippers: A review, Frontiers Robotics AI, **3**(NOV), 2016, doi:10.3389/frobt.2016.00069.

[17] S. Zaidi, M. Maselli, C. Laschi, M. Cianchetti, "Actuation Technologies for Soft Robot Grippers and Manipulators: A Review," Current Robotics Reports, **2**(3), 355–369, 2021, doi:10.1007/s43154-021-00054-5.

[18] G. Alici, "Softer is harder: What differentiates soft robotics from hard robotics?," in MRS Advances, Materials Research Society: 1557–1568, 2018, doi:10.1557/adv.2018.159.

[19] L. Wang, F. Iida, "Deformation in Soft-Matter Robotics: A Categorization and Quantitative Characterization," IEEE Robotics and Automation Magazine, **22**(3), 125–139, 2015, doi:10.1109/MRA.2015.2448277.

[20] Z. Wang, K. Or, S. Hirai, "A dual-mode soft gripper for food packaging," Robotics and Autonomous Systems, **125**, 2020, doi:10.1016/j.robot.2020.103427.

[21] W. Zhou, J. Lin, D.S. Balint, T.A. Dean, "Clarification of the effect of temperature and strain rate on workpiece deformation behaviour in metal forming processes," International Journal of Machine Tools and Manufacture, **171**, 2021, doi:10.1016/j.ijmachtools.2021.103815.

[22] M.-L. Wang, R.-Y. Chang, C.-H.D. Hsu, Molding simulation: Theory and practice, Carl Hanser Verlag GmbH Co KG, 2018.

[23] H. Schlichting, K. Gersten, H. Schlichting, K. Gersten, "Fundamentals of boundary-layer theory," Boundary-Layer Theory, 29–49, 2000.

[24] M.R. Barone, C. Tucker, Fundamentals of computer modeling for polymer processing, Hanser, 1989.

[25] Y. Hao, T. Wang, Z. Ren, Z. Gong, H. Wang, X. Yang, S. Guan, L. Wen, "Modeling and experiments of a soft robotic gripper in amphibious environments," International Journal of Advanced Robotic Systems, **14**(3), 2017, doi:10.1177/1729881417707148.

[26] S.E. Mousavi, N. Sukumar, "Generalized Gaussian quadrature rules for discontinuities and crack singularities in the extended finite element method," Computer Methods in Applied Mechanics and Engineering, **199**(49–52), 3237–3249, 2010, doi:10.1016/j.cma.2010.06.031.

[27] H. Kabir, M.M. Aghdam, "A generalized 2D Bézier-based solution for stress analysis of notched epoxy resin plates reinforced with graphene nanoplatelets," Thin-Walled Structures, **169**, 2021, doi:10.1016/j.tws.2021.108484.

[28] B.N. Peele, T.J. Wallin, H. Zhao, R.F. Shepherd, "3D printing antagonistic systems of artificial muscle using projection stereolithography," Bioinspiration and Biomimetics, **10**(5), 2015, doi:10.1088/1748-3190/10/5/055003.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 9, No. 1, 15-32 (2024)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Improved Candidate-Career Matching Using Comparative Semantic Resume Analysis

Asrar Hussain Alderham*, Emad Sami Jaha

Faculty of Computer and Information Technology, Department of Computer Science, King Abdulaziz University, Jeddah, 22233, Saudi Arabia

ARTICLE INFO

Article history:
Received: 12 November, 2023
Revised: 27 December, 2023
Accepted: 27 December, 2023
Online: 20 January, 2024

Keywords:
Semantic Resume Attributes
Comparative Description
Ranking

ABSTRACT

A resume is a prevalent and generally employed method for individuals to showcase their proficiency and qualifications. It is typically composed using diverse customized, personalized methods in multiple inconsistent formats (such as pdf, txt, doc, etc.). Screening candidates based on the alignment of their resume with a set of job requirements is typically a laborious, challenging, time-intensive, and resource-intensive endeavor. This work is crucial for extracting pertinent information and valuable attributes that indicate acceptable applicants. This study aims to improve the candidate career-matching process for human resource (HR) departments by implementing automation and increasing efficiency. Machine learning (ML) and natural language processing (NLP) techniques are applied to infer and analyze comparative semantic resume attributes. Using semantic data comparisons, the ranking support vector machine (RankSVM) algorithm is subsequently applied to rank these resumes based on attributes. RankSVM detects tiny differences among candidates and assigns unique scores, resulting in an improved ranking of candidates based on their suitability for job requirements and from the best to worst match for the vacancy. The experiment and performance comparison results show that the proposed comparative ranking, which relies on semantic descriptions, outperforms the standard ranking based on regular scores in distinguishing candidates and distributing resumes across the ranks with an accuracy of up to 92%. Eventually, we obtained a list of the top ten candidates out of 228 technical specialists' resumes.

1 Introduction

This paper is an extension of work initially presented at the 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN) [1], which improves the candidate-career matching process. In the beginning, the Internet has played a leading role in several interactive communication fields. It has become crucial for job seekers and employers to find distinct candidates for their vacancies. Nowadays, job applications are predominantly conducted via online job portals or services. The job requirements are published by a human resources (HR) officer, while candidates upload their resumes via a specialized recruitment website.

Subsequently, the process of selecting an appropriate candidate for the job requirements requires time and human effort. Recruiters search available resume databases using a job-related keyword to select pertinent resumes for a specific job, but the traditional engine search is based on keyword filtering techniques without understanding the related semantic information of different resumes. This can lead to similar resumes, which may not be closely related [2].

For example, in the technical domain (computer-related technologies), special information and common resume features such as skills, tools, and technologies may exist in the same single domain as computer science. Therefore, the capability of extracting all of these special skills and attributes can distinguish each resume from the others resumes, which is usually performed manually and is preferable to be partially/fully automated because it is an important, complex, and time-consuming task within any HR [3].

A resume is a document, either in printed or electronic form, that presents job-related knowledge and information to employers. The resume includes personal information, education information, work experience, technical skills under hard skills, leadership, creativity, time management, qualifications, and preference sections under soft skills. Each is written openly based on the author's personality and writing style. Furthermore, a resume typically comprises a document in which the information is presented in various file formats and structures. Therefore, it is probable that the format and content of numerous resumes submitted by applicants vary, necessitating a

*Corresponding Author: Asrar Alderham, King Abdulaziz University, KSA, +966559813248 & aalderham@stu.kau.edu.sa

review, evaluation, and filtration process to ensure improved candidate selection. HR job sections may, therefore, encounter difficulties and challenges when attempting to extract useful information and select potential candidates.

It is, therefore, imperative to keep pace with the astounding advances in machine learning (ML) technologies, as computing power becomes challenging when job requirements require further analysis and re-filtering of resumes by setting more precise criteria for extracting relevant resumes.

Much research has been done in resume analysis, including classification, summarization, and extracting information using various techniques such as ML, natural language processing (NLP), and Ontology. Such research efforts have faced many problems regarding the accuracy of selecting the most eligible candidates' resumes. Extensive resume analyses and further research studies are still needed to address these problems.

In this work, we aim to assist HR officers in improving the extraction, filtering, and selection of the most job-suited resumes in an automated and more accurate manner based on analyzing semantic resume attributes using NLP tools. Our contribution considers two groups of per-attribute semantic descriptions used for each resume as relative and comparative labels. This comparative description is more precise and informative than categorical descriptions. However, comparative labels must be derived from all possible comparisons between every two resumes per attribute. Moreover, using a ranking method like the ranking support vector machine (RankSVM) algorithm [4], results in discriminative ranking for all resumes per single attribute, per multiple attributes, or by all attributes.

The rest of this paper is organized as follows. The following sections 2, 3, 4, and 5 presents background in brief, the process of extracting information from resumes and candidate-career matching, relative and comparative descriptions. In addition, the motivations, the research paper question, and the work objectives. A reference to related works on resume analysis that employ various methodologies and strategies is provided in section 6. Section 7 subsequently provides a comprehensive outline of the methodology employed in this study and the proposed approach. After that, in section 8, the dataset, the experiments, and the results are described in detail. The paper is concluded in the 9 section, and in the 10 section, we suggest directions for future work.

2 Background

A brief background on candidate-career matching, resume parsing, and the information extraction process will be given in this section. Then, the relative and comparative descriptions will be highlighted.

2.1 Candidate-Career Matching

Job/career matching is an operational issue that is ever more important in every society. A candidate-career (job-candidate) matching task aims to assign the right job to the right candidate (applicant). Candidate-career matching is a tedious process that HR officers carry out. There are so many resumes to analyze in a limited time, with a lot of work to pass through each candidate's application and identify the best matches for the job. Hence, unfortunately, any possible negligence in this process may lead to the wrong candidate being put in the wrong job by a human mistake [5], which in turn can cause a loss for both employers and candidates.

There has been a rapid increase in jobs and employment in recent years through numerous internet recruitment platforms. Moreover, nowadays, because of the COVID-19 pandemic, millions of employers and job seekers would prefer or be obliged to conduct their hiring or job-seeking attempts besides job interviews through an online recruitment platform [6].

In [7], Russell stated, "Core personality is made up of traits that have been conditioned over many years. Such traits are critical in assessing a candidate's ability to perform virtually any aspect of any job". He also stated in the candidate's selection process, "The selection process is clearly the most critical and controllable variable in the development of a productive and successful work team."

The proper match between the candidate and the job is crucial for HR in hiring the right candidates, and it is also beneficial for those candidates to avoid involvement in work that is not right for them. As a result, increasing the matching process of candidates with occupations through proper resume selection and analysis is crucial for improving accurate applicant employability and developing a successful employment process, which in turn helps enhance employee job performance.

Three factors can affect a person's job performance capabilities as mentioned in [8], which can be valuable and insightful for this research work:

1- Organization Match

The degree to which a candidate's attitudes, values, ethics, and grooming match those required by the job is called an organizational match. Face-to-face interviews are commonly used to assess these factors. However, relying solely on face-to-face interviews is risky. The halo effect is an issue that occurs when interviewers see a part of themselves in a job candidate. On the other hand, the opposite of the halo effect is another potential pitfall of the interview process: unconscious bias. The more a candidate differs from the interviewer, the more conscious effort the interviewer must make to view the candidate positively or neutrally.

2- Skills Match

The degree to which a candidate's educational background, technical skills, previous job experience, and specific expertise match those required for the job is referred to as a skills match. Many job positions necessitate specific sets of knowledge or technical skills. According to research, those in charge of selecting these positions are frequently biased toward believing that expertise and intelligence are significant. However, more than expertise and skill intelligence are required. According to job matching research, people perform better when fully engaged in the challenges of the job.

3- Job Match

Job match is a major component of a candidate's success on the job. It refers to how much a person's cognitive skills, such as how fast and effectively they learn. Interests include whether a person desires to work with people, data, or things and personality, such as the ability to be part of a team, make decisions, manage clients, etc.

All three factors or cornerstones of job performance are essential in matching the candidate to the job. Therefore, in this research, inspired by these pivotal factors, we focused on the applicant selection

process based on the job requirements by analyzing semantic attributes provided in their resumes and finding the differential aspects between them.

2.2 Resume Parsing and Information Extraction

Resume parsing is the automated extraction of information from websites or inconsistently formatted documents, such as resumes, using complex pattern matching or language analysis tools. This procedure aims to create a possible recruiting database in various formats [9].

Resumes can contain semi/unstructured text, vary in information type, order, and writing style, and represent various file formats ('.txt,".pdf,".doc,".docx,".odt,".rtf,' etc.). A resume is typically divided into sections reflecting the candidate's competencies. Due to this diversity, resumes are difficult to parse, and extracting useful information is challenging. The recruitment or HR staff spends significant time and effort parsing resumes and extracting relevant data [10]. So, it is crucial to have an accurate parsing system for the resume section. To effectively and efficiently analyze the data from various forms and structures of resumes, the automation model must not rely on orders or types of information [11].

Information extraction (IE) technologies, such as NLP, take natural language text as input and help analyze the text efficiently and effectively to discover valuable and relevant knowledge that can be used to produce usefully structured information.

Since resumes are written in human languages, the computer used to parse candidate resumes must be constantly trained and adapted to deal with human languages and different expressions in the writing of resumes. The capabilities of machine/deep learning techniques, considered among the most effective approaches under the artificial intelligence (AI) umbrella, allow a model to learn patterns in data without being programmed [12]. Furthermore, NLP can also be employed for such purposes and used to understand human language in resumes and extract useful information embedded in them.

NLP can be described as using computational methods to process free text in spoken or written form, which serves as a mode of communication commonly used by humans [13]. The main objective of NLP is to analyze unstructured text and represent its meaning using pipeline processing steps. These processes operate at two levels: the syntactic level, which deals with the structure and grammar of the text, and the semantic level, which focuses on the meaning and interpretation of the text. The syntactic level involves dividing the raw text of a document into sentences using a sentence segmenter. Each sentence or statement is then further divided into words and punctuation, known as tokens, using a tokenizer. Subsequently, every token is assigned part-of-speech tags (such as nouns, verbs, adjectives, adverbs, etc.), which will play a significant role in the named entity recognition (NER) process. This step identifies all occurrences of a specific entity type in the text. The final step involves using relation recognition to search for possible links between the different entities in the text at the semantic level, where each word is analyzed to determine the meaningful representation of the sentence [14]. Figure 1 shows the architecture for a simple information extraction system.

We have this background to use as a basis for our goal. To improve the process of matching candidates to job requirements, we propose a new approach that differs from existing work on the topic of resume analysis, where many of these existing related works will be reviewed in detail in section 6. Unlike traditional approaches, our proposed approach aims to enhance resume analysis using comparative descriptions. This involves extracting relevant information from resumes using ML and NLP techniques. These comparative descriptions are expected to be more precise and informative in distinguishing between resumes based on their semantic attributes. Previous research in various domains [4, 15], and [16] has shown that such comparative descriptions are superior to other forms of description.

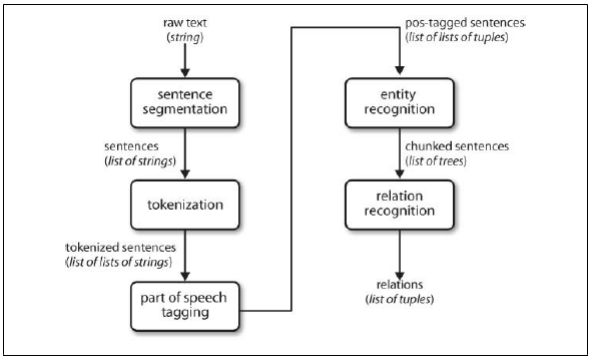


Figure 1: Simple pipeline architecture for an information extraction system [17]

2.3 Relative and Comparative Description

The parsing may fail and not extract high-quality data. If the search is more specific in describing job requirements, the search in most approaches is typically achieved based on an exact match of the search keyword or an explicit, predefined list of related words. However, other valuable implicit and semantic information is not analyzed or considered in the resulting retrieved data, which HR officers can usually infer and consider while manually searching and selecting candidates. Therefore, to fill this semantic gap between how humans and machines analyze resumes, we propose a new method to define several semantic attributes that describe different aspects of the process, like personal information, education level, experiences, certifications, etc. These semantic attributes may be explicitly or implicitly included in a resume. We need to parse them by searching keywords and giving each attribute a set of descriptive comparative labels demonstrating its significance.

Such semantic attributes can either be binary attributes associated with categorical labels or relative attributes associated with comparative labels [4]. Categorical (absolute) labels can be defined as nameable descriptions used to describe semantic attributes, such as education level, which can be labeled with any of the categories (BSc, MSc, Ph.D., etc.). Relative attributes are used to describe the degree of presence using labels such as "Very high," "High," "Average," "Low," "Very Low," and "None." Whereas comparative labels are named descriptions used to describe only relative attributes; for instance, they can be used to describe the education level of one resume based on the comparison with the education level in another resume. In other words, these labels describe the degree of comparison of relative attributes, using suitable labels such as

"Much higher," "Higher," "Same," "Lower," and "Much lower." The proposed relative attributes are expressed using a bipolar five-point scale from 1 to 5. Corresponding comparative labels are assigned on a scale from 2 to -2. such that 2 is associated with the " Much higher" label, while -2 is associated with the "Much Lower" label. Thus, the relative attributes can represent the strength of the measured attributes but can also be comparable, making it easier to more accurately differentiate the minor underlying differences in the descriptive attributes between one resume and another [18]. As indicated in [19], there may be several expected advantages to describing the resume features in a relative (comparative) format:

1. It makes resume descriptions clearer and more meaningful (for example, resume A is better than resume B).
2. It enables comparisons with a reference resume object (e.g., resume A is higher than Asrar's resume).
3. It makes attribute-based searching more efficient (e.g., search for a much higher resume with respect to skills).

3 Motivations

In our lives, getting a job that fits our ambition, qualifications, and abilities is very important, as matching our experience and knowledge with the job opportunities increases the efficiency of our production and progression in the labor market.

Since the advent of the Internet and the accelerating increase in online employment opportunities, matching candidates with job requirements is a tedious process that takes much time, effort, and cost to achieve. With many candidates seeking jobs, the need for a workforce increases to sort and analyze their resumes and choose the most suitable with high accuracy. For instance, when only two vacancies are required, and the applicants are more than a thousand candidates, each has his resume written differently to suit his thinking and how he presents his qualifications. It is often an inconsistently structured document that is difficult to analyze with the required accuracy automatically, ensuring distinguishing and the candidate matches the job from hundreds of others.

Proposing a comparative description-based approach adds to the existing resume analysis works a new viability expected to increase the current accuracy and efficiency of the selection of the candidates. It differs from merely comparing one resume with the other using traditional comparison methods. It compares the attributes of all resumes with each other based on a range that contains five points, making the selection and differentiation process for resumes more precise and informative.

4 Paper Research Question

Does the resume analysis based on semantic attributes described using comparative descriptions along with NLP and ML techniques can improve the process of candidate-career matching and selection?

5 Objectives

The main objectives of this work can be summarized as follows:

- To assist HR officers in improving the extraction, filtering, and selection of the most job-suited resumes in an automated, more accurate manner.
- To enhance existing work on resume analysis and candidate-career matching.
- To extract and organize the resume information related to a group of semantic attributes using relative /comparative semantic descriptions.
- To score the resume using initial relative labels and then compare them using comparative labels based on evaluation metrics for each semantic attribute.
- To rank resumes per attribute and then per all/some attributes based on comparative descriptions by matching the extracted semantic attributes from the resume to the job requirements.
- To improve the accuracy of candidate selection based on more accurate comparative descriptions, enabling the detection of tiny differences between candidates.

6 Related Work

Much work has been published on automatically extracting and analyzing information from resumes to improve candidate-career matching and selection. While many techniques and approaches are used in unstructured resume analysis, most researchers focus on ML and NLP to extract entities from resumes to ensure the best match between applicants and job requirements.

6.1 Resume Analysis and Candidate Selection using NLP

Regarding resume analysis and the selection of candidates, in addition to the job domain allocation process and the classification of applicants in the field of ML and NLP techniques, several research studies have been summarized from 2018 to 2022 since they were interested in the current research scope, as will be illustrated about the studies [10], [20]–[43]. Started with [20], targeted at the challenge of resume data extraction using supervised and rule-based methods, which rely on hierarchical knowledge and large amounts of labeled data that are overly difficult to collect. They proposed a two-step of resume information extraction method. First, lines of raw resume text are segmented using a Naive Bayes (NB) classifier, which predicts each phase label in the resume. Second, after reducing text noise, term frequency-inverse document frequency (TF-IDF) cosine similarity is computed, and K-means clustering shows the resume attribute cluster. Writing style is a new feature proposed, whereas each line's writing style comprises a word index, punctuation index, lexical attribute, and classifier prediction results. So, in the second step, the writing style is utilized to determine each semi-structured text relevant block and module components to get the structured resume text. The technique is practical and efficient, as demonstrated by experiments on a real-world dataset on information technology (IT) professional resumes.

To create a model that could extract valuable information from resumes, the authors of [21] used both Rule-based and Deep-learning approaches. Their work focuses on Vietnamese resume files in the IT domain. The approach of this study contains 4 phases: First, text segmentation; second, named entity recognition (NER); third, a combination of convolutional neural networks (CNN), bidirectional long short-term memory (Bi-LSTM), and conditional random field (CRF) to find name entities, and finally, text normalization. Above 81% accuracy in NER is a significant achievement for their model. This work demands a lot of time and people if they want an extensive enough data set.

In [22], the authors explained the proposed system used, which improves the decision-making process regarding candidates, as it consists of multiple modules: the section-based segmentation module, the filtration module, and the category-based matching module. They deal with the problem of selecting candidates who match the job requirements, and the human resources staff also exerts manual effort. They also addressed the issue of extracting information from unstructured resumes. The system uses an ML algorithm to build the applicant learning model and NLP to extract information. Note that this work covers resume analysis with all its sections and information like personal, education, experience, technical skills, internship, hobbies, etc. However, a limitation of the system is the absence of a research sample to evaluate the accuracy and efficacy.

In [23], they addressed two issues; the most significant was obtaining employer-interesting information from resume content. The second issue is structuring resumes automatically for database entry. They constructed a computer software that answers any question about resume content. A question is posed as a string of keywords that may not need to be grammatically correct. The response is supplied as a text block or a single resume clause. Their process of resume-text processing has four steps. Convert a resume from Word or PDF to text by Tika system, identify text keywords, create two directories, each with a ¡key, value¿ pairs, and finally, text diapason definition. This paper gives examples of resume text extraction to answer arbitrary queries. Their approach allows for a structured resume through clustering and keyword labeling. The shortcoming of these methods is that they apply to process only any short text.

The study in [24] developed a web application that predicts the best resumes for a particular job description. The web app's design has two types of users: candidate and recruiter users. On the candidate side, the SpaCy model and regular expression (RegEx) of NLP extract different sections of data from the uploaded resume. On the recruiter side, Job descriptions, like resumes, must be scanned and parsed to extract essential data. Then, the TF-IDF cosine similarity compares a candidate's resume to the job description. Finally, the output is a ranked list of applicants based on total resume scores visible only to the recruiter. There were problems with parsing unstructured resumes; the candidate's education years were added to their job experience years, which were calculated. So, the applicant had to provide manual input for their job experience to count toward their overall resume score, in addition to the web application problems in working on .doc files only.

To avoid the time-consuming and redundant process of manually allocating projects to new hires by opening and analyzing their resumes one by one. In [25], they discussed developing and implementing a resume classifying application that uses NLP to obtain only the essential and relevant information. In addition, an ensemble learning-based voting classifier consisting of 6 individual classifiers to classify a candidate profile into a suitable domain based on his interest, work experience, and expertise mentioned in the profile. The resume classifier application produces a bar graph showing candidates' suitability for various domains. The model uses Stack-Overflow REST API topic modeling techniques to add a new domain to the list of domains if the candidate profile fails to meet the confidence threshold value. The model results are encouraging, and their studies are distinguished by their focus on IT.

In [26], the authors offered a two-step process to rank applicants according to how well they match the requirements of a job vacancy. The first step is to create a resume parser that extracts complete information. This parser is a web app. In the second step, they employ bidirectional encoder representations from transformers (BERT) as sentence pair classification to rank job candidates based on matching the job descriptions by using a candidate's past resume experience descriptions. Their experiments first extracted LinkedIn resume text into predefined sections and achieved 100%

In [27], they created a resume analysis and position recommendation system that included competitiveness, personality traits, and job recommendation analysis with resume diagnosis. The diagnosis function uses ML and text mining to analyze the user's uploaded resume. It helps job applicants understand their competitiveness in the market and generates a list of recommended job openings based on their resumes. Job-seekers and companies immediately summarize their analysis results in reports and charts for job fairs. The job vacancy recommendation results were evaluated using a questionnaire survey on the applicant side. The experimental results confirmed that the job vacancies recommended by the created system met the expectations of job candidates.

One research work using NLP techniques for extracting information from legal documents is [28]. They have suggested using open information extraction from legal documents. Open information extraction involves extracting data from large datasets without a predefined data set. A sample case from the indiankanoon.com website is put into the system as a text document. As a first step, they performed data cleaning, including removing ambiguities, tagging different parts of speech (POS), chunking names, recognizing them using NER, and extracting relationships between them. No numerical results demonstrate the efficiency and accuracy of their proposed approach.

As the traditional recruitment process lacks speed and accuracy in selecting the best match of candidates for the job, a new model was proposed by [29]. They used the NLP method to summarize resumes in different formats by extracting important information for the job requirements focused on the technical skills and listing each summarized resume in a .txt file. The summarization model specification consists of four integrated modules implemented in Python: conversion, extraction, general, and resume list. The research concluded that the auto summarization model of resumes is not an efficient solution because it might not work well in all industries regarding the different skill sets needed in the different areas of recruitment.

In [30], they suggested an automated method for 'resume categorization and matching' to speed up the candidate selection and

decision-making process and further overcome obstacles to the proper screening of candidates. Their model is built on ML, which makes resume recommendations to HR based on job descriptions. This model works in two phases: extracting the features using NER of NLP and then classifying the resumes into the correct categories using LinearSVM—moreover, finding similarities between the resume and the job description using TF-IDF cosine similarity. The authors demonstrated the result of their model in the resume classification phase, with an accuracy of 78.53%.

Traditional hiring requires manual resume screening, which takes time and energy. Advanced machine learning-based NLP automates the resume screening and hiring process in this paper [31] by extracting entities from resumes using SpaCy and NER models and generating a graph showing each resume's score. The model is evaluated on 20 resumes, and the predicted summary resumes are saved as .txt files for each. The experiment achieved an F1-score of 91.35, a precision of 90.14, and a recall of 92.60. Their entity-recognized model achieved a 91.35 F1 score, 90.14 precision, and 92.60 recall.

In [32], they proposed a resume parser system using NER of Stanford CoreNLP and pattern matching to convert the resume data into a structured format and extract information about the candidates for the recruiting process. Based on skills extracted using TF-IDF and logistic regression classification techniques, the system predicted the candidate's genre, such as computer science, accounting, finance, statistics, business development, etc. The experiment was conducted with 100 resumes from students and job seekers from various universities. The result of the overall resume prediction accuracy of their system was 91.47%.

In this proposed system of [33], the authors made the entire enlistment process more practical and economical. As a result, they have put in place a system that compares applicants' resumes with job descriptions and displays the similarity percentage. Cosine similarity is the basis of the system under consideration. The NLP technology extracts the knowledge and vital abilities required for the specific employment position. It shows the recruiter the candidates' similarity results, which aids in selecting the best candidates. Additionally, candidates are ranked based on their similarity score to obtain the most cost-effective outcome possible. HR departments would reduce their burden if the system were implemented across various industries with a high demand for qualified workers.

Because resumes are unstructured, human-written documents, it is urgent to understand the context of the words. So, in [10], they addressed the problem of the time and effort expended to analyze and extract information, especially the parser of special skills from resumes. They proposed semantical and contextual rich IE using the advanced NLP library SpaCy, which has a feature called "Phrase Matcher.". Their approach identifies a table or dictionary with the various skill sets, parses the resume to search all skill sets, and counts the frequency of those words of different categories. To select the appropriate candidates, they used the Matplotlib tool to represent the information visually after parsing. From 250 resumes and job requirements specified in the dictionary used in the result, they found two candidates' resumes satisfying the job requirements. Note that their research was limited to extracting special skills from the resume, which is only one of several other important parts such as experience, education, etc. Therefore, selecting candidates is done visually, not using ranking algorithms.

In [34], they developed a job portal where employees and applicants can post their resumes, making recruiting easy and efficient. Their portal aids in the organization of resumes according to the needs of specific employers. Also, they used data from social media like Linkedin to help recruit high-quality candidates worldwide while avoiding unfair and discriminatory practices. The result involves converting resumes from /pdf/doc/RTF to plain documents and tokenizing data entities using optical character recognition (OCR), NLP, and ranking algorithm tools. When comparing extracted entities and needed keywords, resumes were ranked based on their technical skills, and the results were supplied in the form of a pie chart and bar graph.

The research work of [35] was to develop a system that automates the eligibility examination and evaluation of candidates in recruiting students for job vacancies or higher education programs. This system handles the manual analysis of resumes with all its tedious tasks and provides accurate and practical evaluation results. They used ML approaches, NLP, and three classification algorithms to implement the system. The evaluation of this work was divided into several steps conducted on personal information, professional experience, academic background, and soft and technical skills. The results obtained in the classification are cross-validated by the results of online video interviews. This work used acceptance criteria for candidates based on the scores given after extracting the attributes, which, as a result, is an indicator in which resumes are classified between acceptable and rejected. Hence, this is what mainly differs from our proposed approach, which systematically ranks the candidates starting with the most matching the job requirements based on relative and comparative scoring instead of merely judging them with binary labels like (acceptable or rejected).

In [36], the authors address the problem of resume analysis and the complex use of the parser when supporting languages other than English. They used ML techniques in the context of NLP that achieved high accuracy in extracting information from resumes in an arbitrary format and five different languages. They created a system with many interconnected models, using state-of-the-art NLP models as a basis. Their approach used a new deep model architecture for sequential input data, called a transformer, with parallel processing of the input sequence in the form of the BERT language model. BERT is designed to learn deep bidirectional representations on unnamed text. Their models extract and categorize relevant resume sections (personal information, skills, education, previous job) and corresponding specific information at the lower hierarchical level (names, dates, addresses, competencies, etc.). Models were evaluated on a data set of 1,686 resumes. For Norwegian, Swedish, Finnish, Polish, and English languages, the system achieves F1-scores of 0.86, 0.88, 0.86, 0.87, and 0.82 at the section level, respectively, and 0.75, 0.80, 0.80, 0.81, and 0.83 at the item level, respectively.

One of the essential parts of a resume is the educational qualifications section, which captures the knowledge and skills relevant to the job. In [11], they solved the problem of a large amount of annotated data required to determine educational institutions' names and grades from a resume's education section. They proposed a semi-supervised model for accurately identifying degrees and insti-

tute names on a resume based on NER. It consists of CNN, word embeddings, and BI-LSTM. This model is used to predict unclassified education section entities and is corrected using a correction unit. This model was evaluated on 500 training and 50 test resumes, achieving an overall F1-score of 73.28 and an accuracy of 92.06%

Higher-level skills can be deduced from lower-level ones, and vice versa, rather than simply pulling out terms associated with competence. Extracting skills is a vital step in developing job recommendation systems. In [37], they proposed using CNN to create an explainable model that can extract high-level skills from resumes in their raw text format. The resulting model can predict the high-level skills mentioned in a resume and highlight the underlying low-level skills that led to that prediction. Experiments were conducted on anonymous IT resumes collected from various websites. The overall model achieves 98.79% recall and 91.34%, and more than 99% accuracy for specific skills.

In [38], they have proposed a two-stage embedding-based recommender system for matching available jobs with suitable candidates. A component for candidate retrieval using fused embedding and a module for fine-tuning and reranking candidates. Deep learning using CNN, representation learning, job-skill information graph, and geolocation calculator are fused for the job and candidate embedding. They have also implemented the Faiss index for clustering and compressing embeddings, which allows conducting runtime nearest neighbor searches. The final ranking score is calculated by a weighted linear equation that aggregates the first-stage relevancy score and contextual job and candidate features. Their job-to-candidate matching system has a satisfying title, description, requirement, and location-matching quality results. Their system has vastly improved the matching quality between jobs and applicants.

Discovering top candidates with few resources in a short time is the most pressing problem facing businesses today. In [39], the best candidates could be ranked using content-based suggestion, which employs NLP techniques to parse resumes. With an average parsing accuracy of 85% and a scoring accuracy of 92%, the system performs effectively. In addition, cosine similarity is used to discover the resumes most relevant to the job description provided, and the k-nearest neighbor (KNN) algorithm is used to select and rank Job applications.

People may have a better chance of finding jobs that will allow them to have a good life if they have the education level appropriate for the professional environment. In support of the impact of education on job applicants, the authors of [40] addressed gaps between the job market, job seeker, and educational institution skills. Recruiters want to check resumes for required skills automatically. Job applicants want to know what skills must be added to their resumes. Educational institutions recommend study programs and assist students in ensuring their courses cover job posting skills. Since all three of these users have skills in common, they made an app called "Skill Scanner," which uses NLP techniques to analyze, vectorize, cluster, and compare skills. It then makes reports with statistics and suggestions for all three users of which skills are covered and missing. They analyzed the master's program in data science and data scientist positions from Indeed.com and Kaggle.com. As part of a questionnaire, 108 representatives from their three parties were provided reports generated by Skill Scanner. Most users report that their solution improves the efficiency, speed, fairness, explainability, autonomy, and support of skill-related processes. They proved that 89% of those who tried out their recommendation system had no problem using it.

Categorizing job applications received as resumes against open vacancies requires a significant amount of time and effort from an employer. This study [41] aims to develop an automated resume classification system (RCS) that classifies resumes according to their job categories. This study's main contribution is in preprocessing the resumes to a corpus and extracting vectorized representation using NLP techniques appropriate for classification tasks performed by ML algorithms. This work was evaluated on three extraction and representation techniques and nine ML classification models. The TF-IDF vectorizer was the best at extracting features and representation, and the SVM classes performed exceptionally well on parsed resume datasets with more than 96% accuracy.

In [42], authors introduce I-Recruiter, an intelligent decision support system (IDSS) for screening resumes and identifying the most qualified applicants for available IT sector vacancies. The semantic similarity between a resume and the job description determines an applicant's ranking in I-Recruiter. It provides information about the best applicants so that the hiring process can move forward. ML and NLP constitute the core of the system's functionalities. The essential components of this system are the training, matching, and extracting phases. The training block produces domain-specific word embeddings that have been trained. At the same time, the matching section identifies top candidates by comparing resumes and job postings for semantic similarity. In the last phase, It extracted some primary data on the top applicants. I-Recruiter showed excellent results, with an average accuracy of 96% and a short amount of working time.

Recently, In [43], they use NLP to extract useful information from resumes. The system will then go to the person's profiles on LinkedIn and GitHub, scrape information from those sites, and feed it into the ML Models to make a more accurate prediction. The first ML model, either KNN or SVM, predicts what kind of job role their resumes are best suited for it. The second model recommends improving their resumes by using cosine similarity, which compares the user's input to the model's prediction. They proved that the accuracy of the models ranges from 78% to 98% depending on the datasets utilized, the learning methods' complexity, and the dataset's size.

Several studies like [44, 45, 46, 47, 48, 49], and [50] suggest using ontology-based knowledge for resume parsing, an ontology is a knowledge base representation that can generate a semantic model of data connected with specific domain knowledge. Furthermore, ontology is utilized to define linkages between various types of semantic knowledge in a domain. Semantic web models use ontology to define the meaning of target data, information, or knowledge. The most fundamental comprehension of the semantic web is surpassed by ontology, which provides the ability for standard reasoning, typically based on the specification of inference rules. Ontologies let users organize information in taxonomies of concepts and their attributes and describe how these concepts relate to each other [51]. The authors of [52] say that there are two main reasons to use ontologies: to help people and software agents understand how information is structured and to make it possible to reuse existing

domain knowledge.

In [44], a resume ontology based on skills was created, which refers to the most necessary parts of a job to see how well the applicant matches the job description. Applicant and job profiles were annotated with a common vocabulary. The semantic concept similarity algorithm was modified to accurately compute and rank matching scores between profiles when a query was run. Based on the results of the system's experiments, the approach improved the search matching accuracy.

The author in [45] built an ontology that contains the most important factors to consider while hiring for IT positions. The primary purpose of the proposed approach is to reduce errors in the first stage of personnel selection by manually filtering hundreds of resumes/profiles to choose applicants for interviews. Their suggested system consists of a mobile application that automatically chooses online profiles from professional websites (like Indeed, LinkedIn, and Monster) and ranks them to eventually display the appropriate candidates for a given vacancy to the recruiter. They created an ontology focused on resume skills to match candidate resumes to job description requirements. GATE and Apache Tika tools were used to automatically extract skills from unstructured resumes in the implementation's first step. The specified information will be used to construct a structured RDF document, saved in a triple store, and queried for each job offer.

A supplement work that used the ontology for resume analysis is presented in the proposed research of [46]. The authors used text mining and ML tools to make an effective company recommender system (CRS) that could help recruiters find the best candidate for a given job title. Because the data to be extracted is unstructured, the authors used traditional and ontology-based information extraction strategies. Candidates for IT organizations were classified into three categories: low, average, and high, based on their rating score, which was interpreted as their competency level in the programming language mentioned in their project description, according to the model proposed. Finally, the organization can use the ranking to select the best possible candidates for the job openings.

The exploration in [47] also sought to address the issue of analyzing unstructured resumes. In addition to the limited availability of language dictionaries and the complexity of Polish linguistic dependencies, this drives their interest. The proposed prototype designed a hybrid resume parser service using text mining algorithms to extract information from Polish resume documents for IT recruitment. The resume parsing system combines three NER tools (Liner2, Nerf, Babelfy) with the anchor NER service and dictionary methods such as Fuzzy Dict and Competence Lexem. The proposed system also included a promising CRF method. Data from an IT recruiting firm was used to conduct the research, and the results showed that the hybrid approach proposed improves single results obtained for detecting educational institutions. The result of the combined tools (hybrid approach) was more than 60% better than the best single solution (30.92% vs. 19.23% for Babelfy).

The authors in [48] have considered the effects of ontology-based use in the resume search system for job applicants. They provided an ontology structure for representing resume and job description (JD) contents, extracting information from collected resumes and JD using OCR, and classifying extracted information using an unsupervised approach called computer science ontology (CSO) classifier. This ontology searches resumes for job skills that match JD criteria. The challenge of candidate ranking in an automated recruitment system is solved using ML and NLP. This solution is based on an IT ontology called Job-Onto. Moreover, the planned system is implemented as a recruiting website that helps candidates and employers find opportunities and streamline their recruitment processes. The suggested recruitment system uses automated resume ranking based on criteria, with the recruiter controlling the weighting. Each time a job is posted, the system calculates the resume correlation score for each candidate. This score is the sum of their domain skill, general skill, and soft skill matching scores. Finally, a list of the best resumes for the front-end engineer job is shown. The proposed method proved effective in finding resumes that correspond with a particular JD in the IT industry, according to test results.

In [49], the authors proposed an approach to improve selecting the best candidates, including three primary processing steps: text extraction, text block classification, and resume facts identification via NER. A custom-built ontology is used to augment the extraction of technical skills. This research is unique because it focuses on many areas, including diversity and improving electronic recruitment re-usability. However, there are no numerical findings for the accuracy they intended to improve in extracting information from personal resumes.

Every job has distinct requirements: some require the most experience, education, or skills. Finding the proper applicants in a large resume pool is difficult, especially when we are looking for specific skills. So, in [50], their goal is to create a system to analyze resumes and job postings based on how well they match one another. Then, rank them in ascending order of score. They used ruled-based methods, developed their own ontology to facilitate matching and scoring, and employed NLP methods for information extraction. Their ontology contains 37493 distinct nodes, and 54632 relations were created. They did real case studies to evaluate the AI scoring and other features. For actual job vacancies, recruiters and the system simultaneously searched for candidates. The system's top candidates were ranked.

Finally, we highlighted a study that focused on sentiment analysis regarding personal interviews with candidates and used the SVM ranking algorithm that we will apply in our research. All users on social media can write their sentiments to express their emotions and opinions. Three processes comprise sentiment analysis: sentiment extraction, keyword preparation, review analysis, and classification. Moreover, Covid-19 has shifted all processes online, such as video conferencing. This study in [53] provides a job candidate rank (JCR) model for job candidates based on the interviewer's sentiment analysis. The candidate rank model has stages—first, sentiment text parsing, stopword removal, and stemming. Second, select features using document frequency (DF), expand using the WordNet database, and focus groups on ranking them. SVM and NB of ML classifications are included. Using the NB classifier, the model achieved 93% accuracy, while the SVM classifier achieved 89% accuracy.

6.2 Comparative Description Approach

This research was motivated by the successful use of comparative descriptions in other applications; therefore, we aim to apply and extend such effective capability as a novel contribution to resume ranking and selection processes for effective candidate-career matching. The ability of comparative descriptions regarding resume analysis has yet to be extensively investigated. Thus, we widen the exploration to embrace other prominent applications utilizing comparative descriptions, such as human identification by semantic/soft biometrics.

One of the challenges of human identification and semantic characterization is the quality of soft biometric description [19]. A search in soft biometrics showed better accuracy of comparative traits or attributes than the accuracy of categorical traits [4, 15, 16, 19].

In [16], the authors compared subject comparisons in the Soton gait database by comparing one target to multiple subjects. When comparing one target to more than one subject, they found that the comparative descriptions performed 17% better than the absolute/categorical descriptions.

Also, in [4], the authors used semantic clothing traits as soft biometrics for human identification. They further explored their validity and efficiency through corresponding comparative descriptions, allowing for more accurate differentiation.

The study in [15] analyzed a data set using gender as a comparative attribute and found that comparative annotations are more discriminatory than categorical labels. The study's approach on 100 annotated subject images showed correct-match reliability in the top 7% with ten comparisons or the top 13% with only five comparisons.

Additionally, in [19], the authors studied human identity through comparative facial soft biometrics using the labeled faces in the wild (LFW) dataset. Comparative soft biometrics allows each person to be uniquely identified in the database by creating a biosignature with their exact physical traits compared to others. Such comparative descriptions improved searches based on a given comparative trait (e.g., searching for someone younger).

The last study [54] focused on enhanced human-machine communication and demonstrated the benefits of relative attributes on four applications. The applications include active learning of discriminative classifiers, zero-shot learning from relative comparisons, automatic image description, and image search with interactive feedback. Relative attributes compare an image's attribute strength to others rather than predicting its presence. Relative attributes would be more natural and allow for richer communication, more detailed human supervision, higher recognition accuracy, and more informative descriptions of new images. The SVM ranking function is learned and used for each attribute to predict the attribute strength. They demonstrate that relative attributes lead to higher performance across all applications by comparing them to numerous strong baselines using image datasets of scenes and faces.

All previous research works aim to bridge the semantic gap between biometrics and human description by using relative and comparative descriptions of attributes. In particular, to the best of our knowledge, no study has used comparative descriptions for semantic attribute analysis of resume documents for improving job-candidate matching.

Table 1 presents our proposed approach compared with several previous studies.

7 Methodology

As in the literature, there is a need to improve the process of automatically extracting data from resumes and mining their information [55]. The improvement can assist in selecting candidates and ranking them according to the job's requirements. Recruiters may have subjective ideas about which aspects they want to emphasize, depending on the circumstances and nature of their company's requirements at that particular time. Therefore, in this research, we proposed a novel approach to improve the candidate-career matching process based on comparative semantic resume analysis. Our approach comprises three phases: the semantic attribute extraction phase using NLP techniques, the relative and comparative labeling phase, and the ranking phase using the ranking SVM algorithm. The proposed model can generally analyze the resume by finding the exact keywords for each resume section, which will be used to assign a suitable label for 13 semantic attributes each. Thus, our proposed model lets the recruiters modify the attribute-based searching as they want to depend on their own requirements. After constructing a table or dictionary, which covers various semantic sets of attributes, count the occurrences of the words belonging to different semantic attributes and assign scores for each attribute based on evaluation metrics, which are dynamic to change and determined by recruitment staff based on the job requirements. Then, aggregate all points for each attribute as regular scores and show the related normalization score. After that, we come to the comparative description part, to which we contribute to enhancing the resume analysis domain by inferring a suited relative descriptive label per attribute for each resume based on evaluating their given scores. Then, inferring corresponding comparative descriptive labels by describing a pairwise comparison per attribute for all of the attributes between each resume and all other resumes in the dataset. Finally, ranking all resumes per attribute using the Rank SVM algorithm through extracting usable relative measurements from those comparative descriptions as resume scoring per attribute and, consequently, leading to overall resume scoring concerning all sets or a selected subset of attributes in descending order. This method helps HR query the list of candidates ranked based on each semantic attribute in the resume or multiple semantic attributes.

Our study aims to provide an ML-based model that does not solely rely on training data to match candidate resumes to career requirements. However, it also enables dynamic learning and ranking per any combination of the proposed resume semantic attributes. We summarize all of the methodology phases and present the flowchart of the proposed approach of semantic resume analysis in the next subsection.

Table 1: Comparing our proposed approach with previous studies

Study	ML	NLP	Ontology	Scoring	Ranking	Comparative	Language
[10]	√	√	-	√	-	-	English
[11]	√	√	-	-	-	-	English
[20]	-	√	-	-	-	-	English
[21]	√	√	-	-	-	-	Vietnamese
[22]	√	√	-	√	√	-	English
[23]	√	√	-	-	-	-	English
[24]	√	√	-	√	√	-	English
[25]	√	√	-	-	-	-	English
[26]	√	√	-	√	√	-	English
[27]	√	-	-	-	-	-	English
[28]	√	√	-	-	-	-	English
[29]	√	√	-	-	-	-	English
[30]	√	√	-	√	√	-	English
[31]	√	√	-	√	√	-	English
[32]	√	√	-	-	-	-	English
[33]	√	√	-	√	√	-	English
[34]	√	√	-	√	√	-	English
[35]	√	√	-	√	-	-	English
[36]	√	√	-	-	-	-	English, Norwegian, Swedish, Finnish, Polish
[37]	√	√	-	-	-	-	English
[38]	√	√	-	√	√	-	English
[39]	√	√	-	√	√	-	English
[40]	√	√	-	-	-	-	English
[41]	√	√	-	-	-	-	English
[42]	√	√	-	√	√	-	English
[43]	√	√	-	√	√	-	English
[44]	√	-	√	√	√	-	English
[45]	√	-	√	√	√	-	English
[46]	√	√	√	√	√	-	English
[47]	√	√	√	-	-	-	Polish
[48]	√	√	√	√	√	-	English
[49]	√	√	√	√	√	-	English
[50]	√	√	√	√	√	-	English
[53]	√	√	-	√	√	-	English
Our work	√	√	**-**	√	√	√	**English**

7.1 Summary of Proposed Approach

Our approach is proposed to assist HR officers in improving the extraction, filtering, and selection of the most job-suited resumes in the technical domain (or specialties) in an automated and more accurate manner based on analyzing semantic resume attributes. For each resume, two groups of per-attribute semantic descriptions are used as relative and comparative labels, which are more precise and informative than categorical descriptions. The following scenario will be followed in this research study:

1. Construct a table or dictionary that covers various semantic sets of attributes with each keyword in the resume, such as (personal information, education level, technical skills level, professional experience, personal skills, additional qualification, etc.).

2. An NLP-based tool is used to parse and extract information from the entire resume to search the words in the table or dictionary.

3. Count the occurrences of the words belonging to different semantic attributes and assign scores for each attribute based on evaluation metrics determined by recruitment staff based on the job requirements.

4. Aggregates all points for each attribute as regular scores and shows the related normalization score.

5. Inferring a suited relative descriptive label per attribute for each resume based on evaluating their given scores.

6. Inferring corresponding comparative descriptive labels by describing a pairwise comparison per attribute for all of the attributes between each resume and all other resumes in the dataset.

7. Ranking all resumes per attribute using the SVM ranking algorithm through extracting usable relative measurements from those comparative descriptions as resume scoring per attribute and, consequently, leading to overall resume scoring concerning all sets or a selected subset of attributes in descending order.

8. Analyzing results based on four comparison aspects which compare the efficiency in distribution and discrimination of resumes rankings.

Figure 2 overviews the proposed semantic resume analysis approach, and we will explain each phase in detail in the following subsections.

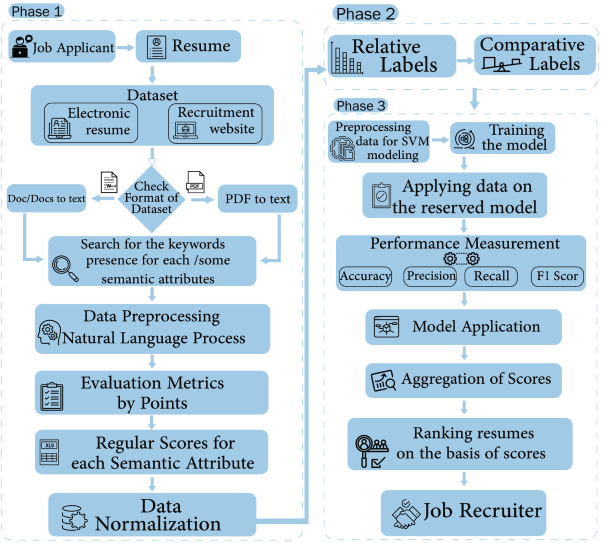


Figure 2: Overview of the proposed semantic resume analysis approach

7.2 Semantic Information Extraction Phase

An automated system for extracting information and selecting the potential candidates who best meet the position's requirements can significantly improve the efficiency of HR agencies [56]. However, resumes are unstructured documents with different file formats (pdf, txt, doc, etc.) and contain ambiguous and variable language. This heterogeneity makes it challenging to extract useful information.
To make the recruitment process simple, efficient, and automated, some advanced studies, like [10] and [33], solved the parsing problem with specific searches and unstructured data. They searched the whole document for words in predefined tables, dictionaries, or a BoW to list important words and count the number of detected words of interest. Therefore, we proposed a candidate selection system based on searching the keywords. If the keywords exist, it will count their presence and allocate the given points as evaluation metrics or as HR-specified job requirements.
Before we move to define the proposed semantic attributes, we will overview the NLP techniques in general and the related techniques we used in our approach:

7.2.1 NLP with Python

By "natural language," we refer to a human language used for everyday conversation, such as English, Arabic, Hindi, and Portuguese. In contrast to programming and formal languages such as computer and mathematical notations, natural languages have evolved as they are passed from generation to generation and are challenging to define with specific rules. NLP aims to "understand" complete human sentences so that appropriate replies can be provided. More and more technologies are appearing based mainly on natural language processing. In today's increasingly globalized and interconnected information world, language processing has assumed critical importance [17].

7.2.2 Proposed Semantic Attributes

Table 2 demonstrates the proposed semantic attributes that will be extracted from resumes and all relevant keywords and suggested evaluation matrices that will be defined and used in our approach.

7.2.3 Keywords Detection or Information Extraction

- Extracting files from the archive directory of resumes dataset and generating a list using ***extractall()*** python function.

- Checking the format of the input file from the dataset (pdf, doc/docx) and converting them to text using ***str()*** python function.

- Extracting semantic attribute information. We construct three functions (a function to check keywords, a function to check keywords and count their presence, and a function to check how many keywords are present). First, the personal information, name, email, and mobile number are parsed separately. Then, allocate the weights that the requirements illustrated using a simple resume parser existing in [57], which is used for extracting information from resumes by using the NLTK and SpaCy models of NLP. Since resumes have rather varied structures and formats, the keywords are similarly searched for the ten remaining semantic attributes. If the keywords exist, then the given points of evaluation metrics would be allocated. Aggregating all points for each attribute as regular scores and showing the related normalization score.

7.2.4 Data Normalization

For many procedures to work properly, the data must be normalized or standardized; since resumes have unbalanced scores after extracting attributes, we used normalization (MinMaxScaler) to transform features (the regular scores of the attributes) by scaling each attribute to a corresponding range (between zero and one). Then, we used the normalization score to get the relative scores, which range from 0 to 5, as explained in the next section.
The transformation imported using Scikit-learn of ML library is given by:

$$X_std = \frac{(X - X.min(axis=0))}{(X.max(axis=0) - X.min(axis=0))}$$
$$X_scaled = X_std * (max - min) + min \quad (1)$$

where $min, max = feature_range$.
Table 3 shows an example of evaluation metrics by scores given for the personal information attribute, where the total is eight points for the whole personal information attribute and shows the related normalization score, which is calculated by "(1)" where $feature_range$=(0,8), X=8, min=1, and max=8.

Table 2: Proposed resume semantic attributes to be extracted and analyzed

Semantic attributes	Keywords
1- Personal information completeness	Name, Nationality, Mobile No., Email, Address, Phone, Contact No, Country, ...etc.
2- Education level	GPA, Bachelor, Master, PhD.
3- Technical skills level	Programming Languages, Tools, Framework, Databases, Operating systems, ...etc.
4- Professional experience	Experience, Participated, Worked, ...etc.
5- Personal skills (Soft skills)	Communication, Teamwork, Time Management, Problem-Solving, Creativity, Leadership, ...etc.
6- Awards	Awarded, Winner, ...etc.
7-Hobbies and interest	Play, Games, Surf, News, Design, Read, Video, Sports, ...etc.
8- Additional Qualification	Course, Program, Trained, Level, ...etc.
9- Professional Certificate	Certified, Cisco, Scrum, Professional, Management, Administrator, ...etc.
10- Suitable age requirement	Age, Years old, Birth, Birth year.
11- Career objective	Gain, Objective, Opportunity, Looking for, Looking forward, ...etc.
12- Project Experience	Project, Role, Client, Responsibilities, Environment.
13- Languages	Arabic, English, ILETS, STEP, ...etc.

Table 3: Example of evaluation metrics by scores, regular score, and normalization score for the personal information attribute

Personal information	Score
Name	2
Nationality	1
Address	1
Mobile No.	2
Email	2
Total (Regular score)	8
Normalized score	1

7.3 Relative and Comparative Labeling Phase

After we extract the important semantic attributes from the resume, we come to the comparative description part, to which we contribute to improving the resume analysis domain. In our work, apart from simple traditional evaluation metrics (a regular scoring system), we use the proposed comparative and relative labels to score or rate a resume more precisely.

After normalization, the points are totaled for each attribute in the regular scoring system to infer the relative scores. We multiply each point by five and then round it down. The labels of relative descriptions are defined as "Very high," "High," "Average," "Low," "Very low," and "None." Table 4 shows an example of relative labels and their corresponding scores.

Table 4: Relative labels and corresponding scores

Relative score system	Relative labels
5	Very high
4	High
3	Average
2	Low
1	Very low
0	None

After assigning the most appropriate scores based on relative labels for each semantic attribute in the resume, we will accordingly use comparative labels to describe a pairwise comparison per attribute for all of the attributes between each resume and all other resumes in the dataset (one subject with all others). The comparative labels include "Much higher," "Higher," "Same," "Lower," and "Much lower." For example, when the relative score of personal information is 5 for resume A and 4 for resume B, then the comparative label for resume A will be labeled as "Higher" than resume B, with the new corresponding score equal to 1, as listed in Table 5, which shows an example of comparative labels and their corresponding scores.

Table 5: Comparative labels and corresponding scores

Comparative score system	Comparative labels
2	Much higher
1	Higher
0	Same
-1	Lower
-2	Much lower

We put the comparative scores as follows: the score will be "1" or "-1" if the difference between the relative scores of two resumes is 1 or 2. The score will be "2" or "-2" if the difference between the relative scores of the two resumes is more than 2, and the score will be "0" if the relative scores of the two resumes are the same. As a result, resume A will be compared to each of the other resumes, one by one, per attribute.

Note that when we compare resumes A and B, we do not need the opposite comparative label comparing B and A because we only compare the possible pairwise combinations, not the permutations. The combination relations refer to the combination of elements n taken k-combination at a time without repetition, and the order of elements selection does not matter here. Therefore, a k-combination of the set S is hence a subset of k different elements from S. If there are n elements in the set, the number of k-combinations is equal to:

$$\frac{n!}{k!\,(n-k)!} \tag{2}$$

whenever $k \leq n$, and which is zero when $k > n$

7.4 Ranking SVM Phase

Using the ranking SVM algorithm, the final resume ranking per semantic attribute will be based on comparative labels concerning the other N-1 resumes. Rank SVM is a technique for sorting lists of objects using pairwise difference vectors to adaptively arrange given comparable peers based on a certain criterion. Rank SVM performs rankings using Standard SVM. The ranking SVM's objective is to sort the list of resumes for each attribute, and we can infer the ordering list of resumes for all attributes. This ranking will enable further search capabilities based on how relevant the retrieved objects are to a particular search query.

In our work, we used **(train_test_split)** function of sklearn that split input resumes and values data into train and test dataset according to **test_size** ratio. The **test_size** is equal to 0.33 mean, then 33% is test data, and 67% is train data. The output is **resume_train, resume_test**, which is **(x_train x_test)**) and **(y_train, y_test)**. All resume comparisons are placed as pairwise inputs into the SVM ranking algorithm for the training, with each attribute labeled. We used hyper-parameter optimization in SVM for both C and Gamma values to get their optimized values. Then, pass training data to SVM to predict the comparative labels against each attribute and resume pairs by using the three following functions [58, 59]:

- **svm = SVC(kernel = 'rbf', c=10, gamma=0.1)**
- **svm.fit(x_train, y_train)**
- **predict_eachAttribute = svm.predict(x_test)**.

Rank resumes based on predicted ranks so that in each iteration, the resume with the highest rank is selected and removed for the next iteration. Ultimately, we will get the ranks of resumes based on scores against individual attributes. Finally, the output will be the corresponding usable relative measurements and the ordering list based on comparisons of all resumes.

This method helps HR query the list of candidates ranked based on single or multiple semantic attributes in their resumes. As such, we can apply a search to retrieve the resume that is "Much higher" than the others in a certain attribute. To rearrange the resumes based on the resulting rankings, we can count (sum) their nascent relative measurement scores based on multiple or all attributes for each resume. The ranking is given in descending order to present the resumes from the best to the worst. A soft-margin ranking SVM method is used for a given set of attributes A to learn a ranking linear function r_a for each attribute, similar to the way used in [4]:

$$r_a(x_i) = w_i^T x_i \tag{3}$$

Where w_a is the coefficient of the ranking function r_a and x_i is a feature vector of attributes of a resume being ranked. Rearranging a set of comparisons into two groups can be considered a representation of the pairwise relative constraints needed to learn a ranking function. The first group is a set of dissimilarity comparisons D_a of ordered pairs so that $(i, j)\epsilon D_a \Rightarrow i > j$. The second group is a set of similarity comparisons S_a of non-ordered pairs so that $(i, j)\epsilon S_a \Rightarrow i = j$. Then, the following formula is used to get the w_a coefficients of r_a from the D_a and S_a sets:

$$\begin{aligned} &\text{minimize} \quad \left(\tfrac{1}{2}\left\|w_a^t\right\|^2 + C\sum \xi_{ij}^2\right) \\ &\text{subject to} \quad w_a^T(x_i - x_j) \geq 1 - \xi_{ij}; \quad \forall (i,j)\epsilon D_a \\ &\qquad\qquad \left| w_a^T(x_i - x_j) \right| \leq \xi_{ij}; \quad \forall (i,j)\epsilon S_a \\ &\qquad\qquad \xi_{ij} \geq 0 \end{aligned} \tag{4}$$

ξ_{ij} is the misclassification bias, and C is the trade-off between maximization of margin and minimization error. The resulting optimal w_a function can then be used to (explicitly) rank all training samples according to a. "Equation(3) is used to map a feature vector x_i to a feature vector consisting of several real-value relative measurements."

7.5 Performance evaluation

For SVM performance evaluation, we assess the accuracy of the proposed model, along with a classification report that quantifies the quality of the predictions of each semantic attribute.

A classification report assesses the quality of a classification algorithm's predictions. How many predictions are correct, and how many are incorrect? True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN), to be exacted, as in Figure 3 [60].

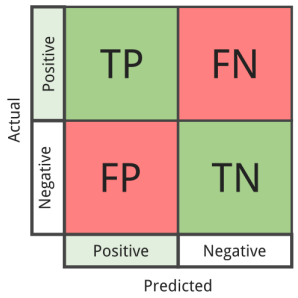


Figure 3: Simple representation of confusion matrix [60]

The input of the classification report is **(X, Y)** where X is predicted values and Y is actual values. We used **(classification_report)** function of sklearn library to print precision, recall, and F1-score. Similarly, used **(accuracy_score)** function of sklearn library to print the accuracy of the SVM model. Whereas the accuracy is the proportion of accurate predictions to the total number of data samples [61]. The precision shows how many predictions for a certain class are of the same class. The recall shows how many of certain class predictions were right, and the F1-score is the geometric mean of precision and recall [62]. The numeric implementation of the classification report is proven in the next section of the experiment results.

7.6 Practical Example of the Three Proposed Phases

The graph in Figure 4 shows our approach example with the following information:

-Inputs: 3 resumes (resume A, resume B, and resume C) with semantic education attributes extracted from resumes.

-Processing: Describing the relative and comparative labels with

corresponding scores for comparing between resumes.
-Output: Extract usable relative measurements in which these values are set and adjusted based on comparing each resume with the other resumes. Then, get the ordering list of all resumes using the SVM ranking algorithm.

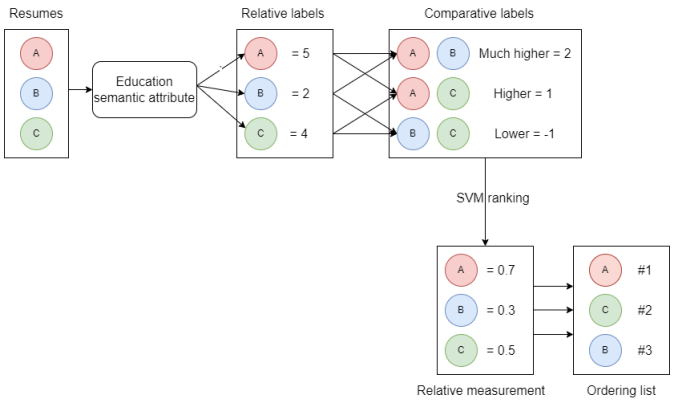


Figure 4: A simple example clarifies the main three phases of our proposed approach

8 Experiments and Results

8.1 Dataset

Our proposed approach's experiment was conducted on a dataset containing 228 resumes in various text file formats, such as doc, docx, and pdf. We collected resumes from various technical (computer-related) specialists from various sources, including Github, Kaggle, LinkedIn, etc. Most resumes contain information about technical skills, such as programming in Python or Java, and other technical experience. We used the resume dataset for analysis, and each resume was assigned relative and comparative labels.

8.2 Implementation of three proposed phases

After the three phases of our approach, we generated 25,878 resume comparisons between resumes for each attribute. To select the candidate who matches the job requirements, we aggregated the scores and ranked the resumes according to each attribute and all attributes, as detailed in the methodology chapter. The resumes were ranked in descending order of total score, with the highest resume coming first and the lowest coming last. Next, we analyzed the results using four comparison-related aspects:

1. Ranking based on the regular scores. Figure 5 shows the histogram distribution of resume rankings for all attributes based on the regular scores.

2. Ranking based on the comparative scores. Figure 6 shows the histogram distribution of resume rankings based on the comparative scores for all attributes.

3. Ranking based on the comparative scores and regular scores. Figure 7 shows the histogram distribution of resume rankings for all attributes based on the comparative and regular scores.

4. Ranking based on the comparative scores and relative scores. Figure 8 shows the histogram distribution of resume rankings for all attributes based on the comparative and relative scores.

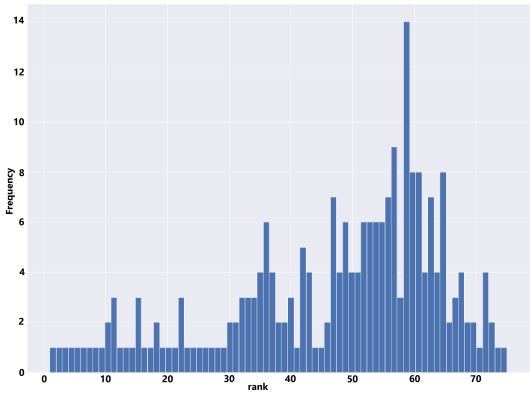


Figure 5: Resume rankings for all attributes based on the regular scores

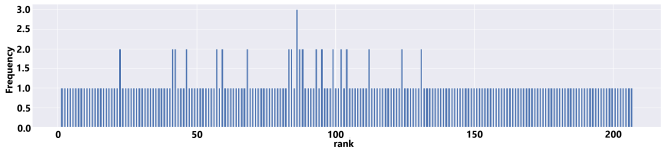


Figure 6: Resume rankings for all attributes based on the comparative scores

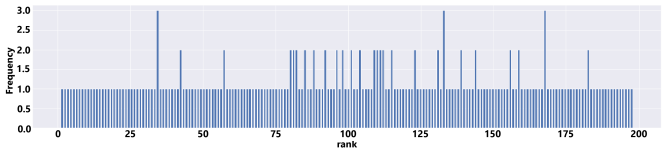


Figure 7: Resume rankings for all attributes based on the comparative scores and regular scores

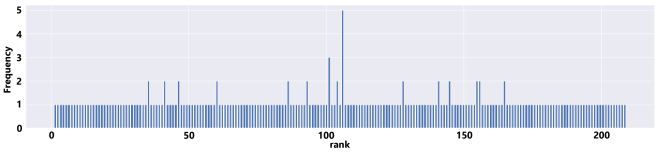


Figure 8: Resume rankings for all attributes based on the comparative scores and relative scores

8.3 Analysis of Results

Table 6 summarizes the four ranking results. The table compares the effectiveness of distributing and discriminating in the rankings of resumes with high unique values. It is observed that all three rankings based on comparative scores outperform the ranking based on regular scores, which offers very weak resume differentiation with high (undesired) redundant values and highly (confusable) similarities in resumes' scores. It is important to observe that when it comes to resume distributions (as indicated in Table 6, Figure

6, Figure 8, and Figure 7), the more unique scores, the better the ranking result.

Table 6: Results of rankings with different four basis

Basis of rankings	Max Rank	No. of resumes that has unique scores	No. of resumes that has redundant scores
Regular scores	75	30	198
Comparative scores	207	187	41
Comparative and Regular	198	171	57
Comparative and Relative	**209**	**194**	**34**

Let us demonstrate the theoretical distribution of the duplicated rankings between resumes on each basis of rankings. Figure 9 shows the histogram distribution of the duplicated rankings between resumes for all attributes based on the regular scores. Figure 10 shows the histogram distribution of the duplicated rankings between resumes for all attributes based on the comparative scores. Figure 11 shows the histogram distribution of the duplicated rankings between resumes for all attributes based on the comparative and regular scores. Figure 12 shows the histogram distribution of the duplicated rankings between resumes for all attributes based on the comparative and relative scores.
As we proved in Table 6, the distribution of the duplicated rankings between resumes for all attributes based on the comparative and relative scores outperforms all other ranking bases.

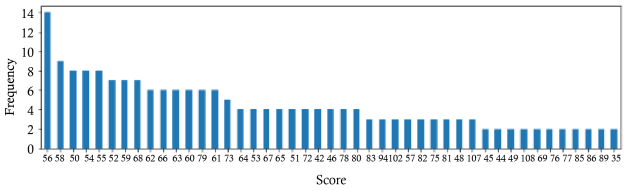


Figure 9: Duplicated rankings between resumes for all attributes based on the regular scores.

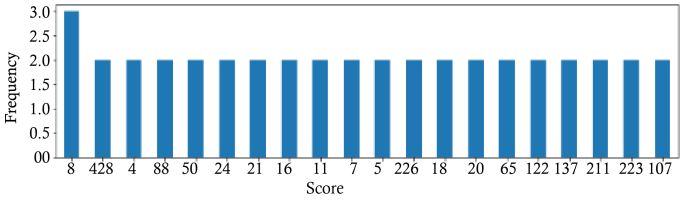


Figure 10: Duplicated rankings between resumes for all attributes based on the comparative scores.

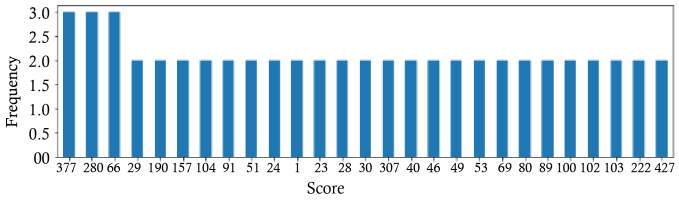


Figure 11: Duplicated rankings between resumes for all attributes based on the comparative and regular scores.

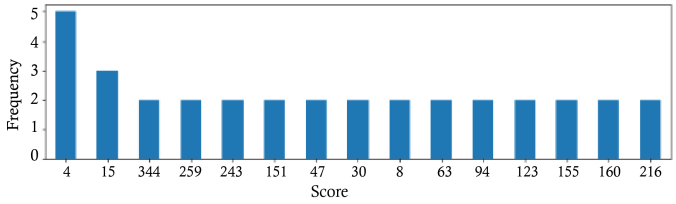


Figure 12: Duplicated rankings between resumes for all attributes based on the comparative and relative scores.

We proved that the method of ranking based on comparative and relative scores is very effective by comparing its accuracy with all other basis rankings as shown in Table 7. The ranking quality can further be measured as the accuracy of differentiating (as much as possible) between N of compared resumes. Resulting in as many as possible different unique scores S, which is, in the best-case scenario, equal to the number of compared resumes (i.e., $S = N$), implies that each resume has a unique (distinct) score. This approach is made by computing the percentage of the total sum of all unique and redundant scores concerning their frequencies using the following formula:

$$Accuracy = \frac{1}{N}\left(\sum_{i}^{S} \frac{1}{f_i}\right) * 100 \quad (5)$$

Where N is the number of resumes to be ranked. S is the total number of unique scores assigned to N number of resumes to be ranked, and f_i is the corresponding ith frequency of the ith unique score.

Table 7: Accuracy of rankings with different four basis

Basis of rankings	Accuracy
Regular scores	33%
Comparative scores	90%
Comparative and Regular scores	87%
Comparative and Relative scores	**92%**

As such, it is seen that all methods that rank resumes using comparative scores outperform those that rely on regular scores for ranking. The highest accuracy in ranking the resumes is achieved using ranks derived from comparative scores with the relative scores as the basis. This approach can be applied to each attribute.
Figure 13 shows the rankings' histogram of the resumes for the personal information attribute, taking the comparative and relative scores as the basis.

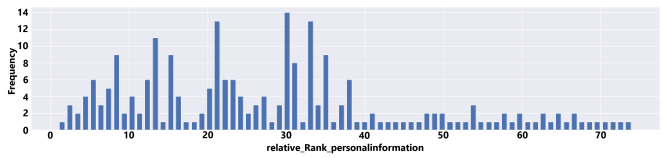


Figure 13: Resume rankings for personal information attribute based on the comparative scores and relative scores

Finally, Figure 14 depicts the top ten resumes' (candidates) rankings with totaling scores on a comparative and relative basis.

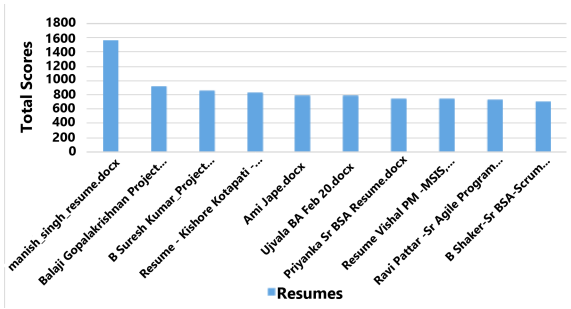


Figure 14: Top ten resume' rankings

8.4 *Evaluation of Resume Ranking per Attribute*

Table 8 reports the accuracy [61] and classification report that contains a precision [63], recall [64], and F1 score [65] of the SVM ranking algorithm which quantifies the quality of the predictions of each of all semantic attributes.

Table 8: Evaluation metrics of the SVM ranking algorithm for each of all semantic attributes

Semantic Attribute	Accuracy	Precision	Recall	F1-Score
1- Personal information	99.168%	0.97751	0.97162	0.97456
2- Education level	99.496%	0.98852	0.99625	0.99238
3- Technical skills	99.461%	0.99106	0.98162	0.98634
4- Professional experience	99.754%	0.99516	0.99756	0.99636
5- Personal skills	99.841%	0.98810	0.99747	0.99278
6- Awards	99.660%	0.98212	0.99108	0.98660
7- Hobbies and interest	99.332%	0.98624	0.96987	0.97805
8- Additional Qualification	99.601%	0.98980	0.96697	0.97838
9- Professional Certificate	99.344%	0.97351	0.99453	0.98402
10- Suitable age requirement	99.941%	0.99099	0.99979	0.99539
11- Career objective	99.976%	0.99984	0.99963	0.99973
12- Project Experience	99.742%	0.98223	0.99774	0.98998
13- Languages	99.964%	0.99964	0.99975	0.99969
Average score	**99.652%**	**0.98805**	**0.98952**	**0.98878**

Through the SVM algorithm performance table, we note that the career objective attribute is higher than other attributes in accuracy, precision, and F1-score, and the suitable age requirement attribute is higher than other attributes in the recall. In general, the average values of all evaluation metrics represent high performance up to 99%. The table results prove the effectiveness of using an SVM algorithm to rank resumes for each semantic attribute separately.

9 Conclusions

This study presents a model to improve the candidate ranking and selection process for career matching. The model consists of three phases: firstly, the extraction of semantic information for 13 proposed attributes using NLP tools; secondly, inferred of relative and comparative descriptions by comparing every two resumes for each attribute; and finally, the use of the ranking SVM algorithm to order all resumes for each attribute. Implementing a comparison description in the candidates' selection process has been suggested and proven more accurate and informative. It helps to distinguish the tiny differences between resumes. The experiment results were analyzed using four comparison aspects: ranking based on regular scores, comparative scores, comparative and regular scores, and comparative and relative scores. The histogram results prove that all three ranking methods based on comparative scores with high unique rankings values outperform the ranking using mere regular scores, which offers very weak resume differentiation with high redundant rankings values and high similarities in resume scores. Also, the ranking attained by the comparative scores with the relative scores as the basis achieves the highest accuracy of 92%. We reported evaluation metrics of the SVM ranking algorithm with an average accuracy of 99.652%, an average precision of 0.98805, an average recall of 0.98952, and an average F1-score of 0.98878, which quantifies the quality of the predictions of each of all semantic attributes. Finally, we have obtained a list of the top ten candidates out of 228 technical specialists' resumes.

10 Future Work

In the future, a proposed model should consider keyword matching for 13 attributes and focus on the semantic relations and meanings of the resume language to account for variations in resume authoring. We will explore additional parts that individuals may incorporate into their resumes to enhance the evaluation and comparison process for more effective candidate-career matching. Furthermore, we will explore the possibility of applying the proposed methods to many domains beyond the technological field and incorporating languages other than English, such as Arabic. Finally, we will evolve our work to analyze the job description document to calculate its matching with the resume as another way different from calculating evaluation metrics of job requirements.

Acknowledgment This work is selected as a special issue of a conference paper published in the 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN) to be extended into a journal paper.

References

[1] A. H. Alderham, E. S. Jaha, "Comparative Semantic Resume Analysis for Improving Candidate-Career Matching," in 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN), 313–321, IEEE, 2022, doi:10.1109/CICN56167.2022.10008255.

[2] Y. Lin, H. Lei, P. C. Addo, X. Li, "Machine learned resume-job matching solution," arXiv preprint arXiv:1607.07657, 2016.

[3] C. Saxena, "Enhancing productivity of recruitment process using data mining & text mining tools," 2011.

[4] E. S. Jaha, M. S. Nixon, "Soft biometrics for subject identification using clothing attributes," in IEEE international joint conference on biometrics, 1–6, IEEE, 2014, doi:10.1109/BTAS.2014.6996278.

[5] D. Lee, M. Kim, I. Na, "Artificial intelligence based career matching," Journal of Intelligent & Fuzzy Systems, **35**(6), 6061–6070, 2018, doi:10.3233/JIFS-169846.

[6] L. Columbus, "Remote Recruiting In A Post COVID-19 World," https://www.forbes.com/sites/louiscolumbus/2020/03/30/remote-recruiting-in-a-post-covid-19-world/?sh=7c4ce7e599fd, 2020.

[7] C. Russell, Right Person–Right Job: Guess Or Know: the Breakthrough Technologies of Performance Information, Human Resource Development, 2002.

[8] A. Swenson, "The new art of hiring smart: Matching the right person to the right job," Technical report, 2000.

[9] A. K. Sinha, M. A. K. Akhtar, A. Kumar, "Resume Screening Using Natural Language Processing and Machine Learning: A Systematic Review," Machine Learning and Information Processing: Proceedings of ICMLIP 2020, **1311**, 207, 2021.

[10] Y. Suresh, A. M. Reddy, et al., "A Contextual Model for Information Extraction in Resume Analytics Using NLP's Spacy," in Inventive Computation and Information Technologies, 395–404, Springer, 2021, doi:10.1007/978-981-33-4305-4_30.

[11] B. Gaur, G. S. Saluja, H. B. Sivakumar, S. Singh, "Semi-supervised deep learning based named entity recognition model to parse education section of resumes," Neural Computing and Applications, **33**(11), 5705–5718, 2021, doi:10.1007/s00521-020-05351-2.

[12] A. Kulkarni, A. Shivananda, "Advanced Natural Language Processing," in Natural Language Processing Recipes, 97–128, Springer, 2019.

[13] H. Assal, J. Seng, F. Kurfess, E. Schwarz, K. Pohl, "Semantically-enhanced information extraction," in 2011 Aerospace Conference, 1–14, IEEE, 2011.

[14] S. Singh, "Natural language processing for information extraction," arXiv preprint arXiv:1807.02383, 2018.

[15] D. Martinho-Corbishley, M. S. Nixon, J. N. Carter, "Analysing comparative soft biometrics from crowdsourced annotations," IET Biometrics, **5**(4), 276–283, 2016, doi:10.1049/iet-bmt.2015.0118.

[16] D. A. Reid, M. S. Nixon, S. V. Stevenage, "Soft biometrics; human identification using comparative descriptions," IEEE Transactions on pattern analysis and machine intelligence, **36**(6), 1216–1228, 2013, doi:10.1109/TPAMI.2013.219.

[17] S. Bird, E. Klein, E. Loper, Natural language processing with Python: analyzing text with the natural language toolkit, " O'Reilly Media, Inc.", 2009.

[18] D. Parikh, K. Grauman, "Relative attributes," in 2011 International Conference on Computer Vision, 503–510, IEEE, 2011.

[19] N. Y. Almudhahka, M. S. Nixon, J. S. Hare, "Comparative face soft biometrics for human identification," in Surveillance in Action, 25–50, Springer, 2018, doi:10.1007/978-3-319-68533-5_2.

[20] J. Chen, C. Zhang, Z. Niu, "A two-step resume information extraction algorithm," Mathematical Problems in Engineering, **2018**, 2018, doi:10.1155/2018/5761287.

[21] L. Pham Van, S. Vu Ngoc, V. Nguyen Van, "Study of Information Extraction in Resume," Conference, 2018.

[22] R. Nimbekar, Y. Patil, R. Prabhu, S. Mulla, "Automated Resume Evaluation System using NLP," in 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), 1–4, IEEE, 2019, doi:10.1109/ICAC347590.2019.9036842.

[23] Y. O. German, O. V. German, S. Nasr, "Information extraction method from a resume (CV)," 2019.

[24] S. Amin, N. Jayakar, M. Kiruthika, A. Gurjar, "Best Fit Resume Predictor," International Research Journal of Engineering and Technology, **6**(8), 813–820, 2019.

[25] S. T. Gopalakrishna, V. Vijayaraghavan, "Automated Tool for Resume Classification Using Sementic Analysis," International Journal of Artificial Intelligence and Applications (IJAIA), **10**(1), 2019.

[26] V. Bhatia, P. Rawat, A. Kumar, R. R. Shah, "End-to-end resume parsing and finding candidates for a job description using bert," arXiv preprint arXiv:1910.03089, 2019.

[27] Y.-C. Chou, H.-Y. Yu, "Based on the application of AI technology in resume analysis and job recommendation," in 2020 IEEE International Conference on Computational Electromagnetics (ICCEM), 291–296, IEEE, 2020, doi:10.1109/ICCEM47450.2020.9219491.

[28] P. Kadu, V. Z. Ashwini, "Knowledge Extraction from Text Document Using Open Information Extraction Technique," International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE) Volume, **9**, 2020.

[29] F. O. Oladipo, A. A. Ayomikun, "A MODEL FOR AUTOMATIC RESUME SUMMARIZATION," Journal of Information Systems & Operations Management, **14**(2), 147–159, 2020.

[30] P. K. Roy, S. S. Chowdhary, R. Bhatia, "A Machine Learning approach for automation of Resume Recommendation system," Procedia Computer Science, **167**, 2318–2327, 2020.

[31] K. Satheesh, A. Jahnavi, L. Iswarya, K. Ayesha, G. Bhanusekhar, K. Hanisha, "Resume Ranking based on Job Description using SpaCy NER model," International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395–0056, **7**(05), 2020.

[32] V. Mittal, P. Mehta, D. Relan, G. Gabrani, "Methodology for resume parsing and job domain prediction," Journal of Statistics and Management Systems, **23**(7), 1265–1274, 2020, doi:10.1080/09720510.2020.1799583.

[33] N. Khamker, "Resume Match System," International Journal of Innovative Science and Research Technology, 2021.

[34] S. Bhor, H. Shinde, V. Gupta, V. Nair, M. Kulkarni, "Resume parser using natural language processing techniques," 2021.

[35] R. Haddad, E. Mercier-Laurent, "Curriculum Vitae (CVs) Evaluation Using Machine Learning Approach," in IFIP International Workshop on Artificial Intelligence for Knowledge Management, 48–65, Springer, 2021.

[36] D. Vukadin, A. S. Kurdija, G. Delač, M. Šilić, "Information Extraction from Free-Form CV Documents in Multiple Languages," IEEE Access, 2021.

[37] K. F. F. Jiechieu, N. Tsopze, "Skills prediction based on multi-label resume classification using CNN with model predictions explanation," Neural Computing and Applications, **33**(10), 5069–5087, 2021.

[38] J. Zhao, J. Wang, M. Sigdel, B. Zhang, P. Hoang, M. Liu, M. Korayem, "Embedding-based Recommender System for Job to Candidate Matching on Scale," arXiv preprint arXiv:2107.00221, 2021.

[39] K. Tejaswini, V. Umadevi, S. M. Kadiwal, S. Revanna, "Design and Development of Machine Learning based Resume Ranking System," Global Transitions Proceedings, 2021.

[40] K. Bothmer, T. Schlippe, "Skill scanner: connecting and supporting employers, job seekers and educational institutions with an AI-based recommendation system," in The Learning Ideas Conference, 2022.

[41] I. Ali, N. Mughal, Z. H. Khand, J. Ahmed, G. Mujtaba, "Resume classification system using natural language processing and machine learning techniques," Mehran University Research Journal Of Engineering & Technology, **41**(1), 65–79, 2022.

[42] B. Amro, A. Najjar, M. Macido, "An Intelligent Decision Support System For Recruitment: Resumes Screening and Applicants Ranking," 2022.

[43] B. Kinge, S. Mandhare, P. Chavan, S. Chaware, "Resume Screening Using Machine Learning and NLP: A Proposed System," 2022.

[44] O. Dada, A. Kana, S. Abdullahi, "An ontology based approach for improving job search in online job portals," Journal of Computer Science and Its Application, **25**(1), 34–43, 2018.

[45] M.-I. Enăchescu, "Screening the Candidates in IT Field Based on Semantic Web Technologies: Automatic Extraction of Technical Competencies from Unstructured Resumes." Informatica Economica, **23**(4), 2019.

[46] B. Kelkar, R. Shedbale, D. Khade, P. Pol, A. Damame, "Resume Analyzer Using Text Processing," Journal of Engineering Sciences, 353–361, 2020.

[47] A. Wosiak, "Automated extraction of information from Polish resume documents in the IT recruitment process," Procedia Computer Science, **192**, 2432–2439, 2021.

[48] T. T. Phan, V. Q. Pham, H. D. Nguyen, A. T. Huynh, D. A. Tran, V. T. Pham, "Ontology-based resume searching system for job applicants in information technology," in International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, 261–273, Springer, 2021.

[49] H. Sajid, J. Kanwal, S. U. R. Bhatti, S. A. Qureshi, A. Basharat, S. Hussain, K. U. Khan, "Resume Parsing Framework for E-recruitment," in 2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM), 1–8, IEEE, 2022.

[50] N. Tülümen, G. S. A. Genç, A. Öztaş, T. Güngör, A. Nohutcu, "Ontology based job and resume matcher," .

[51] M. N. Uddin, T. H. Duong, K.-J. Oh, J.-G. Jung, G.-S. Jo, "Experts search and rank with social network: An ontology-based approach," International Journal of Software Engineering and Knowledge Engineering, **23**(01), 31–50, 2013.

[52] I. Abuhassan, A. M. AlMashaykhi, "Domain ontology for programming languages," Journal of Computations & Modelling, **2**(4), 75–91, 2012.

[53] L. Mostafa, S. Beshir, "Job candidate rank approach using machine learning techniques," in International Conference on Advanced Machine Learning Technologies and Applications, 225–233, Springer, 2021.

[54] D. Parikh, A. Kovashka, A. Parkash, K. Grauman, "Relative attributes for enhanced human-machine communication," in Proceedings of the AAAI Conference on Artificial Intelligence, volume 26, 2153–2159, 2012.

[55] A. Sainani, P. Reddy, "Extracting special information to improve the efficiency of resume selection process," 2011.

[56] F. Yasmin, M. I. Nur, M. S. Arefin, "Potential candidate selection using information extraction and skyline queries," in International Conference on Computer Networks, Big data and IoT, 511–522, Springer, 2019.

[57] O. Pathak, "A simple resume parser used for extracting information from resumes," https://omkarpathak.in/pyresparser/, 2019.

[58] "sklearn.svm.SVC," https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html.

[59] "Scikit Learn - Support Vector Machines," https://www.tutorialspoint.com/scikit_learn/scikit_learn_support_vector_machines.htm.

[60] MLee, "Visual Guide to the Confusion Matrix," https://towardsdatascience.com/visual-guide-to-the-confusion-matrix-bb63730c8eba, 2021.

[61] D. Cournapeau, "sklearn.metrics.accuracy_score," https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_{}score.html, 2007-2022.

[62] S. Solanki, "Model Evaluation & Scoring Metrics," https://coderzcolumn.com/tutorials/machine-learning/model-evaluation-scoring-metrics-scikit-learn-sklearn.

[63] D. Cournapeau, "sklearn.metrics.precision_score," https://scikit-learn/stable/modules/generated/sklearn.metrics.precision_{}score.html, 2007-2022.

[64] D. Cournapeau, "sklearn.metrics.recall_score," https://scikit-learn/stable/modules/generated/sklearn.metrics.recall_{}score.html, 2007-2022.

[65] D. Cournapeau, "sklearn.metrics.f1_score," https://scikit-learn/stable/modules/generated/sklearn.metrics.f1_{}score.html, 2007-2022.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 5, 17-27 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Feedback Controller for Longitudinal Stability of Cessna182 Fixed-Wing UAVs

Veena Phunpeng[1,*], Wilailak Wanna [1], Sorada Khaengkarn[1], Thongchart Kerdphol[2]

[1]*School of Mechanical Engineering, Institute of Engineering, Suranaree University of Technology, Nakhon Ratchasima, 30000, Thailand*

[2]*Department of Electrical Engineering, Faculty of Engineering, Kasetsart University, Bangkok, 10900, Thailand*

ARTICLE INFO

Article history:
Received:05 July, 2023
Accepted:08 August, 2023
Online: 24 September, 2023

Keywords:
Unmanned Aerial Vehicles
Fixed Wing
Fuzzy Logic Controller
PID Controller
Feedback loop controller

ABSTRACT

Unmanned aerial vehicles (UAVs) are becoming increasingly popular for both civil and military applications. Unmanned aerial vehicles can be categorized into two categories: rotary-wing and fixed-wing. Due to its capacity to fly long distances and carry substantial payloads, fixed-wing UAVs are gaining popularity and are currently utilized for various tasks. However, when confronted with disturbances such as weather or wind gusts, fixed-wing UAVs can rapidly lose stability, leading to a loss of lift and stalling. Consequently, it is vital to ensure the stability of fixed-wing UAVs. For the longitudinal stability management of a fixed wing unmanned aerial vehicle, the design and modeling of a feedback controller, including a PI controller, PID controller, and Fuzzy logic controller, are discussed in this article. MATLAB/SIMULINK©2021 will be used to design the control system and compare the response of each controller during the simulation. The controller's response to various input formats will indicate its capacity to regulate the system's behavior. Our results indicate that the fuzzy logic controller was superior to the PI and PID controllers at controlling the system's response according to the desired or input behavior.

1. Introduction

Unmanned aerial vehicles (UAVs) are becoming increasingly popular in a variety of civil and military applications, including aerial photography, shipping and delivery, geo-graphic mapping, disaster management, precision agriculture, search and rescue, weather forecasting, wildlife monitoring, law enforcement, and entertainment [1–3]. Unmanned aerial vehicles are small aircraft without a pilot on board. They can be remotely commanded and operated to accomplish operations in inaccessible and extremely dangerous regions. Unmanned aerial vehicles can be categorized into two basic categories: rotary-wing and fixed-wing [4]. Due to their capacity to fly long distances and carry heavy payloads, fixed-wing UAVs are currently gaining popularity and are employed for a variety of tasks [5]. Yet, fixed-wing UAVs quickly lose their stability when disturbances arise (weather or wind gusts). The instability of unmanned aircraft will cause the wings to lose lift, resulting in the stalling of fixed-wing UAVs [2,6–8]. Thus, the fixed-wing UAV's stability is vital.

Accidentally, Lion Air Indonesia's Boeing 737 MAX passenger aircraft crashed into the ocean in October 2018, resulting in 189 fatalities. Additionally, Ethiopian Airlines' Boeing 737 MAX caused 157 deaths in March 2019 [9]. Clearly, the breakdown of the flight control system caused the airplane to crash. The error-generating automatic flight control system is known as the Maneuvering Characteristics Augmentation System (MCAS) [10]. MCAS is a Boeing Company-developed system fitted in Boeing 737 MAX aircraft. MCAS is a mechanism that aids in the maintenance of stopped flights caused by an excessive angle of attack. MCAS is software coupled to the sensor that measures the angle of attack of an aircraft. The resultant MCAS mistake is the result of a faulty measurement of the rise of impact, in which the system calculates the impact angle to be greater than it actually is. The MCAS is activated when the angle of attack is greater than the angle of attack set by the system. This causes the elevator to adjust the aircraft to a progressiveMCASly decreasing angle when the pilot detects a change in the aircraft's pitch angle. The MCAS problem causes the plane to crash. Based on actual events aboard the Boeing 737 MAX, the significance of flight control systems in actual aircraft during flight is emphasized. Therefore, flight control systems are essential for unmanned aircraft. It relates directly to flight control and safety.

*Corresponding Author: Veena Phunpeng, veenap@g.sut.ac.th

The stability of fixed-wing UAVs can be broken down into three categories: directional stability, lateral stability, and longitudinal stability, with longitudinal stability being an essential type. It relates to the nose-up or nose-down position, takeoff or landing, and direct flying altitude of the UAV.

Fixed-wing UAVs employ a standard industrial feedback controller to reduce UAV expenses, such as a Proportional, Integral, Derivative, Proportional-Integral, and proportion-al-derivative controller, Proportional-Integral-Derivative controller, etc. [7,8,11]. Commercial controllers such as the Pixhawk and Paparazzi are commonly utilized in the flight control system of fixed-wing UAVs [4,12]. This present work describes control systems for fixed-wing UAVs.

In this study, the longitudinal stability management of a fixed wing unmanned aerial vehicle is constructed and simulated using a feedback controller comprising PI controller, PID controller, and fuzzy logic controller. MATLAB/SIMULINK©2021 will be used to simulate the control system and compare the responses of each controller. Modeling of aviation systems is explained in this study's summary. P Controller, I Controller, D Controller, PI Controller, PID Controller, and Fuzzy Logic Controller define the feedback controller. In the last section, the design of the control system is implemented, and MATLAB/SIMULINK©2021 simulation results are presented.

2. Aircraft System Modeling

The movement of a fixed wing unmanned aerial vehicle can be divided into three actions according to the imaginary rotation around the X, Y, and Z axes: roll, pitch, and yaw [8], as shown in Figure 1.

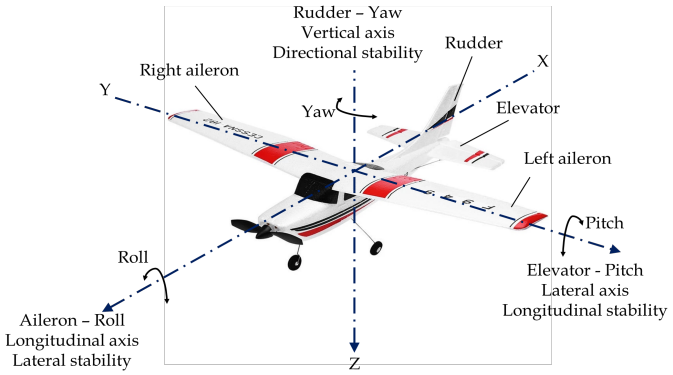


Figure 1: The movement of fixed-wing UAV.

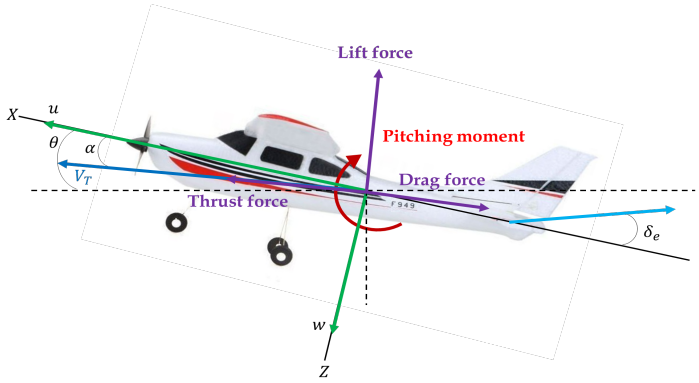


Figure 2: Pitch movement of fixed-wing UAV.

Pitch is the nose up or down of fixed-wing UAVs. It revolves around the Y-axis, an imaginary axis that goes from one wing tip to the other. It is controlled by a control surface known as an elevator, which is mounted on an empennage (horizontal stabilizer); when the elevator turns up and down (i.e., angle of the elevator: δ_E), which causes a pitch angle: θ as shown in Figure 2.

The pitch of the fixed-wing UAV can be written as a transfer function. In this research, the pitch transfer function of the fixed-wing UAV can be written as a transfer function, as shown in equation 1 [6,7,13].

$$\frac{\theta_s(s)}{\delta_{E_s}(s)} = \frac{-19893s^2 - 105510s - 15567}{86.1189s^4 + 1985.9478s^3 + 16150s^2 + 2082.5s + 945.7337} \quad (1)$$

where $\theta_s(s)$: Pitch angle
$\delta_{E_s}(s)$: Elevator angle

Equation 1 is a model of a fixed-wing Cessna182 UAV, with a ratio of 1:6.65 compared to a real scale, of which the fundamental dynamics of a mathematical model are similar to the real-sized aircraft. The wingspan of the model is 5.4135 ft. The dynamic of the pitching control system includes an actuator that can force the control surface to move as needed. The mathematical model used to control the elevator up and down is expressed in equation 2 [6,7,13].

$$\frac{\delta_E}{v_E} = -\frac{10}{s+10} \quad (2)$$

Where δ_E : Angle of the elevator
v_E : Input voltage

3. Feedback controller

Many systems in modern industries and other systems need a controller, such as an automobile steering control system, speed control systems, temperature control systems, tracking systems, flight control systems, etc. A controller controls the process or plant that responds to the method according to the desired output response. The control system can be divided into two systems 1) an Open-loop control system and 2) a Closed-loop control system.

The open-loop control system [12] is used in processes that do not require much control precision. Therefore, it is a process that does not require a back-measurement or feedback signal to be used to calculate constants for use in control. An open control process has system components, as shown in Figure 3. The application of the open-loop control system, such as fan control systems, washing machine control, etc.

Figure 3: Open-loop control system.

A closed-loop control system [14,15] or feedback control (see Figure 4) is a controller in which the actual signal received from the system is measured and compared with the desired output signal to calculate errors. The error can be used to determine the system control constants to adjust the system to the desired output.

The block diagram of the closed-loop control system is shown in Figure 5.

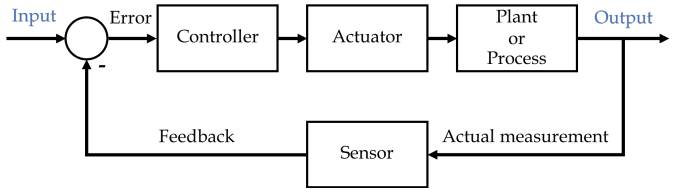


Figure 4: Closed-loop control system or feedback control system.

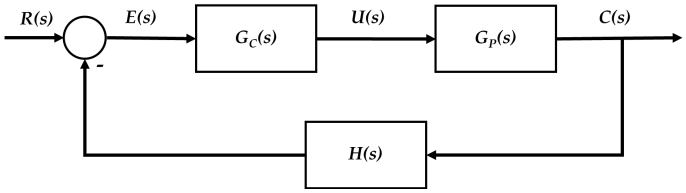


Figure 5: Block diagram of a closed-loop control system.

where R(s) : Input signal
E(s) : Error signal
GC(s) : Transfer function of the controller
U(s) : Control signal
GP(s) : Transfer function of plant or process
C(s) : Actual response signal
H(s) : Transfer function of the measurement sensor

General controllers for closed-loop control systems can be divided into Proportional-Integral (PI) controllers, Proportional-Derivative (PD) controllers, and Proportional-Integral-Derivative (PID) controllers. These controllers have used various industrial systems such as temperature, pressure, chemical, and vehicle manufacturing. The type of controller will have a different effect on the system response. Selecting a control type in the control design will depend on the response requirements of that system. The controllers can be divided into the following types.

3.1. P-Controller

A proportional controller [16] (P controller) is a control that determines the controller's gain or the proportion of the output signal to the input signal of the controller. The proportional controller has proportional gain to be used as an extension of the controller's input signal. If the proportional gain is high, the system will have a fast response. The result is that the response has an oscillation. On the other hand, when the proportional gain is low, the system will slowly respond. The proportional equation of the proportional control system can be found in equation 3.

$$G_C(s) = \frac{U(s)}{E(s)} = K_P \tag{3}$$

where K_P: Proportional Gain.

3.2. I-Controller

An integral controller is a working principle of controlling the output signal by signal integration. The integral controller can be described in equation 4.

$$G_C(s) = \frac{U(s)}{E(s)} = \frac{K_I}{s} \tag{4}$$

where K_I: Integral Gain

3.3. D-Controller

A derivative controller is a control that represents the derivation of the input signal of the controller. The derivative controller can be described in equation 5.

$$G_C(s) = \frac{U(s)}{E(s)} = K_D s \tag{5}$$

where K_D: Derivative Gain.

3.4. PI-Controller

A proportional-Integral [16] (PI) controller consists of a proportional or "P" controller and an integral or "I" controller. The P controller can be found in this equation.

$$G_C(s) = \frac{U(s)}{E(s)} = K_P + \frac{K_I}{s}$$

3.5. PID-Controller

The proportional integral derivative controller combines the proportional, integral, and derivative controllers shown in this equation.

$$G_C(s) = \frac{U(s)}{E(s)} = K_P + \frac{K_I}{s} + K_D s$$

Each controller has a different behavior, which can be summarized in Table 1.

In the control process, the gain must be selected according to the instability of the process. Select the controller gain that can be obtained from the Ziegler–Nichols method [17–19]. The procedure for finding the controller gain is as follows: The first step is to set K_I and K_D to the value 0, then increment K_P from zero until the maximum value is reached, resulting in K_U (Ultimate Gain). The system output will have a fixed vibration with the oscillation period or P_U. As shown in Table 2, the Ultimate gain (K_U) and Oscillation period (P_U) can be used to calculate the controller gain (K_P, K_I, and K_D).

Table 1: Characteristics of PID Gains K_P, K_I, and K_D.

Characteristic	K_P	K_I	K_D
Rise time	Decreases	Minimal Decrease	Minimal Decrease
Settling time	Minimal Decrease	Increase	Decrease
Overshoot	Increase	Increase	Decrease
Steady-state error	Decrease	Decrease	No impact
Stability	Decrease	Decrease	Increase

Table 2: Ziegler–Nichols method using Ultimate gain (K_U) and Oscillation period (P_U).

Controller	K_P	K_I	K_D
Proportional (P)	$0.5K_U$	-	-
Proportional-integral (PI)	$0.45K_U$	$\frac{0.54K_U}{P_U}$	-
Proportional- integral- derivative (PID)	$0.6K_U$	$\frac{1.2K_U}{P_U}$	$\frac{0.6K_UP_U}{8}$

3.6. Fuzzy Logic Controller

A fuzzy logic controller is a type of feedback controller that can be applied to a wide variety of systems [13,20–26]. Due to the ability to control nonlinear systems, complex systems Include systems that require immediate response (Real-time system) by a fuzzy logic controller is a controller that uses reasoning and decision-making principles similar to human decision-making with a work process structure as shown in Figure 6, which consists of four components as follows:

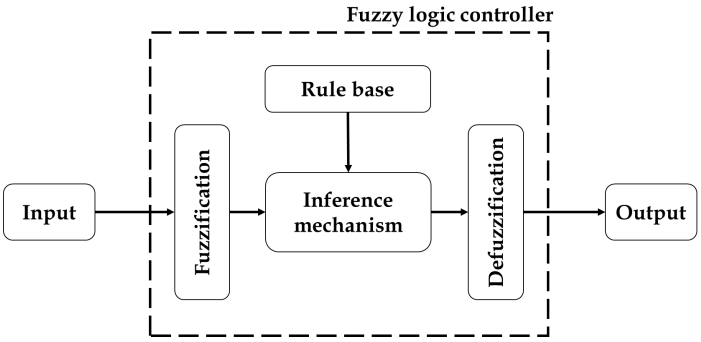


Figure 6: Fuzzy logic controller architecture.

- Fuzzification It is the part to convert the crisp input signal (Input) that is in the form of the original set with data values 0 and 1 (which can be interpreted as "false" and "true" as shown in Figure 7) into the form of a fuzzy set whose value is between 0 and 1 (the truth value is between false (0) and true (1), as shown in Figure 7).
- Rule base or rule base that was created as a tool used to evaluate the results of the input signal. The nature of the rule is in the form of "If-Then".
- The inference mechanism is used to make decisions or interpret problems related to the rule base.
- Defuzzification This is the process of converting the fuzzy set data back to the original data set.

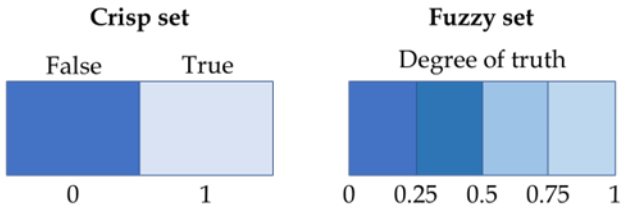


Figure 7: Crisp set and fuzzy set.

4. Simulation

This section shows the design of the control system through the program MATLAB/SIMULINK©2021. For the PI and PID controllers, the controller gains of PI and PID controllers are tuned by using Ziegler–Nichols method as shown in Figures 8 and 9. Where K_U = 3.162 and P_U = 0.384 s. The gain controller (K_P, K_I, and K_D) is shown in Table 3.

To design a fuzzy logic controller for longitudinal stability control of fixed-wing UAVs, able to design a control system through the program MATLAB/SIMULINK©2021, as shown in Figure 10.

From the structure of the fuzzy logic control system, as shown in Figure 10, it can be seen that there are two input signals for the fuzzy logic controller, the error signal (e) is the difference between the actual measurement pitch angle (output) is compared to the desired pitch angle. The error rate (Δe) is the error change rate over time. The fuzzy logic controller provides an output signal that can be used to control the transfer function of the pitch angle.

To design a fuzzy logic controller, the data value must be changed to a fuzzy set when receiving an incoming signal (crisp input signal) as a number or in the original set format. It uses the semantic characteristics of linguistic variables, which are linguistic variables for the input signal (error and error rate). The output signal is {NB, NM, NS, ZO, PS, PM, PB}, which means a fuzzy set. They have the following meanings: Negative Big, Negative Medium, Negative Small, Zero, Positive Small, Positive Medium, and Positive Big, respectively.

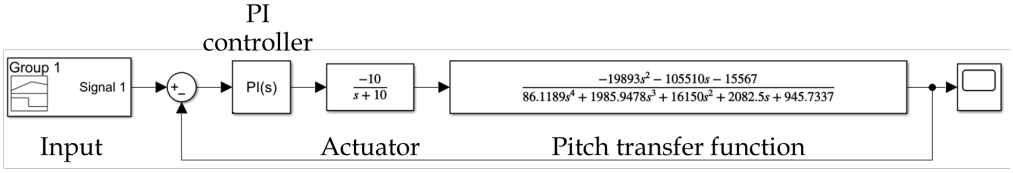


Figure 8: The PI control system.

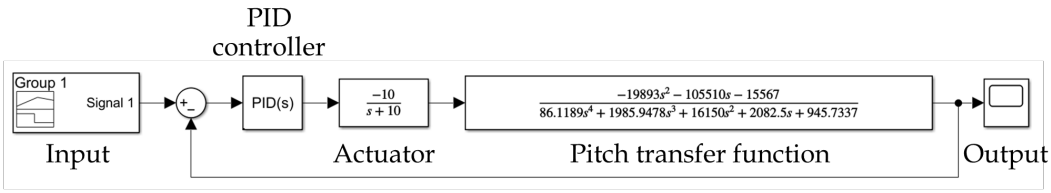


Figure 9: The PID control system.

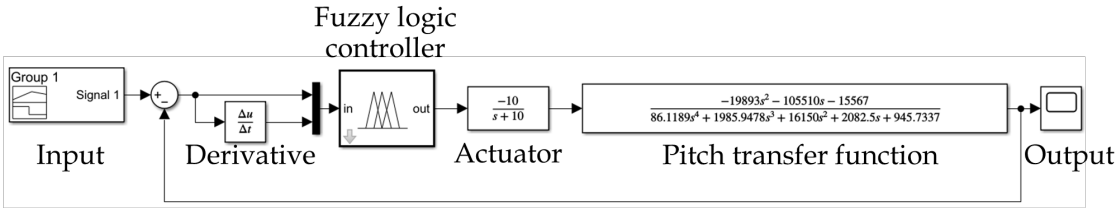


Figure 10: Fuzzy logic control system.

Table 3: The Ziegler–Nichols tuning PI and PID controller method uses Ultimate gain (K_U) and Oscillation period (P_U).

Controller	K_P	K_I	K_D
Proportional-integral (PI)	1.423	4.447	-
Proportional- integral- derivative (PID)	1.892	9.881	0.091

This research uses two types of membership functions for fuzzy sets: The triangular membership function for {NM, NS, ZO, PS, PM} and the trapezoidal membership function for {NB, PB} are shown in Figure 11 – 13.

The fuzzy rule base provides an output signal for the system's control. The fuzzy rule base is set according to the "If-Then" rule. The fuzzy rule base has 49 rules, as shown in Table 4.

Figure 14 shows the architecture of a Fuzzy-PID controller. This sort of controller combines a fuzzy logic controller with a PID controller employing an adaptive controller to select the best appropriate controller gains. Figure 15-16 illustrates the membership function employed by the Fuzzy-PID controller, whilst Table 5-7 displays the fuzzy rule base.

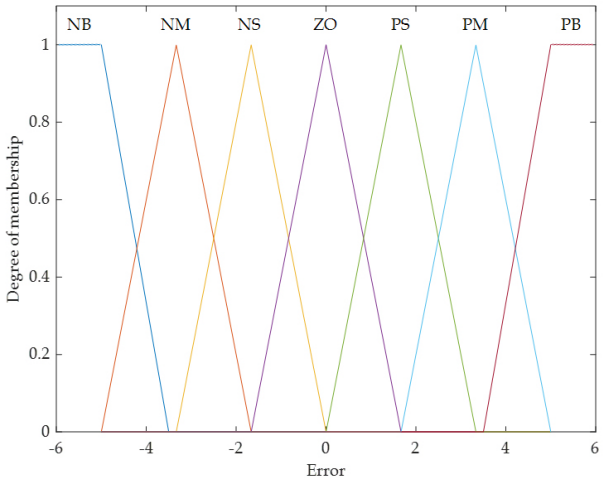


Figure 11: Membership function of error.

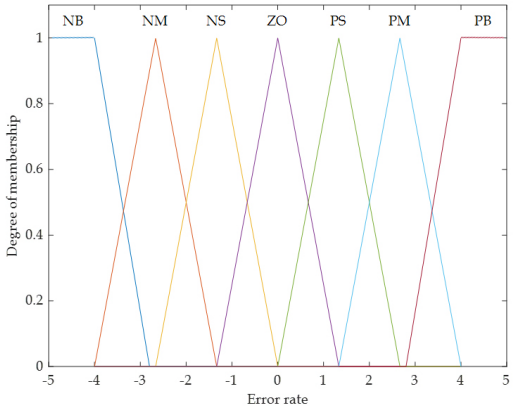


Figure 12: Membership function of error rate.

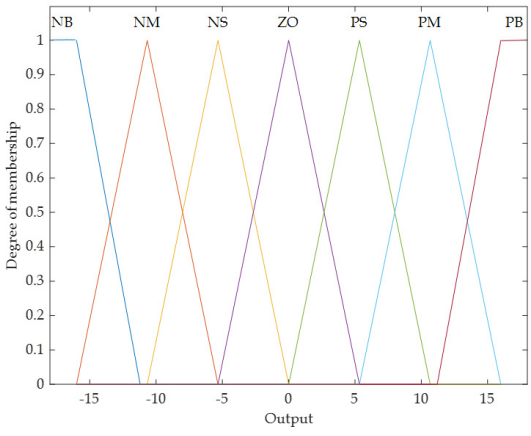
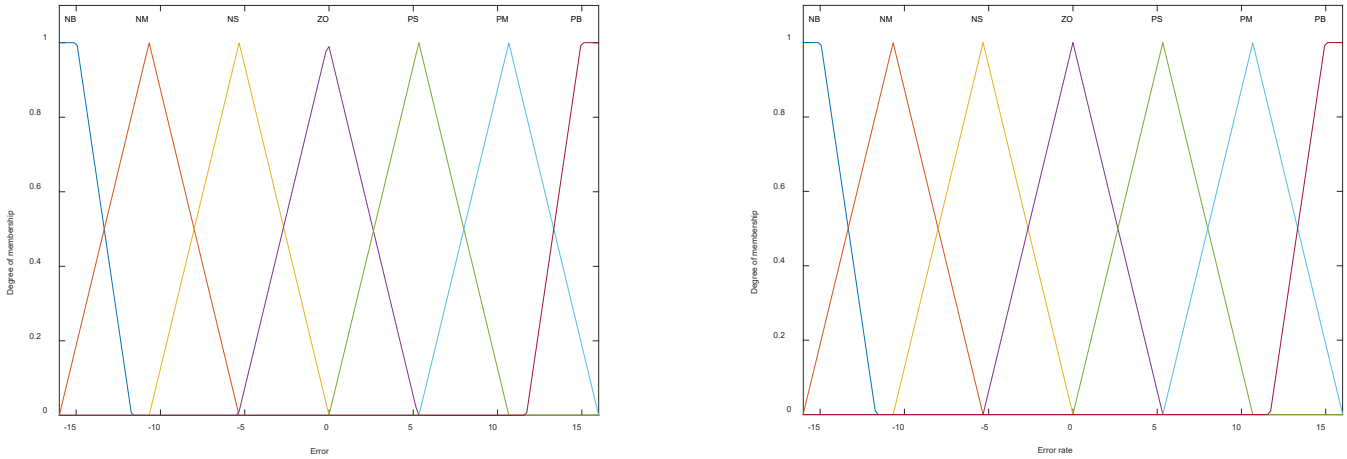


Figure 13: Membership function of output.

Table 4: Fuzzy logic rules base for fuzzy logic controller.

Error \ Error rate	NB	NM	NS	ZO	PS	PM	PB
NB	NB	NB	NB	NB	NB	NB	NM
NM	NB	NB	NB	NB	NB	NM	NS
NS	NB	NB	NB	NB	NM	NS	NS
ZO	NM	NS	NS	ZO	ZO	PS	PM
PS	PS	PS	PM	PB	PB	PB	PB
PM	PS	PM	PB	PB	PB	PB	PB
PB	PM	PB	PB	PB	PB	PB	PB



(a) (b)

Figure 15: Membership function of input for Fuzzy-PID controller (a) Error (b) Error rate.

Table 5: Fuzzy logic rules base for K_P.

Error \ Error rate	NB	NM	NS	ZO	PS	PM	PB
NB	PB	PB	PM	PM	PS	ZO	ZO
NM	PB	PB	PM	PS	PS	ZO	NS
NS	PM	PM	PM	PS	ZO	NS	NS
ZO	PM	PM	PS	ZO	NS	NM	NM
PS	PS	PS	ZO	NS	NS	NM	NM
PM	PS	ZO	NS	NM	NM	NM	NB
PB	ZO	ZO	NM	NM	NM	NB	NB

Table 6: Fuzzy logic rules base for K_I.

Error \ Error rate	NB	NM	NS	ZO	PS	PM	PB
NB	PB	PB	PM	PM	PS	ZO	ZO
NM	PB	PB	PM	PS	PS	ZO	NS
NS	PM	PM	PM	PS	ZO	NS	NS
ZO	PM	PM	PS	ZO	NS	NM	NM

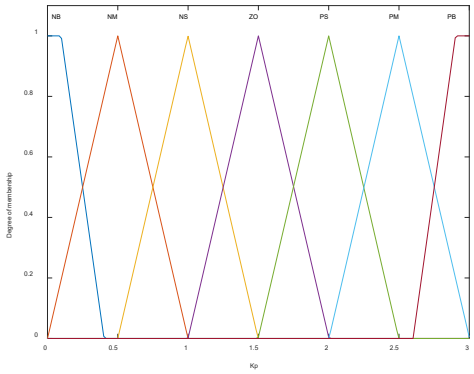
PS	PS	PS	ZO	NS	NS	NM	NM
PM	PS	ZO	NS	NM	NM	NM	NB
PB	ZO	ZO	NM	NM	NM	NB	NB

Table 7: Fuzzy logic rules base for K_D.

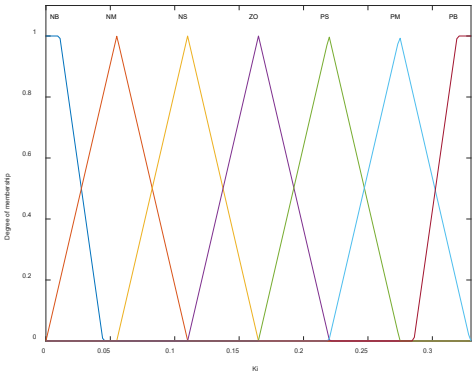
Error / Error rate	NB	NM	NS	ZO	PS	PM	PB
NB	PB	PB	PM	PM	PS	ZO	ZO
NM	PB	PB	PM	PS	PS	ZO	NS
NS	PM	PM	PM	PS	ZO	NS	NS
ZO	PM	PM	PS	ZO	NS	NM	NM
PS	PS	PS	ZO	NS	NS	NM	NM
PM	PS	ZO	NS	NM	NM	NM	NB
PB	ZO	ZO	NM	NM	NM	NB	NB

5. Results and Discussion

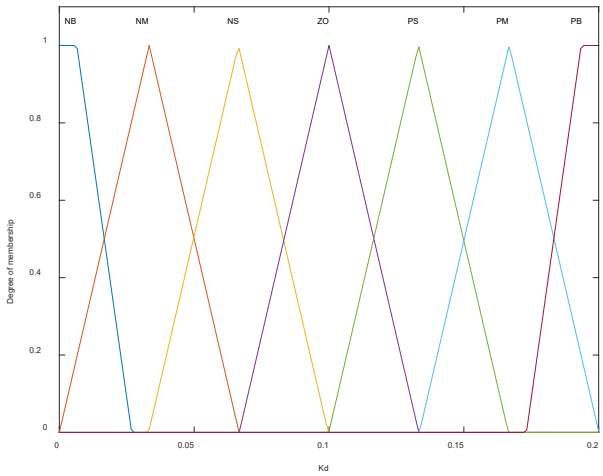
This section shows the simulation results of the designed control system for the pitch control system of the Cessna182 fixed-wing UAV through MATLAB/SIMULINK©2021. The results show the comparison of the response of the original system (no controller), PI controller, PID controller, and fuzzy logic controller.



(a)



(b)



(c)

Figure 16: Membership function of output for Fuzzy-PID controller (a) K_P (b) K_I (c) K_D.

For the original system (no controller), it was found that the system had zero position for open-loop transfer function at -0.1519 and – 5.152 and pole position at -0.0618 ± 0.2359i, – 10 and -11.4685 ± 7.2878i, as shown in Figure 17. For the pole at position -11.4685 ± 7.2878i, the system has a high overshoot and takes a long time to reach equilibrium (settling time).

Therefore, to improve the system's behaviour, reduce the oscillation (overshoot), and increase the stability of the pitch system. The control system (PI controller, PID controller, and fuzzy logic controller) was designed and simulated through MATLAB/SIMULINK©2021, with the following input format that can be divided into 3 cases.

Figure 18 shows the response of step-up input. The noncontrolled system is slow to respond and significantly overshoots with high errors. When the PI controller controls the system, the system has a faster response but a high overshoot. The PID controller provides a quick response and lower overshoot

than the PI controller. Comparing these results with the fuzzy logic controller shows that the reaction does not have an overshoot.

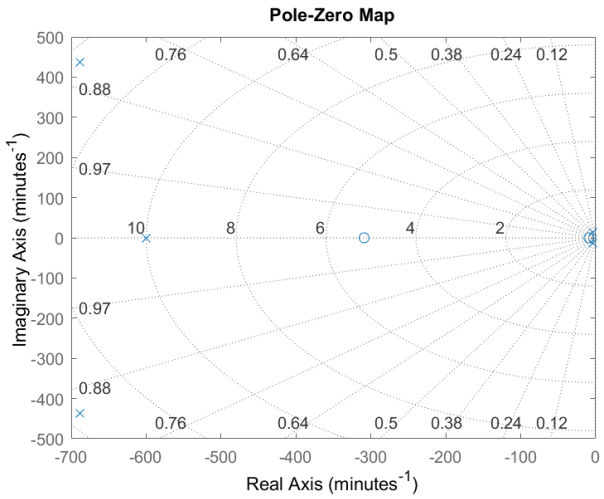


Figure 17: Pole and zero of system.

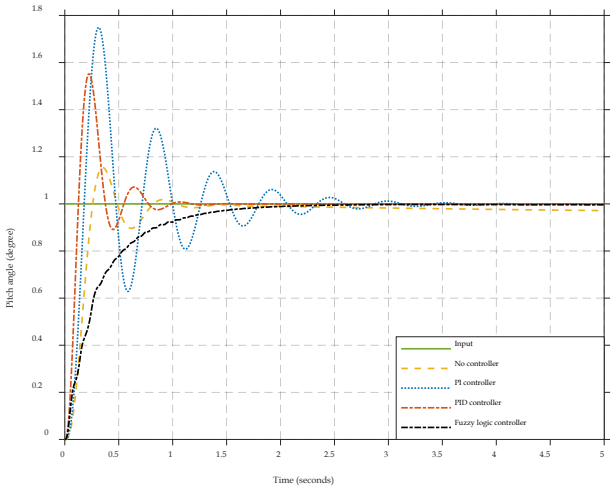


Figure 18: Input as a step up.

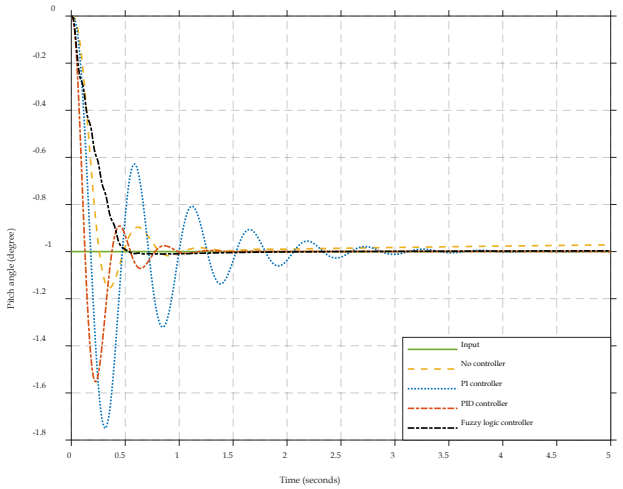


Figure 19: Input as a step-down.

Figure 19 shows that the system response with step-down input has the same response as a system with step-up input. Systems with no controller have slow response and overshoot. The PI and PID controller systems have a faster response but also a high overshoot. The fuzzy logic controller provides the best response without overshoot and fast system response.

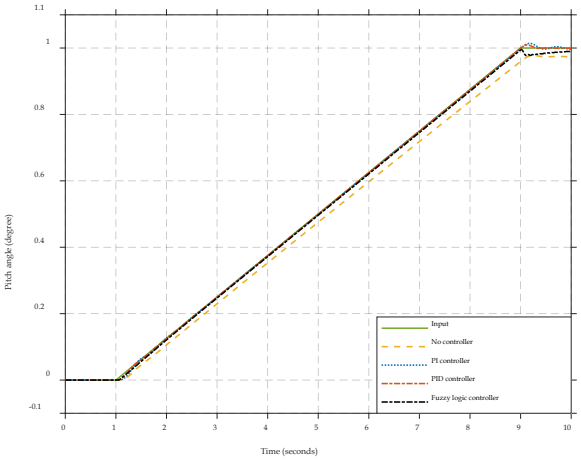


Figure 20: Input as ramp up.

Figure 20 illustrates the response of the system to a ramp-up input. Without a controller, the system malfunctions. The system's response demonstrates that both PI and PID controllers result in overshoot. Nonetheless, the fuzzy logic controller responds adequately to the desired input without overshooting.

For the ramp-down input, the response is shown in Figure 21 The system response result clearly indicates that the system has an error when there is no controller system. The PI and PID controllers decrease the system's error but have high overshoot. The fuzzy logic controller provides a satisfactory response with reduced error and overshoot.

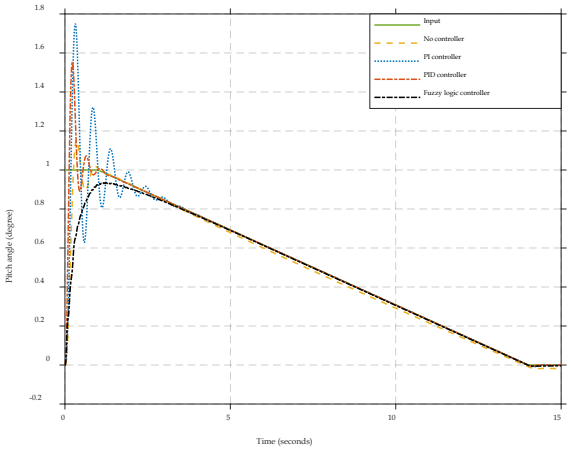


Figure 21: Input as ramp down.

In the second case, the input is a combined disturbance divided into three modes. The response in the second case is shown in Figure 22 – 26.

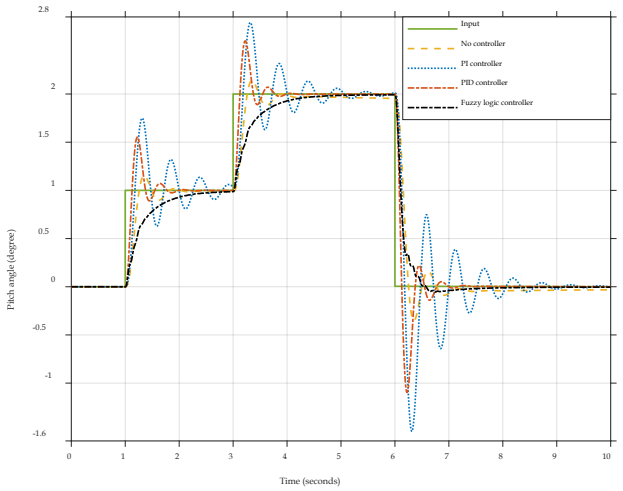


Figure 22: Input mode 1.

The first mode's response is shown in Figure 22. The system response shows indicates system without a controller has errors and overshoots. PI controllers provide high system response with high overshoot compared to PID controllers and fuzzy logic controllers. The fuzzy logic controller provides a system response that is fault-free and overshoot-free, meaning the system has no oscillation.

For the second mode (Figure 23), this input combines the step input and the ramp input. The system response is as close as possible to the input. From the system's response, it was found that no controller system has the most valuable error. Fuzzy logic controllers provide the best system response compared to PI and PID controllers.

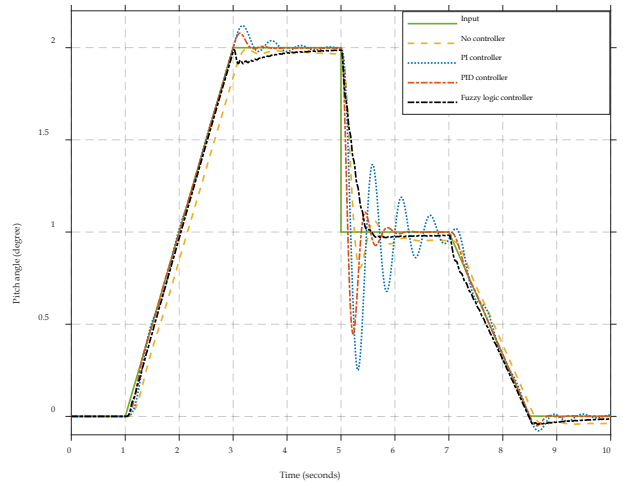


Figure 23: Input mode 2.

The input is a sine wave in the third mode (Figure 24). From the system's response, it was found that no controller system has the most valuable error. The PID and Fuzzy logic controllers provide the best system response and are as close as possible to the input.

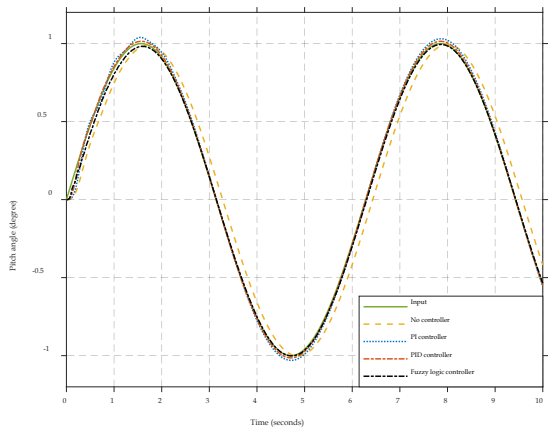


Figure 24: Input mode 3.

For the third case, set the system input as a continuous disturbance, as shown in Figure 25, and zoom-in of input as a constant disturbance shown in Figure 26. From the system's response, the no-controller system has a high error. PI and PID controllers, a high overshoot is provided. The fuzzy logic controller provides the most input-compliant system response. The system response was no error and overshoot.

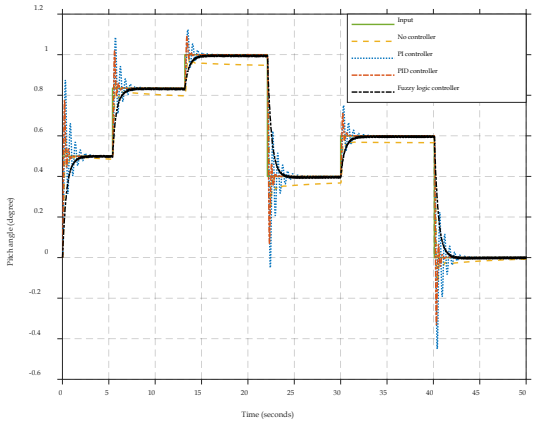


Figure 25: Input as a continuous disturbance.

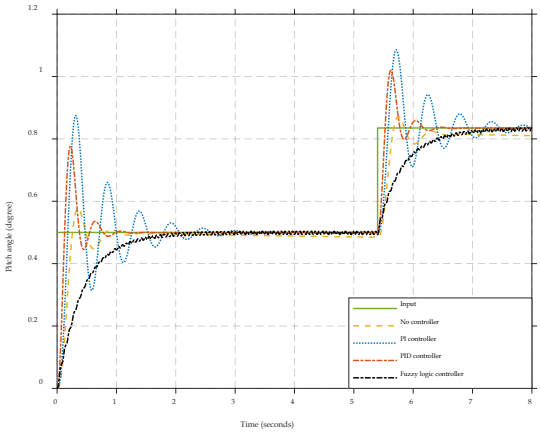


Figure 26: Zoom-in of input as a continuous disturbance.

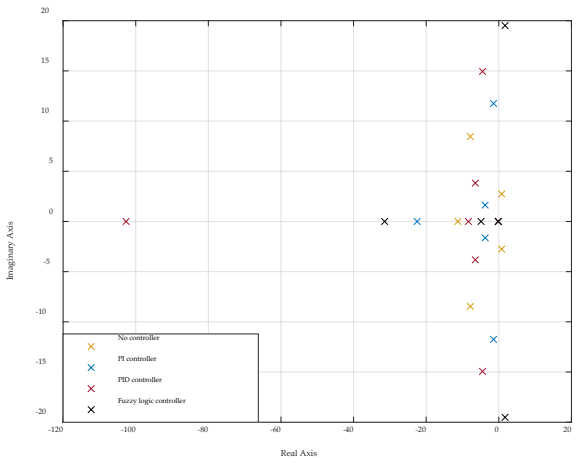


Figure 27: Pole displacement of input as a continuous disturbance.

Figures 27 show the pole displacement of each input mode. The pole displacement of each input mode is the same position because each input mode uses the same controller gains and rules (no controller, PI controller, PID controller, and fuzzy logic controller). Therefore, pole displacements are the same. From the pole position of no control system, PI control systems and PID control systems have pole displacement on imaginary axes that show the system has overshoot or oscillation in the system. For fuzzy logic control systems, make the system not have a pole on an imaginary axis. The system has a fuzzy logic control system without overshoot or oscillation.

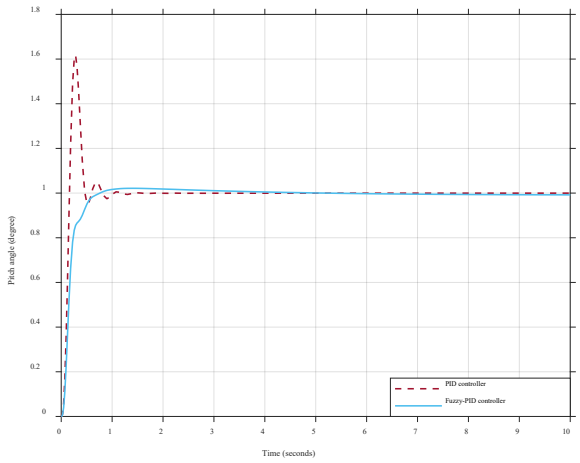


Figure 28: Response of PID and Fuzzy-PID controller.

Figure 28 compares the PID controller's response to that of the fuzzy PID controller. The control results show that a controller with the application of a fuzzy logic controller has a better system response than a system using a PID controller. As observed from the control effect, when using a controller with the application of a fuzzy logic controller, the overshoot and rise time can be reduced. The steady-state errors are removed.

6. Conclusion

UAVs tend to lose stability easily when they are disturbed by disturbances (Weather or wind gusts). This can cause the UAV to lose control and possibly stall easily. Therefore, UAV systems need a controller that is used to control the stability of the UAV. The longitudinal stability is the most important in many types of stability of UAVs due to the longitudinal stability directly influencing the pitch motion of the UAVs. The pitch motion directly involves UAVs' take-off, landing, and holding altitude. In this research, the design and simulation of longitudinal stability control of a fixed-wing unmanned aerial vehicle with a feedback controller include PI controller, PID controller, and fuzzy logic controller. The simulation uses MATLAB/SIMULINK©2021 to design the control system with each controller and compare the response of each controller. The controller's response according to the input format shows the ability to control system behavior. The comparison of the PI, PID, and fuzzy logic controllers show the fuzzy logic controller was able to control the system response as desired or according to the behavior of the input to the system. In the future, in addition to longitudinal stability control, fuzzy logic controllers can also be applied to control other systems in UAVs.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research was made possible by the Suranaree University of Technology (Grant Number FF7-707-66-12-82), Thailand Science Research and Innovation (TSRI), and National Science Research and Innovation Fund (NSRF) (NRIIS number 179358).

References

[1] H. Shakhatreh, A.H. Sawalmeh, A. Al-Fuqaha, Z. Dou, E. Almaita, I. Khalil, N.S. Othman, A. Khreishah, M. Guizani, Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges, IEEE Access, **7**, 48572–48634, 2019, doi:10.1109/ACCESS.2019.2909530.

[2] Y. Cheng, S. Xu, Z. Huang, B. Xiong, C. Wang, "Stability control of small fixed-wing UAV," in Proceeding of the 11th World Congress on Intelligent Control and Automation, IEEE: 1056–1060, 2014.

[3] V. Phunpeng, T. Kerdphol, "Integrated Piezoelectric and Flexible Solar Systems for Enhancing Inflight Operation Time of Small UAV," Suranaree Journal of Science and Technology, In press, 1–25, 2022.

[4] M.-C. Wen, C.-H. Yang, E.-X. Sung, T.-H.R. Wu, Optimized integration of UVAs surveys, and image-based modeling strategies for digital terrain model reconstruction Program of Smart Living View project, 2014.

[5] J.L. Hernandez, I. González-Hernández, R. Lozano, "Attitude and altitude control for a fixed wing UAV applied to photogrammetry," in 2019 International Conference on Unmanned Aircraft Systems (ICUAS), IEEE: 498–502, 2019.

[6] J.H. Yang, H.K. Xu, "Robust Controller Design for Non-Minimum Phase UAV System and System Analysis," IEEE Access, **6**, 70734–70769, 2018, doi:10.1109/ACCESS.2018.2879649.

[7] B.M. Albaker, M. Majidi, B.M. Albaker, N.A. Rahim, "Flight path PID controller for propeller-driven fixed-wing unmanned aerial vehicles Related papers Flight Cont rol Syst em for Small High-Performance UAVs EXPERIMENTAL FRAMEWORK FOR EVALUAT ION OF GUIDANCE AND CONT ROL ALGORIT HMS FOR UAVs Flight path PID controller for propeller-driven fixed-wing unmanned aerial vehicles," International Journal of the Physical Sciences, **6**(8), 1947–1964, 2011, doi:10.5897/IJPS11.162.

[8] J. Yang, X. Wang, S. Baldi, S. Singh, S. Fari, "A software-in-the-loop implementation of adaptive formation control for fixed-wing UAVs,"

IEEE/CAA Journal of Automatica Sinica, **6**(5), 1230–1239, 2019, doi:10.1109/JAS.2019.1911702.

[9] J. Herkert, J. Borenstein, K. Miller, "The Boeing 737 MAX: Lessons for Engineering Ethics," Science and Engineering Ethics, **26**(6), 2957–2974, 2020, doi:10.1007/s11948-020-00252-y.

[10] S. Makó, M. Pilát, P. Šváb, J. Kozuba, M. Čičváková, "Evaluation of MCAS System," Acta Avionica Journal, 2020, doi:10.35116/aa.2020.0003.

[11] Z. Zheng, Z. Jin, L. Sun, M. Zhu, "Adaptive Sliding Mode Relative Motion Control for Autonomous Carrier Landing of Fixed-Wing Unmanned Aerial Vehicles," IEEE Access, **5**, 5556–5565, 2017, doi:10.1109/ACCESS.2017.2671440.

[12] E. Ebeid, M. Skriver, J. Jin, "A Survey on Open-Source Flight Control Platforms of Unmanned Aerial Vehicle," in Proceedings - 20th Euromicro Conference on Digital System Design, DSD 2017, Institute of Electrical and Electronics Engineers Inc.: 396–402, 2017, doi:10.1109/DSD.2017.30.

[13] M.J. Patyra, J.L. Grantner, K. Koster, Digital Fuzzy Logic Controller: Design and, 1996.

[14] Electrical4U, "What is a Control System? (Open Loop & Closed Loop Control Systems Explained)," Https://Www.Electrical4u.Com/, 2020.

[15] M. Saoudi, A. El-Sayed, H. Metwally, "Design and implementation of closed-loop control system for buck converter using different techniques," IEEE Aerospace and Electronic Systems Magazine, **32**(3), 2017, doi:10.1109/MAES.2017.150261.

[16] V. Jain, "Types of Controllers | Proportional Integral and Derivative Controllers | Electrical4U," Https://Www.Electrical4u.Com/, 2021.

[17] V.V. Patel, "Ziegler-Nichols Tuning Method," Resonance, **25**(10), 2020, doi:10.1007/s12045-020-1058-z.

[18] V.V. Patel, "Ziegler-Nichols Tuning Method: Understanding the PID Controller," Resonance, **25**(10), 2020, doi:10.1007/s12045-020-1058-z.

[19] V. Kumar, A. Patra, "Application of Ziegler-Nichols Method for tuning of PID controller," International Journal of Electrical and Electronics Engineers, **8**(2), 2016.

[20] M.J. Patyra, J.L. Grantner, K. Koster, "Digital fuzzy logic controller: Design and implementation," IEEE Transactions on Fuzzy Systems, 4(4), 1996, doi:10.1109/91.544304.

[21] C. Elmas, O. Deperlioglu, H.H. Sayan, "Adaptive fuzzy logic controller for DC-DC converters," Expert Systems with Applications, **36**(2 PART 1), 2009, doi:10.1016/j.eswa.2007.11.029.

[22] J. Carvajal, G. Chen, H. Ogmen, "Fuzzy PID controller: Design, performance evaluation, and stability analysis," Information Sciences, **123**(3), 2000, doi:10.1016/S0020-0255(99)00127-9.

[23] A.A. Thorat, S. Yadav, S.S. Patil, "Implementation of Fuzzy Logic System for DC Motor Speed Control using Microcontroller," Journal of Engineering Research and Applications (IJERA), **3**(2), 2013.

[24] A.C. Soh, E.A. Alwi, R.Z.A. Rahman, L.H. Fey, "Effect of Fuzzy Logic Controller Implementation on a Digitally Controlled Robot Movement," Kathmandu University Journal of Science, Engineering and Technology, **4**(1), 1970, doi:10.3126/kuset.v4i1.2881.

[25] B. Hamed, M. Almobaied, "Fuzzy PID Controllers Using FPGA Technique for Real Time DC Motor Speed Control," Intelligent Control and Automation, **02**(03), 2011, doi:10.4236/ica.2011.23028.

[26] V. Phunpeng, S. Khaengkarn, T. Hengmeechai, T. Kerdphol, "Longitudinal Stability Assessment of Fixed-Wing UAV based on Fuzzy Closed-Loop Control," in Proceeding - 2023 International Electrical Engineering Congress, iEECON 2023, 2023, doi:10.1109/iEECON56657.2023.10126642.

ASTES Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 6, 19-25 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ
ISSN: 2415-6698

Tree-Based Ensemble Models, Algorithms and Performance Measures for Classification

John Tsiligaridis*

Heritage University, Mathematics & Computer Science Department, Toppenish, WA, 98948 USA

ARTICLE INFO

History:
Received:15 July, 2023
Accepted:29 October, 2023
Online: 30 November, 2023

Keywords:
Decision Trees
Ensemble Models
Bagging

ABSTRACT

An ensemble method is a Machine Learning (ML) algorithm that aggregates the predictions of multiple estimators or models. The purpose of an ensemble module is to provide better predictive performance than any single contributing model. This can be achieved by producing a predictive model with reduced variance using bagging, and bias using boosting.
The Tree-Based Ensemble Models with Decision Tree (DT) as base model is the most frequently used. On the other hand, there are some individual Machine Learning algorithms that can provide more competitive predictive power to the ensemble models. It is a problem, and this issue is addressed here. This work has two parts. The first one presents a Projective Decision Tree (PA) based on purity measure. Next node criterion (CNN) is also used for node decision making. In the second part, two sets of algorithms for predictive performance are presented. The Tree-Based Ensemble model includes bagging and boosting for homogeneous learners and a set of known individual algorithms. Comparison of two sets is performed for accuracy. Furthermore, the changes of bagging and boosting ensemble performance under various hyperparameters are also investigated. The datasets used are the sonar and the Breast Cancer Wisconsin (BCWD) from UCI site. Promising results of the proposed models are accomplished.

1. Introduction

Decision Trees (DTs) are an important type of algorithm for predictive modeling machine learning. It's often used to plan and plot business and operational decisions as a visual flowchart. The approach sees a branching of decisions which end at outcomes, resulting in a tree-like structure. This paper is work extended of an original one that appeared at the ICAIIC 2023 [1].

Decision tree induction is the method of learning the decision trees from the training set. The training set consists of attributes and class labels.[2]-[4]. Ensemble method is an algorithm that aims to improve the predictive performance on a task by aggregating the predictions of multiple estimators or models. The goal of the ensemble methods is to combine the predictions of several base estimators to produce improved results. Use of an ensemble model and optimization with parameter tuning can provide higher accuracy.

An easy way to combine the predictions is the majority voting where for each base estimator is assigned an equal weight. If we have m base estimators each base estimator has weight of 1/m. The weighted predictions of the individual base estimators are combined, and the most voted class is predicted. This way the classifier with the higher accuracy is selected. The ensemble models have more abilities to generalize compared to the single DT's predictions since it provides comparable bias and smaller variance. The Tree-Based Ensemble models belong to homogenous ensemble ones and use the same base learning algorithm; the DT classifier which is sensitive to small data variations. Random Forest (RF), an ensemble of randomized DTs, is used to further promote ensemble diversity. It predicts using the majority vote of all DTs [2]-[4].

One of the DTs problems is the creation of over-complex trees with replication and repetition of subtrees that do not generalize the data well. The Projective Decision Tree Algorithm (PA) can avoid this disadvantage by selecting the partition that maximizes the purity of the split with the use of CNN. The ensemble methods [5] produce an optimal predictive model with the combination of several base models. For the creation of the ensemble models the PA is used as base model. For creating and testing the prediction, two bagging methods are generated from PA, Random Forest (PARF) and the Extra Tree (PAET) along with two boosting methods; AdaBoost (AB) and Gradient Boosting (GB). A set of

*Corresponding Author: John Tsiligaridis, Heritage University,3240 Fort Rd,Toppenish,WA98948 ,USA, tsiligaridis_j@heritage.edu

https://dx.doi.org/10.25046/aj080603

individual algorithms; the PA, the k Nearest Neighbor (kNN), and the Support Vector Machine (SVM) are included in the examination of the performance improvement with tuning methods.

SVM models are also used in medical informatics to classify persons with or without diseases and especially for diabetes categories (undiagnosed diabetes, or no diabetes) [6]. In [7] SVM is a strong tool that has been used for cancer genomic classification or subtyping. Logistic regression (LR) estimates the probability of event occurrence given a dataset of independent variables [3].

A set of tree-based ensemble models comprise Random Forest (RF) [8] and the Extra Tree (ET) [9]. Both are based on PA. Since ET trees work randomly, they are faster than RF that looks for optimal split at each node. To decrease bias, ET uses original training samples instead of bootstrap replicas. Recent applications include land cover classification using Extremely Randomized Trees [10]. Tree-Based Ensemble model is used for investment in the stock market facilitating financial decision making. The purpose of the model is to minimize the prediction error and reduce the investment risk [11]. In [12], Tree-based machine learning models predict microbial fecal contamination in beach water for public health awareness. Ensemble methods, RFs and ETsare used for sensitivity analysis of environmental models [13].

The set of individual algorithms consists of PA, the Logistic Regression (LR) [14], the k Nearest Neighbor (kNN) [15], and the Support Virtual Machine (SVM) [16]. In [17] an Ensemble Model with Random Forest (RF), AdaBoost (AB) and XGBoost is used for weather forecasting. The ensemble learning model outperforms the simple Decision Tree (DT) in either calm or stormy environment. An AdaBoost ensemble method with reduced entropy for Breast Cancer prediction is developed in [18]. For this purpose, the target column is created from weighted entropy. In [19] a new ensemble Machine Learning method based on AdaBoost is developed for placement data classification analysis. It increases performance in terms of time complexity and accuracy for the student dataset.

Tuning parameter methods are applied before proceeding with the ensemble models. The comparison of the two sets' components shows individual algorithms could have better performance than the ensemble models after parameter tuning.

The paper's organization is as follows. In section 2, performance evaluation and process description are included. Section 3 deals with the Algorithms (DT, PA, kNN and SVM). Section 4 covers the Ensemble models, PARF, PAET, AB and GB. Section 5 contains the Ensemble Performance Issues. Bagging Hyperparameters, Boosting Hyperparameter and Control Overfitting are included in Section 6,7,8 respectively. Simulation results are provided in Section 9.

2. Performance Evaluation and Process Description

To evaluate learning models' performance the cross-validation technique is used. Cross-validation provides a more robust estimate of the model's performance on unseen data, and it prevents overfitting. The data are randomly divided into k folds almost of the same size. The k-1 folds are used for training while the one-fold is selected for validation. It is a method that generally results in a less biased estimate of the model compared to the simple train/test split. The out of sample testing refers to cross-validation where the model is built on a subsection of data and then tested on data that were not used to build it. The out of sample provides us with the information on how well the model predicts results for the "unseen" data (validation set). For each individual algorithm, the hyperparameters are tuned using the grid search method. The process of this work has two phases as below:

- Preparation phase:

  Prepare: PA (purity measure with the CNN criterion), the algorithms (LR, kNN, SVM), parameters tuning (kNN, SVM), PARF, and PAET.

- Execution phase:

  Accuracy: for algorithms, and ensemble models.

The model process is led by PA (base model) which provides the DTs for RF (PARF) and for ET (PAET). The implementation starts with the PA and then follows the PARF and PAET. Details of the used algorithms are presented in the next section.

3. Algorithms

3.1. Decision Tree (DT)

Inductive inference uses specific examples to make a general conclusion. It is a widely used method for DTs learning and produces a target function with discrete output values (i.e., binary). DTs tend to overfit the training data, in case of very deep or complex tree, mainly due to replication problems. In that case, two or more copies of the same subtree can be created. These DTs fail to generalize since they provide poor performance on new, unseen data [3]. Instability can be created to the structure of the DT due to the sensitivity of the training set when a small change of data (i.e., irrelevant attribute) or noise appears [3].

3.2. Projection Algorithm (PA)

The Projection algorithm (PA) a top-down DT inducer, can create a new model by learning the relationships between the descriptive features and a target feature. In PA, the next splitting node is decided by the CNN criterion based on purity using conditional probabilities. In the splitting process the data partition is achieved so that the highest purity attribute in the new nodes is selected. The dataset splitting process continues with the creation of new subset until pure sets are acquired. CNN uses conditional probability values to define the new internal node.

There are two PA phases. The primary one is to discover the root node that has the feature with the lowest impurity.For each feature (d) the number of instances with feature value t, with target feature value k is given by :

$a_{d,t,k} = |\text{features}_d = t|$ U $|\text{target} = k|$, $b_{d,t} = \text{purity}(a_{d,t,k})$.

The $a_{d,t,k}$ stands for projection of feature d, with value t over the target feature value =k

The feature with the maximum value of purity can be found as

$c = \max_d \sum_{t=1}^{n} b_{d,t}$ where n= the number of feature values.

The second phase is for branch selection. The next node for all feature values is created by the previous node. The new node is determined by the maximum value of conditional probabilities as follows.

$$p' = p(f1_{d1} = t, f2_{d2} = t, f3_{d3} = t/f_n = s)$$

$$= p(f1_{d1} = t, f2_{d2} = t, f3_{d3} = t) / p(f_n = s)$$

$f1_{d1} = t$ means the value of feature $f1$ is t, and $f2_{d2} = t$ means the value of feature $f2$ is t.

CNN determines the next internal node according to the following p' values.

if $p' \neq 1$, next internal node is created (purity ≠1). CNN is valid.
if p'=1 terminal node is created (purity=1). CNN is not valid.
if p'=0 no internal or terminal node created

A simple version of Zoo Animal Classification is used as the dataset containing animals' properties as features, and their species as target feature.

Example 1: (Phase 1: root discovery)

The s1= {toothed (True, False), breathes (True, False), legs (True, False) and the target feature (Mammal, Reptile).

From the dataset counting the total number of instances for feature values and target feature values is as follows.

For toothed: True + mammal: 6 false + mammal: 1 true +reptile: 2 false + reptile: 2 purity (true) = 5/7, purity(false) = 2/3 tot_purity = 1.3802

For breathes: True + mammal: 6 false + mammal: 1 true +reptile: 2 false + reptile: 1 purity (true) = 6/8, purity(false) = 1/2 tot_purity = 1.25

For legs: True + mammal: 6 false + mammal: 0 true +reptile: 1 false + reptile: 3 purity (true) = 6/7, purity(false) = 3/3=1 tot_purity = 1.857.

Root will be the legs feature due to the maximum value of tot_ purity.

Example 2:

The next step after the discovery of the root node from phase 1, the conditional probabilities are estimated from the two feature values of the previous node. Assume that "toothed" will be the next node and the branch is: "legs=true".

For the branch: "legs=true".

The process computes the conditional probability for all the feature values of the branch "legs=true" having any target feature value: Mammal, Reptile.

$$p'(1) = p(f1_{d1} = t, f2_{d2} = t, f3_{d3} = t)/p(f_n = s)$$
$$= p(\text{toothed} = \text{true}, \text{breathes} = \text{true}, \text{species} = \text{mammal})/p(\text{length} = \text{true})$$
$$= (5/10) / (7/10) = 5/7$$

$p'(2) = p$ (toothed = true, breathes = false, species = mammal/ legs = true) = 0.

$p'(3) = p$ (toothed = false, breathes = true, species = mammal/ legs = true) = 1/7.

$p'(4) = p$ (toothed = false, breathes = false, species = mammal/ legs = true) = 0.

$p'(5) = p$ (toothed = false, breathes = false, species = reptile/ legs = true) = 1/7.

The sum of probability values of "legs=true" branch equals 1, with 6 mammal and 1 reptile instances. This is an intermediate node with 7 instances. The PA's counting process examines the number of instances arising with the feature values' projection over a target feature value. The database split is accomplished by each feature's values providing a new subset for the next split. This process continues to attain pure sets.

3.3. kNN

kNN is a distance-based non-parametric algorithm. It classifies objects based on their proximate neighbors' classes. The k value (hyperparameter) selected specifies the examples' number closest to the query. The tuning parameter k considers 7 neighbors from all odd values (1 to 21). The 10-fold cross validation performs evaluation of each k value on the training dataset.

3.4. SVM

In SVM, a data item is represented by a point of n-dimensional space. These points become inputs and outputs of the hyperplane. Each feature value has a certain coordinate. The hyperplane tries to ensure that the margin between the closest points of different classes should be as maximum as possible. The classification is performed discovering the hyperplane that differentiates classes. SVM's effectiveness is apparent in high dimensional cases. For the decision functions different kernel functions can be specified. The C tuning parameter has value 1 (0,…,2.0) with Radial Base Function (RBF) kernel . A grid search is used for 10-fold cross validation as with kNN.

3.5. LR

LR is often used for probability estimation of an instance to belong to a certain class. It is a linear algorithm, and it is used for binary classification, computing the cost function for each instance. This convex function is used with Gradient Descent for global minimum discovery .

4. Ensemble Models

4.1. Ensemble Methods, Bagging

Ensemble models use a combination of multiple other models for prediction that are considered as base estimators. They do better in terms of technical challenges instead of building a single estimator. Ensemble methods use a variety of aggregation techniques depending on the task, including majority vote, model averaging, weighted mean, etc.

Bagging is the most basic homogenous parallel ensemble method we can construct. As an example, for a bagging ensemble with 500 Decision Trees, each of depth 12 and trained on bootstrap samples of 300 size has accuracy of 89,9% compared to a single tree with accuracy of 83,8%.

Bagging has a smoothing behavior due to the model aggregation. In the case of many nonlinear classifiers, where each trained on a slightly different replicate of training data, and then each one might create an overfit, but the difference is that do not all overfit the same way. Hence, the aggregation leads to

smoothing which finally reduces the effect of overfitting. In this way the bagging with aggregation smooths out the errors and improves the ensemble performance.

4.2. PARF

RF, after the creation of multiple DTs, combines them to reach a single accurate and stable result. It provides a higher level of accuracy in predicting outcomes over the DTs and reduces the overfitting of datasets. The DTs created with Bagging can have a lot of structural similarities and finally high correlation in their predictions. On the contrary, the RF changes this procedure. Because of that, the sub-trees from a random sample are learned and, in this way the resulting predictions from all the subtrees have less correlation. For the RF, the bagging ensemble is used. Bagging chooses random sample with replacement from the entire training dataset. PA is applied to each dataset. DTs are created to fit each training set.

PARF is based on PA. Attributes are discovered randomly by bagging from the training set which PA uses to create DTs. To this end, PA works iteratively with the different random subsets. CNN criterion performs splitting operation to produce internal nodes repeatedly up to pure leaves. In this way, for each randomly selected feature, CNN discovers the most appropriate cut-point.

Generalization should be more successful by using ensemble's predictions instead of single PA's predictions. With the training of all the predictors the ensemble model will be able to predict a new instance with better accuracy using aggregation. According to the majority vote, the final choice will be the outcome of the most DTs.

4.3. PAET

The ET is similar to RF because of a random attribute selection but ET uses the whole dataset. For node splitting, the cut-points are randomly selected with the use of random thresholds for each feature. In the RFs, it is time consuming to grow a tree due to the fact that the best possible threshold needs to be found for each feature and for every node. On the contrary, the ETs are considerably faster for a training dataset because the splitting is selected randomly for each feature. In some cases, the obtained PAET results are better than the RFs' ones. The DTs are generated using PA with random splitting. The CNN criterion for purity is applied. ET reduces bias because the sampling refers to the entire dataset and the various data subsets might cause varying bias. Also, it reduces the variance resulted by the random node splitting in DT.

4.4. AB

AB is a boosting technique that aims at combining multiple weak classifiers to build a strong one. Weak learners in AB, named decision stumps, are DTs with a single split. AB puts more weight on hard to classify instances and less weight on the ones operating well. The stumps are produced for every feature iteratively and are stored in a list until a lower error is received. Weight assignment to each training example determines its significance in the training dataset. Weight update with a formula, at each iteration provides the stumps' performance. The AB trains predictors sequentially as happens in most boosting methods, where each predictor tries to correct its predecessor [15]. A major plus for both AB and boosting is that they seem to be very robust against overfitting. AB can be combined with other learning algorithms for performance improvement.

4.5. GB

Another boosting method is GB. This is an alternative to weights on training examples to convey the misclassification using the loss function based on residual (or negative loss gradients). It uses the residual errors to measure the amount of misclassification and define which training examples should be tested in the next iteration. Training examples correctly classified will have small gradients. GB consists of an additive model, a loss function and a weak learner.

5. Ensemble Performance Issues

An ensemble is a ML model that incorporates multiple model predictions. Ensemble learning methods are not always the most appropriate techniques to use or the best methods to use.

The bias and the variance of a model's performance are connected. We have a trade-off of bias and variance, and it is not hard to get a method with extremely low bias rather than high variance or vice versa. The use of hyperparameters can change the high bias or high variance [15] for some models and provide regularization (regularization by hyperparameters). In most cases, ensemble models provide a method to decrease the prediction variance. This reduction of variance provides improved predictive performance [16]. Since bagging tends to reduce variance, it provides an approach to regularization (regularization via bagging). This happens because although each learned classifier from f_1, f_2, ..., f_m is overfit on its own, may also be overfit to other various things. By voting, it can largely avoid overfitting. Bagging reduces variance and minimizes overfitting.

Bagging does not always offer an improvement. In models with low variances that perform well, the bagging can result in degrading the performance. The bagged decision trees are effective since each of them can fit on a different training dataset, which in turn allows to have less differences and, in this way, they make slightly different useful predictions. Bagging with DTs is effective because the trees have low correlation between predictions which means low prediction errors. The randomness used in the model construction can provide a slightly different model by running the same data. Working with bagging using randomness (stochastic learning algorithms) one technique is to evaluate them by averaging their performance with multiple runs or repeat cross-validation method. The latter technique is preferred in our experiments.

6. Bagging Hyperparameters

Bagging and random forest belong to homogenous parallel ensemble methods because they use the learning algorithm on the same dataset. Adaboost and GBoost, LightGBM, XGBoost belong to sequential ensemble learning algorithms.

The main difference between parallel and sequential ensembles is that the base estimators in parallel ensembles can be usually trained independently while for the sequential ensembles, the base estimator in the current iteration depends on the base estimator of the previous one. There are important points related

to the ensemble performance. The ensemble size, the base learner complexity and the learning rate are the most critical.

a. The ensemble size can be measured by the number of estimators. Three algorithms are used: the bagging with decision trees based on PA(PABAG), the PARF and PAET. The test errors can inform how well they do with the future data which is the generalization.

b. The base learner complexity is considered for the base decision trees, and it can be controlled by the maximum number of leaf nodes. The same three algorithms as previously are used as well.

7. Boosting Hyperparameter

The learning rate is another hyperparameter that is used to control the rate of the model learning to avoid overfitting. It shows how fast the model learns. It is time consuming, but it can control how quickly the complexity of the ensemble grows. Apart from avoiding overfitting it also offers generalization after training. The lower the learning rate the slower the model learns which means that the model becomes more robust and generalized.

However, it is possible for the weak learners to increase the tree depth in a sequential ensembles' methods, as boosting, to have a stronger classifier and the improvement of the performance. But to have an arbitrarily increment is not possible because there is the overfit during the training which in turn decreases the performance. A control mechanism is needed to prevent that.

8. Control Overfitting

Generally, increasing the complexity of the base learners it will make more difficult to reduce the variability of the ensemble. The complexity of the ensemble is increasing with the number of the base estimators which basically leads to overfitting. To avoid this bad situation, it is possible to stop the training process before it reaches the limit of ensemble size. This can be applied to the gradient boosted decision trees. The XGBoost is used to provide an efficient implementation of the gradient boosting algorithm. The early stopping decreases the training time, and this can be achieved by using fewer base estimators.

9. Simulation

The experiments are as follows.

9.1. Experiments

For the first 3 tests, various experiments with sonar dataset [20] are used with PA, PARF, PAET, and GB for classification. Through discretization continuous attributes are transformed to categorical ones by first providing the number of categories and then mapping their values to them. This facilitates split points creation for PA. Depiction with box plot diagrams is a streamlined way of summarizing the distribution of groups of data.

- In Figure 1 boosting with AB and default configuration shows higher accuracy compared to PARF, PAET, and GB ensemble methods.
- In Figure 2 kNN with low variance surpasses PA, SVM, and LR individual algorithms.
- In Figure 3 SVM using polynomial kernel offers slightly higher accuracy than ensemble method.

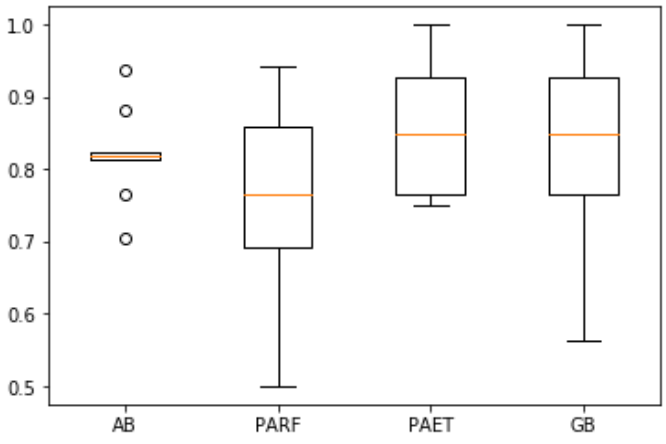


Figure 1: Ensemble algorithms' performance

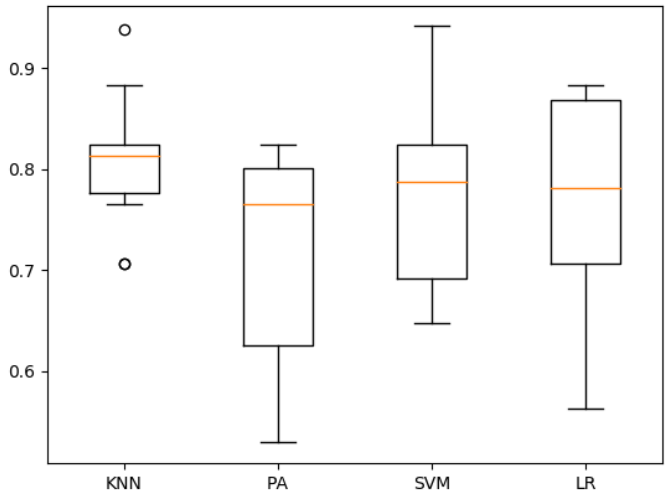


Figure 2: Algorithms performance

- In Figure 4 the size of estimators is examined for the ensemble performance (Figure 4). Three models are used in the experiment: PABAG, PARF, PAET. The Breast Cancer Wisconsin dataset (BCWD) from UCI site is used. The accuracy of each method over various numbers of estimators using a 10-fold cross validation is examined. All methods tend to perform similarly and yield high accuracies for over 10 estimators.
- In Figure 5 for the base estimator (decision tree), which is common for the three models, the tree depth is the most important measure of complexity since deeper trees are more complex. In Figure 5 the complexity of the models against the performance using the depth of the base decision trees is examined. Again here, a 10-fold cross validation shows average performance values in all tested methods over various maximum depths. All three methods tend to yield high accuracy values over all chosen depths but all of them obtain their best accuracies at a depth equal to 8. In general, we see that PABAG tends to yield larger accuracies over all other methods for various depths.
- In Figure 6 learning rate using XGBoost (boosting algorithm) for defining the appropriate rate for the performance. Cross Validation is used to set the learning rate. The XGBoost degrades the performance as the boosting process exhibits the

overfitting behavior. From Figure 6 the value of 1.2 or any value between 1.0 and 1.5 could be appropriate.

- In figure 7 control technique to stop the training with XGBoost. The accuracy is improving while the XGBoost continues training. When there is not any improvement of accuracy XGBoost terminates the training. In this way, the process terminates in a round which will be less than the predefined number of iterations. The training stops when the accuracy has better value after the next five rounds, From Figure 7 it is the 32 rounds when the training stops since the next five values are less than the one of 32 round (0.993449).

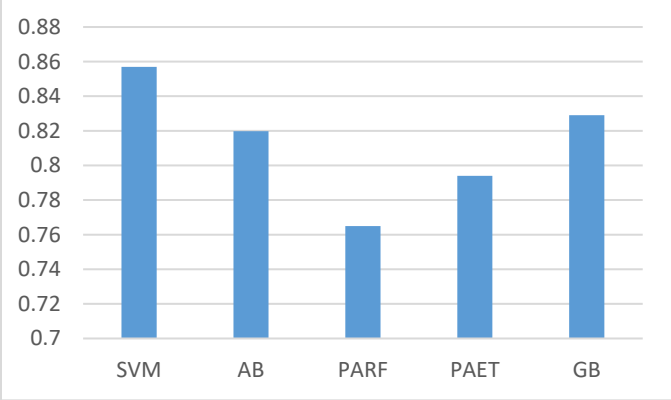


Figure 3: Accuracy with SVM and ensemble models

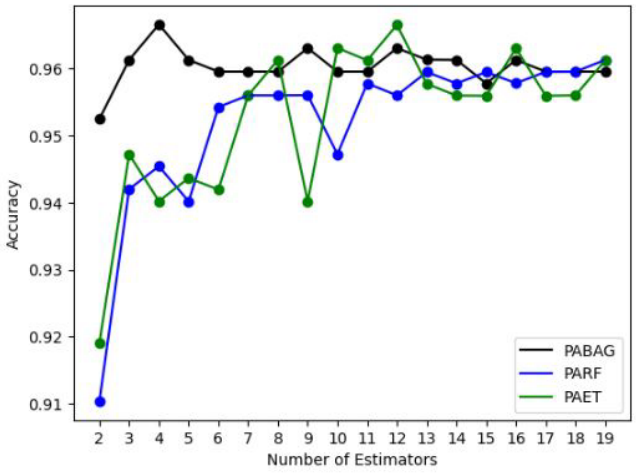


Figure 4: Performance vs Size of Estimators

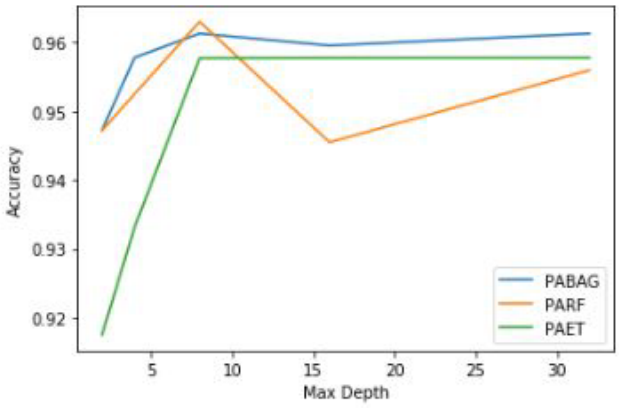


Figure 5: Accuracy vs Depth of trees in Ensemble

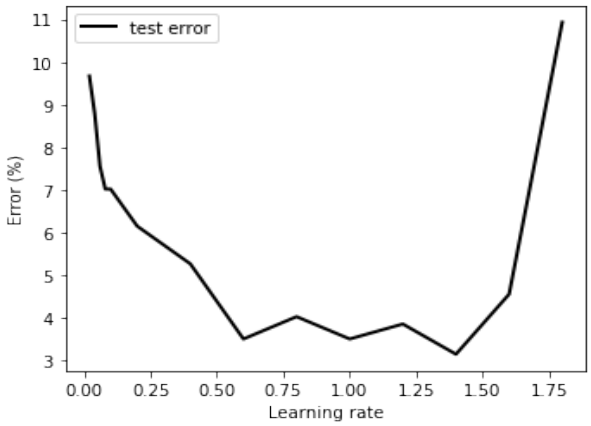


Figure 6: Learning rate discovery

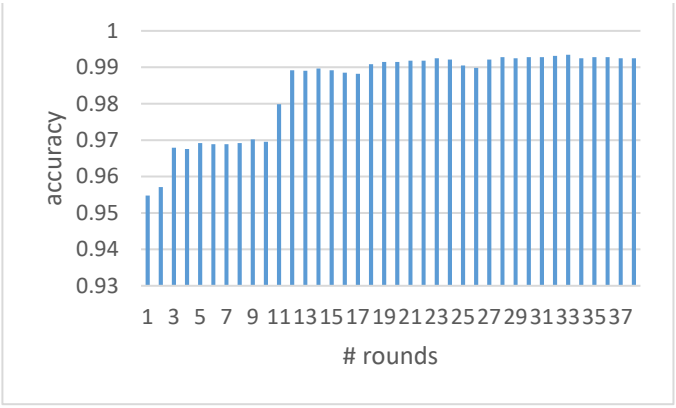


Figure 7: Accuracy vs # of rounds

10. Conclusion

Tree-based ensembles are considered state-of-the-art. The augmentation of the prediction power using individual algorithms or tree-ensemble models and their internal composition is an open issue. For this purpose, PA has been developed as the base model for the proposed ensemble since it avoids the replication problem of DTs by using the CNN criterion.

Ensemble learning helps improve overall accuracy by combining the results from several models. These models are known as weak learners trained to solve the same problem while their combination leads to more accurate and robust models. The kNN surpasses PA, SVM and LR. AB outperforms PARF, PAET, and GB. The SVM with the polynomial kernel exceeds even the ensemble models. Higher accuracy is achieved by an appropriate algorithm rather than ensemble models.

From the comparison of the two sets of algorithms the selection of models with their hyperparameters for creating an ensemble model could be an issue if other algorithms with their tuning parameters can provide good performance. The ensemble size and the depth of the trees affect the performance of the model. For boosting ensemble, the best learning rate with the use of XGBoost allows a good training strategy for the creation of the appropriate model. A reactive approach for terminating the training time by enforcing early stopping is also achieved.

Future work could be based on the stacking of heterogenous ensemble models.

References

[1] J. Tsiligaridis, "Tree-Based Ensemble Models and Algorithms for Classification", in 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIC 2023), 103-106, 2023, doi: 10.1109/ICAIIC57133.2023.10067006

[2] J. Han, M. Kamber, J. Pei, "Data Mining Concepts and Techniques",Morgan Kaufman, 2012

[3] M.Karntardzic,"Data Mining: Concepts, Models, Methods, and Algorithms", IEEE Press, 2003.

[4] L. Rokach, O. Maimon, "Data Mining with Decision Trees". World Scientific , 2008

[5] I. Nti, A. Adekoya, B. Weyon, "A comprehensive evaluation of ensemble learning for stock-market prediction" , Journal of Big Data , **7**(1), 1-40, 2020, doi: 10.1186/s40537-020-00299-5

[6] W. Yu, T. Liu, R. Valdez, M. Gwinn, M. Khoury, " Application of support vector machine modeling for predictiion of common diseases: the case of diabetes and pre-diabetes", BMC Medical Informatics and Decision Making, 10(1): 16, 2010, doi: 10.1186/1472-6947-10-16

[7] S. Huang, N. Cai, P. Pacheco, S. Narandes, Y. Wang, W. Xu, "Applications of Support Vector Machines (SVM) Learning in Cancer Genomics", Cancer Genomics&Proteomics,Journal, **15**(1), 41-51, 2018, doi: 10.21873/cgp.20063

[8] R. Couronne, P. Probst, A. Boulesteix, "Random Forest versus Logistic Regression: a large-scale benchmark experiment", BMC Bioinformatics, **19**, 270, 2018, doi: 10.1186/s12859-018-2264-5

[9] E.K. Ampomah, Z. Qin, G. Nyame, "Evaluation of tree-based Ensemble Machine Learning Models in predicting Stock Price Direction of Movement", Information Journal, MDPI, **11**(6), 332, 2020, doi:10.3390/info11060332

[10] A. Zafari, R. Zurita-Milla, E. Izquierdo-Verdiguier, "Land Cover Classification Using Extremely Randomized Trees: A Kernel Perpective" IEEE Geoscience and Remote Sensing Letters, **17**(10),1702-1706, 2020, doi: 10.1109/LGRS.2019.2953778.

[11] L. Li, J. Qiao, G. Yu, L. Wang, H. Li, C. Liao, Z. Zhu, "Interpretable tree-based ensemble model for predicting beach water qaulity", Water Research, 211, 118078, 2022, doi:10.1016/j.watres.2022.118078

[12] M. Jaxa-Rozen, J. Kwakkel, "Tree-based ensemble methods for sensitivity analysis of environmental models: A performanced comparison with Sobol and Morris technique", Science Direct, **107**, 245-266, 2018 doi:10.1016/j.envsoft.2018.06.011

[13] P.Han, M.Steinbach, A. Karpatne, V. Kumar, "Introduction to data Mining",Pearson, 2019

[14] A. Geron, "Hands On Machine Learning with Scikit-Learn & Tensorflow",O'REILLY,2017

[15] J.Gareth, D. Witten, T. Hastie, R. Tibshirani, ".An Introduction to Statistical Learning with Applications in R", Springer, 2017

[16] L.Rokach,"Ensemble Learning: "Patern Classification Using Ensemble Methods", World Scientific, 2009

[17] R. Natras, B.Soja, M. Schmidt, " Ensemble Machine Learning of Random Forest, AdaBoost, and XGBost for Vertical Total Electron Content Forecasting ", Remote Sensing, MDPI, **14**(15), 3547, 2022 doi:10.3390/rs14153547

[18] M.Ramakrisna,V.Venkatesan,I Izonin, M.Havryliuk, C.Bhat, "Homogenous Adaboost Endemble Machine Learning Algorithms with Reduced Entropy on Balanced Data", Entropy, MDPI, 25, 245,2023, doi:10.3390/e25020245

[19] B. Kalaiselvi, S. Geetha, "Ensemble Machine Learning AdaBoost with NBtree for Placement Data Analysis", in 2nd International Conference on Intelligent Technology (CONIT) , 1-4 , 2022, doi: 10.1109/CONIT55038.2022.9847993

[20] Sonar dataset: http://archive.ics.uci.edu/ml/datasets/connectionist+bench+(sonar,+mines+vs.+rocks)

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 6, 26-36 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Social Media Text Summarization: A Survey Towards a Transformer-based System Design

Afrodite Papagiannopoulou*, Chrissanthi Angeli

School of Engineering, University of West Attica, Athens, 122 43, Greece

ARTICLE INFO

Article history:
Received:19 September, 2023
Accepted:13 November, 2023
Online: 30 November, 2023

Keywords:
Social Media Summarization
Natural Language Generation
Neural Networks
Transformers

ABSTRACT

Daily life is characterized by a great explosion of abundance of information available on the internet and social media. Smart technology has radically changed our lives, giving a leading role to social media for communication, advertising, information and exchange of opinions. Managing this huge amount of data by humans is an almost impossible task. Adequacy of summarizing texts is therefore urgently needed, in order to offer people knowledge and information avoiding time-consuming procedures. Various text summarization techniques are already widely used. Artificial intelligence techniques for automated text summarization are a major undertaking. Due to the recent development of neural networks and deep learning models like Transformers, we can create more efficient summaries. This paper reviews text summarisation approaches on social media and introduces our approach towards a summarization system using transformers.

1. Introduction

Today's abundance of information on the Internet and social media makes it imperative to create summaries to keep readers informed in an accurate and timely manner. Social media posts are based on events. When an event breaks out, a huge amount of posts flood social media platforms. These posts can be documents, articles, discussions and conversations on different topics and events, that are time-consuming to read. Therefore, text summarization is essential for retrieving useful knowledge in a reasonable amount of time.

We can proceed summarization in two ways: a) writing a summary manually, which is time-consuming, b) writing algorithms and performing artificial intelligence techniques, which requires much less time. This method is called Automatic Summarization, whose aim is to create a shorter text of the original document, without losing its meaning intact [1,2]

Many algorithms have been applied to generate text summaries. We can consider two main classes of algorithms: Pre-neural, that do not make use neural networks and b) Deep Learning or Machine Learning techniques that make use of neural networks. The later have been successfully applied to various NLP tasks, yielded excellent results, and have been extensively used in recent years [3]. Various deep learning models have been used, mainly based on both recurrent neural networks (RNN) and convolutional neural networks (CNN). These models have proven highly satisfactory in predicting complex relationships that simple structured or semantic approaches cannot do alone [4]. In recent years, the use of Transfer leaning and Transformer models has gained popularity mainly in the field of Natural Language Understanding, Processing and Generation [5].

1.1 Text Summarization Methods

The aim of text summarization is to convert a large text document into a shorter one by preserving the critical information and ensuring the meaning of the text. Due to large amounts of data available on social media, it is almost impossible for anyone to read all the comments generated by the users under a post. It is therefore necessary to properly code and program machines so that they can create coherent summaries just like humans. The process of text summarization done by machines or artificial intelligence programs is known as "Automatic Text Summarization". Automatic text summarization is the process of generating a text summary which, using artificial intelligence techniques and deep learning algorithms, is based on the original text while preserving the information content and overall meaning [1, 6]. Automatic text summarization presents some challenges: First, appropriate information must be selected from the original document so that the summary preserves the meaning of the document. Second, the output text must be fluid and coherent and also expressed in a way that is direct and understandable to the

*Corresponding Author: Afrodite Papagiannopoulou, apapagiannop@uniwa.gr

reader [2]. We recognize two main categories of text summarization:

- *Extractive summarization* extracts important words and phrases from the original text, uses them as they are and, gathering them together, with a slight rearrangement, generates the summary. [1].
- *Abstractive summarization* abstracts the meaning of the original text and, using paraphrasing, creates a new shorter text which looks completely different but essentially has the same meaning as the original text [1].

1.2 Importance of Social Media Summarization.

The role of social media is becoming increasingly important, especially in public life, as posts and comments made by users are important in shaping public opinion on various issues. These topics refer to economics, politics, entertainment, education, psychology and society. Social media data excels at extracting sentiment and patterns of social behaviour that can be used for social research, economic and political decision-making. So creating summaries seems vital. There is a lot of information flooding the internet and most of it is redundant or repetitive resulting in confusion for readers. For this reason, a mechanism is needed to select the necessary information to provide correct, fast and accurate information.

Recent years have shown that abstractive summarization has achieved great results in the field of document summarization by producing more human-like summaries. Unlike formal documents, online conversations and web communication present three great challenges: 1) tend to be informal, consisting of slang expressions and special characters, 2) show deviations from the original theme and dependencies on previous opinions and, 3) since they are short, they lack lexical richness.

It is important to emphasize that, in the field of social media summarization, few research works [7, 8, 9] have demonstrated transformer-based social media abstractive text summarization systems. Our system aims to optimize the results achieved so far in two main points. First, we focus on creating coherent and meaningful summaries of user comments since they present greater specificity, in terms of: a) the nature of the language, b) their conceptual grouping and c) their dynamic upgrading. Second, unlike other research works that consider only two social media platforms, Reddit and Twitter, our goal is to explore how different data sets and training models can be applied to more social media platforms.

This paper is a survey of social media summarization techniques and approaches as well as a presentation of our work toward a system that generates summaries of user comments on social media posts using transformers. The rest of the paper is organized as follows: Section 2 is a literature review of the research that has been carried out so far in the field of text summarization. Section 3 illustrates the importance of transformers against previous deep learning methods and gives a basic description of transformer model architecture. Section 4 depicts the motivation and design of our system and compares 3 pre-trained transformer models based on a selected dataset. Section 5 shows the results of the comparison between 3 pre-trained model pipelines and Section 6 contains concluding remarks and our future work.

2. Literature review

In the field of Automated Text Summarization on Social Media, several attempts have been made by researchers as seen in Table 1. APPENDIX I. We have grouped them in two main categories: 1) Pre-neural approaches that do not use neural networks and 2) Deep learning approaches that use neural networks. A further division of deep learning approaches has been done, based on transformer architecture, into the following three categories:1) *Social media summarization systems that do not use Transformer architecture,* 2) *Transformer-based Social media summaries* and 3) *Transformer-based projects that do not generate social media summaries.* Finally, recent trends on large language models (LLMs) lead us to prompt engineering that give a new perspective to the fields of text generation and summarization. A more detailed presentation follows:

2.1 Pre-neural approaches

In the early stages of natural language research, summarizing web posts, microblogs, and social networks researchers used probabilistic and optimization methods to generate a summary from different types of data such as text, videos, images, hashtags and so on. In some cases the summary is not only text but a multimedia representation.

A typical example of multimedia representation is the model proposed by [10] & [11]. This model takes as input real-time user interaction in the social media stream and provides a multimedia representation as output. The summarization process takes minimal linguistic information into account. The output is a multimedia representation combining photos, video and some text. The algorithm uses an optimization process under a score-based input to find the topic of greatest importance. Specifically, for each Multimedia Social Network (MuSN), they use graph-based modeling and influence analysis methodologies to identify the most important media objects related to one or more topics of interest. Then a media summary is extracted from a list of candidate objects. This is achieved with the help of heuristic models based on the following properties: Priority, Continuity, Variety and Non-Repeatability. This project does not follow one of the text summary models, either extractive or abstract, because it does not create a text summary.

A second example of multimedia summarization is the model proposed by [12] which is a framework for generating a visualized summary from microblogs with multiple media types. The main goal of this work is to separate the different parts of an event published on social media by creating sub-events. Then for each secondary event it creates a clear and precise summary. Thus a fragmentary summary is produced. Each microblog is taken as a separate information mining unit, so a microblog digest is considered a multi-document digest (MDS). MDS faces two challenges: 1) Word limit and 2) Unstructured social media content. The proposed framework consists of 3 stages: 1) Noise removal in which the authors apply a filtering model to clean data that calculates the probability that the data is irrelevant. 2) Sub-event extraction where the authors propose a multi-media probabilistic model, called Cross-Media-LDA (CMLDA), which explores the associations between the different media types of the published event or separates the sub-events. 3) Creation summary. An algorithm is implemented to identify representative microblog texts considering three criteria: vagueness, importance and

diversity. Based on these criteria, the summarizing ability is also measured. The important point of this work is that the knowledge acquired in the previous stages is used, thereby reinforcing each other for the proper functioning of both the textual and visual summarizing mechanisms.

Another research work that considers each post as a separate text is the work of [13] & [14]. This model generates a summary using Twitter posts by performing a search on the Twitter API based on trending topics. It automatically generates an abstract based on multiple papers published on the same topic by applying a Phrase Reinforcement (PR) algorithm by adding the TF-IDF technique to it. The algorithm performs the following steps: 1) finds a specific phrase of each document that matches the topic. This is done on the one hand because users often use the same word or phrase to describe an event that usually matches the title of the event and on the other hand because very often the most relevant post is "retweeted". 2) Based on this phrase it starts asking Twitter.com to retrieve related posts, 3) filters posts based on related content. The longest sentence from each post is isolated. The collected sentences become the input to the PR algorithm, 4) creates a graph with a common root node representing the topic sentence. Next to the root node there are other nodes, the word nodes and the word count nodes that show how many times the word appears in the sentence, 5) the PR algorithm searches for the path with the highest word frequency (count) starting from the root node to the final. Following this path the words are combined and an initial summary is produced. The algorithm iterates using the partial summary each time as the root node and reconstructs the graph, arriving at the final summary.

In the work [15] the authors create a model that aims to summarize tweets that have sports as their topic. The key point in their research is finding real-time events, which is of particular importance in repetitive and real-time events such as sports. The design of the model includes two stages: 1) Application of the modified Hidden Markov Model according to which the timeline of events is segmented, based on the tweet-stream and the distribution of the words used in the tweets. Each part is a separate "sub-event" of different importance. 2) Selection of tweets that can provide information about the section that was considered most important.

A framework that searches for categories of topics on Twitter and then generates summaries of tweets corresponding to that topic is presented in [16]. To determine the topic categories to use in summarization, the time sequence of the tweet event is taken into account. An important factor is the detection of tweets written at the same time for an underlying event to extract important data. To this end they apply two topic models: the Decomposition Topic Model (DTM) and the Gaussian Topic Model (GDTM) that exploit the temporal correlation between tweets under predicted conditions to produce a concise and information-rich summary of events.

2.2 Deep Learning Models

Since deep learning algorithms require a ton of data to learn from, this increase in data generation is one reason deep learning capabilities have grown in recent years. Additionally, deep learning algorithms benefit from the most powerful computing power available today, as well as the proliferation of Artificial Intelligence (AI) as a Service. Deep Learning has achieved significant success in various fields, such as computer vision and natural language processing. In the following paragraphs we present models that make use of neural networks to generate summaries. Various deep learning models have been employed, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and RNNs with Long-Short Term Memory (LSTM). In addition, recent years have seen the emergence of new techniques in abstractive summarization, such as those based on transformers. The specificity of these techniques lies in the use of an attention mechanism and not a recurrent or convolutional neural network.

The idea of transformers was first introduced in the paper "Attention is All You Need" [5] where transformers achieved state-of-the-art results for translation tasks over previous NLP models. The idea has been extended to all NLP tasks and many researchers have adopted transformer models to build their own applications. The social media summarization field is in the early stages of research. Referring to the literature we distinguish, so far, the following aspects:

- *Social media summarization systems that do not use Transformer architecture*.

The authors of the paper [17] introduce a project which summarizes social media posts by incorporating user comments in order to capture the reader focus. They use a model named reader-aware summary generator (RASG). Their model is based on a sequence to sequence framework rather than transformer architecture to generate abstractive summaries. It consists of 4 main parts: 1) Sequence-to-sequence summary generator, 2) an attention module which captures the reader's focus, 3) a supervisor module that matches the semantic gap between the generated summary and the reader's focus and 4) a goal tracker that defines the goal of each generation step. Specifically, the seq2seq architecture with an attention mechanism is used to generate summaries. First, words from user comments are correlated with those in the document. This correlation is then calculated because it is considered an important indicator of the "reader-centered aspect". At the decoding layer, attention weights are defined as the "decoder focused aspect" of the produced summary. Finally, a supervisor calculates the deviation of the "reader-centered view" and "decoder-centered view" with the goal of reducing the deviation.

Another attempt of text summarization that is related to social media posts and customer reviews is presented in [18]. The main idea is the use of a Sequence-To-Sequence Model with Attention approach to produce abstractive text summarization with no use of transformers. It is based on bidirectional RNN with attention layer on input data. The model uses also LSTM as encoder and decoder to build summaries. LSTM encoders partly extract a fixed length of sequence of the input and LSTM decoder generates a translation of this sequence. With the attention layer in this architecture the decoder is allowed to have more direct access to the input sequence.

The authors of [19] proposed a summary-aware attention model that aims to generate abstractive text summarization in social media. This mechanism helps the decoder to make the right decision. The authors have based this novel proposal on two important limitations that have been faced on previous models: 1)

the original text is not structured, like social network texts, making the decoder's task difficult, and 2) the calculation of attention weight does not take into account the summary information created in previous layers, which negatively affects the consistency of the summary. To overcome the above limitations, a new model of attention computation is proposed, called "summary-aware attention" mechanism for abstractive summary generation. According to this, attention is calculated at each level, taking into account the data of the previous level.

The work of [20] generates an abstractive text summary on social media, in Turkish. The data is collected and the Word2Vec model is used to preserve the meaning of the text, which is a very important factor for generating summaries. The GRU (Gated Recurrent Unit) neural network is then applied to divide user posts into positive and negative. Finally, using the Latent Semantic Analysis (LSA) method, a summary text in Turkish is generated from user comments on Twitter. The LSA algorithm accepts as input the content of a text and reveals the hidden semantic relationship between terms and sentences of that text.

The model of [21] produces an abstract digest of social media text based on Attentional Encoder-Decoder recurrent neural networks. The model is based on the encoder-decoder architecture by enhancing it with an additional hidden layer before the encoder unit. In this layer it is decided which information is useful and which should be removed. In this way, invalid information is better filtered in the encoder unit and through the reinforcement learning policy, better results are achieved. More specifically, this work meets the following three main contributions: (i) They have applied to summarization, the attentional RNN encoder-decoder originally developed for machine translation, showing that it already outperforms technology systems in two different English corpora. (ii) Since the model that is already used for machine translation addresses specific problems when applied in summarization model, the authors propose new models and show how additional performance is improved. (iii) Finally, they introduce a new dataset for the multi-sentence abstract summation task and establish benchmarks. The authors also propose a variety of selective gating methods to better filter the information while preserving the meaning of the original text.

- *Transformer-based Social media summaries*.

An abstractive Event Summarization on Twitter [8] framework uses a pre-trained BERT-based model as an encoder and a not pre-trained transformer model as a decoder. An event topic prediction component helps the decoder to focus on more specific aspects of posts. The most liked comments are used to produce more coherent summaries. The framework consists of the following features: 1) a mechanism that selects the most important tweets to be used by the decoder that is implemented in the following steps: a) tweets are selected based on their timestamp in ascending order. b) After the initial selection, in order to determine their importance, each tweet will be compared to each of those already selected. Selection is based on how representative and informative a tweet is. This gives the summary precision and coherence. 2) a mechanism that predicts the thematic category of the event. Topic information is used by the decoder to generate different summary styles from different events, thereby focusing on very specific aspects of topics. 3) A BERT model as encoder, to better exploit the grammatical and semantic information of tokens pre-trained model and Transformer-based model as decoder. 4) two separate optimizers to smoothly integrate the pre-trained BERT model and the untrained transformer.

A new model is introduced in [22] which generates summaries of each text posted on a Chinese web platform. The model combines BERT, reinforcement learning and other technologies. Their work based on LCSTS, a dataset constructed from a Chinese social media platform named "Sina Weibo".

What is highlighted in [9] is the importance of summarizing user feedback. In their paper, they propose a multi-document text summarization scenario that includes meaningful comment detection, extending previous single-document approaches. They use BERT as the encoder and the transformer-based model as the decoder and produce an abstract summary of the embedded news discussions. This architecture is extended with an attention encoding level that is fed with user preferences. Attention encoding focuses on comments with the highest social impact. Consequently, the model summarizes the most relevant aspects of a discussion gathered from a news article. To encourage the model to pay attention to important comments, they introduce a data-driven strategy that focuses the model on those comments. The key element of this research is that to summarize the discussion there is the approach of many works by many authors. This has the consequence that users are considered co-authors of the original news story, to which they can add new information. Comments are rated as important based on user preferences.

A Transformer-based abstractive summarization is presented in [7] which creates summaries in three languages from posts and comment pools. The datasets are from Reddit and Twitter. The authors fine-tune T5 and LongFormer, test them against BART, and consecutively test them to pools of comments. In addition, they apply enlarged Transformer-based models on Twitter data in three languages (English, German, and French) to discover the performance of these models on data with non-English text. The main part of this work is the comparison of open source models for social media summarization. Three stages are applied: 1) Performance is optimized and one of three models is selected: LongFormer, Bart and T5. 2) The selected model is tested on pools of comments from Reddit and evaluated based on the similarity of posts and comment summaries. 3) The models are also applied to Twitter data in three languages (English, German and French). This study concludes that using transformer-based models, such as LongFormer and T5, give better results in generating summaries on social networks.

In [23] model for extractive summarization with transformers is introduced. This system summarizes a single web document by exploiting relevant information of social media to enhance the document summarization. The intuition behind this model is that they make use of relevant user posts and transformers to enrich the sentences of the output summary. They stack a Convolutional Neural Network (CNN) on the top of transformers (BERT) for classification.

- *Transformer-based projects that do not generate social media summaries.*

In [24] a two-stage transformer-based approach is proposed to generate abstractive summaries from articles in Chinese. The model produces fluent and variable length abstractive

summarization according to the user's demands. It consists of two modules: text segmentation module and two-stage transformer-based module. A pre-trained BERT model and a bidirectional LSTM are used first to divide the input text into segments. Subsequently the extractive based BERTSUM model is constructed in order to extract the most important information of the segments. Then, collaborative training is used to train the two-stage Transformer-based model. The training process begins with the document Transformer in the second stage. The input of the document Transformer is the outputs of the extractive model and the output of the document Transformer is the headline summary.

An abstractive summarization system of meeting conversations using transformers is presented in [25]. The system takes as input human dialogues and by applying the classic Transformer model summarizes the dialogues.

In their work [26] compare three pre-trained transformer-based models in an attempt to produce abstractive summarization of news articles on web. The pre-trained models used in this project are BART, PEGASUS and T5. When the transformer-based pre-trained language models are fine-tuned they give satisfactory results and fluent summaries. After evaluating each model with ROUGE the T5 model outperformed all other models.

Authors of [27] use dataset from Wikihow knowledge base and then deploy bidirectional encoder representations from transformers (BERT) and text-to-text transfer transformer (T5) models in order to produce abstractive summaries of the articles and then use the ROUGE scores to assess the performance of the two models and compare them.

2.3 Prompt Engineering

In the field of natural language processing (NLP), Prompt Engineering has emerged as one of the most innovative and powerful techniques for improving the performance and adaptability of language models. This recent technique has its roots in the development of LLMs whose history dates back to the 1950s, but until the introduction of BERT and GPT models in 2018, LLMs had the main role in NLP tasks. These models still use complex algorithms and massive amounts of training data to understand natural language. Their ability to learn patterns and structures in language has proven invaluable in various NLP tasks. However, due to the significant time and computational resources required to train these models, researchers have turned to prompt engineering. By designing and configuring the appropriate prompts, which can be shaped to improve behaviour and results of these models to achieve specific tasks. These tasks involve extracting information, summarizing text, answering questions, generating code, and classifying text. The detailed reference to this technique is beyond the purposes of this work, as in the current phase of our project we will use transformer models as presented in the following paragraphs.

3. Transformer models and methodology

3.1 Transformers and Model Explanation

Transformer models are a new development in machine learning and neural networks based on transfer learning. The Transformer model was first proposed in the paper "Attention Is All You Need." [5]. These models can handle the context of a text exceptionally well forming a state-of-the-art architecture in various NLP tasks such as translation, text generation, summarization, question-answering, classification and sentiment analysis. Transformer models are significant because:

- They have given exceptional results in many natural language tasks that use sequential data.
- Unlike recurrent neural networks (RNNs), they support parallelization and, they have replaced repetition with attention. Therefore calculations can be done simultaneously.
- Their architecture allows for the creation of better-quality models because they can handle better long distance dependencies. Based entirely on the attention mechanism, they solve the problem of Recurrent Neural Networks with Long-Short Term Memory (LSTM) and CNNs, which is mainly the inability to model larger sequences without data loss.

The architecture of the transformer model inspires from the attention mechanism used in the encoder-decoder architecture in RNNs to handle sequence-to-sequence (seq2seq) tasks. Unlike RNNs the transformer eliminates the factor of sequentiality meaning that, it does not process data in sequence which allows for more parallelization and reduces training time. The transformer works as encoder-decoder architecture because it is composed of two large building blocks: an encoder and a decoder. As we can see in (Figure1) the encoder is on the left and the decoder is on the right. Both encoder and decoder can consist of a stack of Nx identical layers (in original paper Nx = 6). Since the model does not contain any recurrence or convolution, it adds a positional encoding layer at the bottom of the encoder and decoder stacks to take advantage of the order of the sequence. Each layer mainly consists of Multi-Head Attention and feed forward layers. Inputs and outputs are first integrated in an n-dimensional space [5].

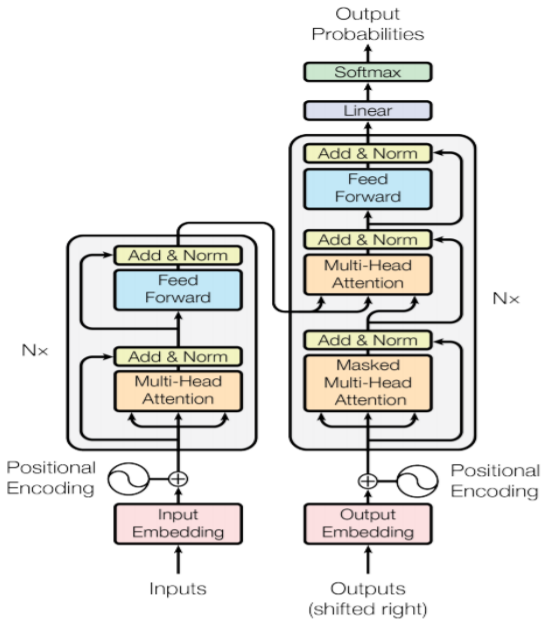


Figure. 1: Transformer Model Architecture

Encoder: The encoder is comprised of two major components: a multi-head attention (self-attention) mechanism which is followed by normalization and a feed-forward neural network.

- Inputs: Since transformers, as any other model, do not understand natural language, the input text is tokenized and processed to convert every word into a unique numeric id.
- Embedding layer: In the embedding layer the transformer uses learned embeddings to transform the input tokens into vectors of dimension d = 512.
- Positional encoding: A positional encoding is a fixed-size vector representation that provides the transformer model with information about where the words are in the input sequence, since there is no use of any recurrence or convolution.
- Multi-head Attention and Self-Attention: Self-attention is a mechanism that captures the contextual relationship between the words of a sentence. Based on a scale dot-product attention, it creates a vector of every input word. The main role of this vector is to understand how relevant every word in the input sentence is, with respect to other words in the sentence. Since self-attention is not only applied once, but also several times (in the original paper it is applied 8 times), is called multi-head attention. The objective is to generate several attention-based vectors for the same word. This helps the model to have different representations of the words' relations in a sentence. The different attention-based matrices generated by the multiple heads are summed and passed through a linear layer to reduce the size to a single matrix.

Decoder: The decoder side has a lot of shared components with the encoder side. The decoder takes in two inputs: (a) the output of the encoder and the output text shifted to the right. Then it applies multi-head attention twice with one of them being "masked". The final linear layer in the decoder has the same size of words as the target dictionary. Each unit will be assigned a score. The softmax function is applied to convert the scores into probabilities indicating the probability of each word to be present in the output.

4 Systems design

4.1 Motivation

Automatic Text Summarization in both formal documents (such as articles and books) and Social Media texts are grouped into two main categories: Extractive and Abstractive summarization. Initially, researchers focused on creating summaries using the extractive summarization method. According to this technique, the summarizer extracts important words or sentences from the original document. In order not to lose its meaning it uses statistical and linguistic methods. It then rearranges slightly these words and phrase, to produce the output summary. This technique, while simple to apply, does not produce satisfactory results, as the summary is not coherent and accurate. This effect is enhanced when the length of the sentence starts to grow.

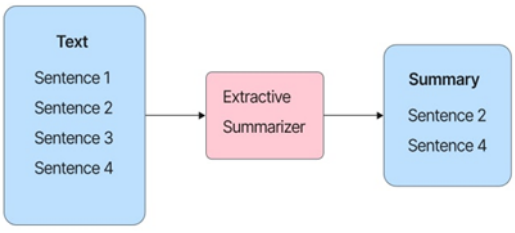


Figure 2: Extractive Text Summarization

In recent years, researchers have turned to the abstractive summarization technique. In this technique the summarizer extracts the meaning of the original document and the output document is produced without losing this meaning. This process adopts the way the human mind produces summaries and despite its complexity, it tends to give better results. All of the research papers that applied the Abstractive Summarization technique used deep neural network methods and attention mechanisms to train their models, adopting the way the human mind might be trained to summarize documents.

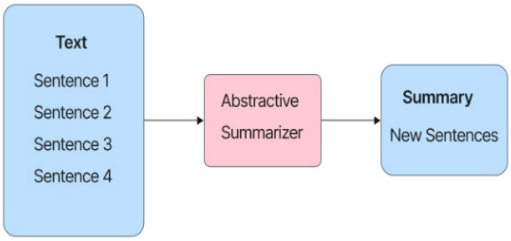


Figure 3: Abstractive Text Summarization

The area of social media summarization is in its early stages and remains challenging due to the particularity of social media content. As already mentioned in the literature review, various models, based either on pre-neural or deep learning techniques, have been proposed giving interesting results, but also facing several limitations. Transformer-based neural networks try to overcome these limitations giving state-of-the-art results. The perspective of our work concerns the following areas:

- Transformer-based models.
- Abstractive summarization techniques
- The discussions of users under the post

4.2 Methodology

As we have mentioned earlier Social Media posts and user comments provide great challenges on text summarization: 1) tend to be informal, consisting of slang expressions and special characters, 2) show deviations from the original theme and dependencies on previous opinions and, 3) since they are short, they lack lexical richness, 4) due to the massive amount of comments generated under each post, many of them are repetitive. It is therefore considered necessary to use data pre-processing techniques in order to improve the performance of the model. The original idea of our approach is to use the pre-trained transformer models. For a correct decision we fed with data 3 pre-trained pipelines of T5, BART and PEGASUS models comparing the generated summaries (Figure. 4). All three of the aforementioned pre-trained models generate abstract summaries and use the

encoder-decoder transformer architecture that performs best in text generation tasks. Our model uses a Facebook dataset of news posts and their corresponding comments and consists of the following steps:

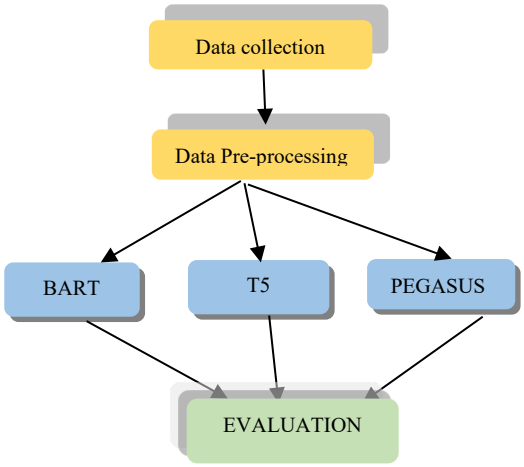


Figure. 4: Pre-trained model evaluation

- Data collection and pre-processing using Python libraries.
- Feed this text into the transformer-based encoder.
- Clustering and attention rendering based on the topics.
- Feed this encoding to the Transformer decoder and
- Produce the summary text

4.3 Data Collection and pre-processing

Finding the right datasets from social media is a really complex task. Especially for data that will be used for summary. Several social media platforms are limited in receiving data, and in those that are open, the data must be reformatted to meet the needs of the project. The dataset we are looking at for this project is Facebook news posts accompanied by user comments below each post. The raw data has 7 columns, namely 'created_time', 'from_id', 'from_name', 'message', 'post_name', 'post_title', 'post_num' and 1,781,576 rows. The dataframe consists of the 3 columns: 'message', 'post_title', 'post_num' where the 'post_title' as well as the 'post_num' identify a specific post while the 'message column represents the user comments under the post. We focused on creating summaries of user comments, as they are more discrete in terms of (a) the nature of the language and (b) their conceptual grouping. After downloading the dataset and reshaping it in a meaningful way for summary generation, the first step is to apply some basic pre-processing before building the model. Using impure and unclean data does not lead to issuing the desired results. Hence in this step, using python packages NLTK and regex, we have removed punctuation, special characters, emoji and link threads, as long as NULL values. The second step is to perform data clustering according to the topic of the discussion. Since we on the first stages of our ongoing research work we need to decide which of the pre-trained models can be better trained and fine-tuned based on our dataset. Therefore, the cleaned and classified data is first fed to the pre-trained-models and the results of the comparison are extracted.

4.4 Pre-trained Transformer models and pipelines

Today excellent results have been achieved in all tasks thanks to the wide variety of pre-trained transformer models available. Pre-trained language models have been trained on large-scale datasets and due to their capabilities and recourses can be reused even with small amount of datasets. PTLMs are differentiated based on their architectures and pre-training tasks. In order to make a correct choice of the appropriate model for a particular task on a particular dataset, knowledge of these details is crucial. For example, BERT is better suited for task understanding than task generation. Furthermore, another differentiation of pre-trained models is their architecture. There are many models whose architectural backbone is Transformers, but some of which pre-train only the encoder, such as BERT and UniLM, while others pre-train only the decoder, such as GPT. Therefore, PTLMs must be chosen wisely based on these details when adapting them to downstream tasks. Based on the above, in order to proceed with the correct model selection, we have set the following criteria:

- Both input and output of the system are text
- The output will be abstractive summary generation, which means that we chose the models that produce abstractive summaries and are based on the Encoder-Decoder architecture [28].
- All the pre-trained models can fine-tuned on our target task according to the limitations of our datasets.

The HuggingFace hub [29], acts as an open source library and provides a huge number of pre-trained models as well as datasets, on a wide range of different NLP tasks. These models can be used to predict a summary and then accurately fit any data set. Due to the large number of models available in Hugging Face, it is quite difficult to decide which one to use as the best one for our dataset in order to produce the desired results. Pipelines are a fast, easy and efficient way to use different pre-trained models and can be applied to many NLP tasks. The pipelines meeting the above criteria are BART, T5 and PEGASUS and are presented thoroughly below:

- BART – stands for Bidirectional and Auto-Regressive Transformers [30]. BART supports tasks involving Natural Language Creation, such as the summarizing task, which requires new text to be generated to summarize a large input text, unlike the traditional summarizing task that extracts some parts of the input text to make the summary. It combines the pre-training processes of BERT and GPT within the encoder-decoder architecture (Figure 5). Input sequences undergo one of many possible transformations,

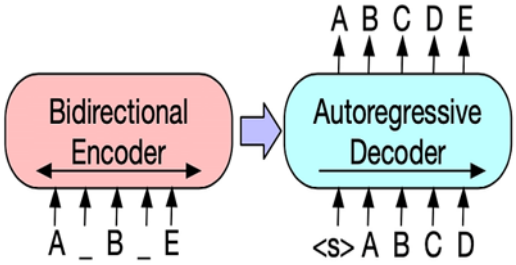


Figure 5: BART (source original paper)

from simple masking to model sentence permutation, token deletion, and document rotation. These modified inputs are passed through the encoder and the decoder must reconstruct the original texts. This makes the model more flexible, as it can be used for NLU as well as NLG tasks, and achieves top performance in both. BART pre-training involves 2 steps: a) the input text is free of unnecessary noise, which may change the length of the text. b) the seq2seq model learns to reconstruct the original text from the corrupted text. BART is known for its excellent performance in tasks that require complex language handling, such as text summarization and machine translation. BART's pre-training task encourages the model to learn representations that are robust to noise and variations in the input text. This makes BART suitable for tasks that require handling text that is noisy, ambiguous, or written in different languages. It has a more complex architecture, which makes it more suitable for tasks that require handling large sequences of text or that require text generation.

- T5 - stands for "Text-to-Text Transfer Transformer" [31]. The T5 model is a task-agnostic model meaning that unifies all NLU and NLG tasks by converting them into text-to-text tasks, such as Translation, Language Inference, Information extraction and Summarization. All tasks are framed as sequence-to-sequence tasks, where adopting encoder-decoder architecture is natural thus it's capable of handling architecture uses the original Transformer architecture different sequence lengths for input and output. The T5 using the large crawled C4 dataset, the model is pre-trained with masked language modelling as well as the SuperGLUE tasks by translating all of them to text-to-text tasks. The largest model with 11 billion parameters yielded state-of-the-art results on 17 of the 24 tasks tested. It achieved high scores on Corpus of Linguistic Acceptability (CoLA), Recognizing Textual Entailment (RTE) and Natural Language Inference on WNLI tasks. It also achieved very high performance on the SQuAD dataset but less satisfactorily on the CNN/Daily Mail abstractive summarization task. The pre-training process involves two types of training: supervised and self-supervised. During supervised training, downstream tasks from the GLUE and SuperGLUE benchmarks are used and tasks, as explained before. On the other hand, self-supervised training is done using corrupted tokens. This converted into text-to-text is achieved by randomly removing 15% of the tokens and replacing them with individual sentinel tokens.The encoder takes the corrupted sentence as input, while the decoder takes the original sentence as input. The target is then the dropped-out tokens, delimited by their sentinel tokens.

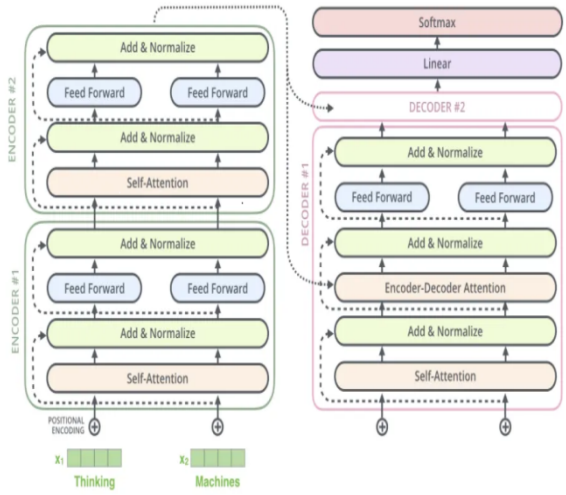


Figure 6: T5 model (source original paper)

- *PEGASUS* – stands for Pre-training with Extracted Gap-sentences for Abstractive Summarization Sequence-to-sequence models [32]. It is also based on the Transformer architecture but with some modifications for abstractive text summarization. It is a task-specific architecture where every component in pre-training closely maps text summarization. PEGASUS uses a gap-sentence generation task, where whole encoder input sentences are replaced by a second mask token and fed to the decoder, but which has a causal mask to hide the future words like a regular auto-regressive transformer decoder. It is designed explicitly for abstractive text summarization, where it generates summaries that have the same meaning but different phrases from the source text. The model is pre-trained on CNN/DailyMail dataset, which consists of large volumes of articles. The pre-training objective, Gap Sentence Generation (GSG), is designed to have a well performance on text summarization and text generation tasks and is not used in other domains. It is adaptable and can be fine-tuned on limited data, achieving human performance.

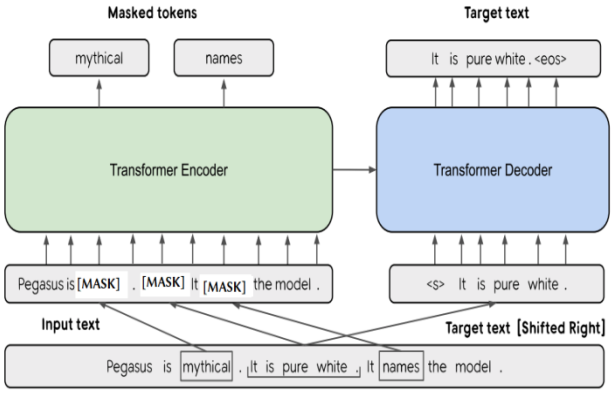


Figure 7: PEGASUS model (source original paper)

5 Results

Measuring performance in a text generation task is not as easy as in other domains such as sentiment analysis or named entity recognition. It is very important to use the appropriate evaluation metrics to illustrate the best performance of the models. To measure the performance of our own work we used one of the most well-known NLP job evaluation mechanisms, ROUGE. ROUGE stands for “Recall-Oriented Understudy for Gisting Evaluation” and evaluates the similarity between a candidate document and a collection of reference documents. The use of the

ROUGE score to evaluate the quality of document translation and summarization models [33]. There are variations of ROUGE metrics such as ROUGE-N, ROUGE-L and ROUGE-Lsum where the N in ROUGE-N stands for the numbers 1 and 2. More specifically, ROUGE-1 compares the monograms between the machine-generated summaries and the human reference digest, ROUGE-2 uses digrams instead of monograms, while ROUGE-L does not compare n-grams, instead treating each digest as a sequence of words and then searching for the greatest common subsequence (LCS). These measures, by comparing the ground truth and the generated text, automatically determine the quality of a generated summary. The number of overlapping n-grams, word sequences, and word pairs between the generated text and the ground truth are used for the metrics. In this work we use ROUGE-1, ROUGE-2, ROUGE-L and ROUGE-Lsum. In Table 2 we show the results of the comparison between 3 pre-trained model pipelines, T5, BART and PEGASUS, on asample text of Facebook news posts and comments. As we have already mentioned above, the use of these particular three models was made because all of them meet the following constraints: they based on Transformers with the Encoder-Decoder architecture and they produce abstractive summarization. T5 pipelines give better results than BART and PEGASUS. We are led to the conclusion that by applying this model to our own data we will produce more accurate summaries. Moreover, T5 is a modern model that has already been successfully used in summarization studies, but in the field of social media it is less explored. This is an initial estimate for our own model design, but by enriching our study with additional assumptions to optimize the results the pre-trained models will be re-evaluated for their performance.

Table 2: Evaluation and Comparison of ROUGE scores

Models	Evaluation Metrics			
	ROUGE-1	*ROUGE-2*	*ROUGE-L*	*ROUGE-Lsum*
T5	0.585034	0.524138	0.544218	0.585034
BART	0.529032	0.483660	0.516129	0.529032
PEGASUS	0.365079	0.306452	0.349206	0.349206

6 Conclusions and future work

The constant increase in the volume of information on social media makes the summarization of texts decisive and necessary to save users' time and resources. Social media is event based. When an event occurs, social networks are flooded with posts related to the event as well as user comments under each post. Since the volume of published data is growing dynamically, reading all these posts in a reasonable time is almost impossible for anyone. Summarizing your social media feed posts is important to keep your readers informed correctly, accurately and in a timely manner. On the other hand, another important issue is the summary of user comments and sub-comments under each post. As their number increases, reading them is a time-consuming and difficult process. User comments are important to read because they express the public's opinion on an important topic. For this reason creating a summary of the comments is of utmost importance. Social media is a challenging area because it offers different types of information, but also because online discussions and chat threads are not formal. They usually contain slang expressions, special symbols such as hashtags or emoticons. Generating summaries from user comments is a difficult task, since they confront language informality and lack lexical richness. Additionally, there is one main question: “How many and which comments will the summary model consider”. Comments are updated dynamically. Working with the most liked, is not an option, as it has been observed that comments readers pay attention to, are the initial comments on the list, finding it tedious to go even deeper. In this paper the overall literature of social media text summarization was reviewed. The result of this literature review shows that various technologies can be used to develop social media summarization models. However, each technology is implemented under different assumptions and constraints. Research around social media text summarization and, in particular, those involving abstractive summarization, are based mainly on Neural Network models. In light of these models, our approach focuses on developing an abstractive, Transformer-based summarization system. The field of text summarization in social media, especially with Transformer models, is still underexplored for both posts and user comments. We focused on creating summaries of user comments since they present greater specificity both in terms of the nature of the language and in terms of their conceptual grouping. In particular, we took user comments on facebook posts and after pre-processing we used them as input to pre-trained language models, based on transformer encoder-decoder architecture. We used ROUGE metrics to compare 3 different pre-trained models and we concluded that T5 gives the highest performance, leading us to the conclusion that T5 will be applied better to our dataset. We proceeded to cluster the data so that T5 can be fed and fine-tuned with the desired text to be summarized. Certainly this is an initial consideration in the design of our model. Through our research we learned about several state-of-the-art models that use a transformer architecture for text summarization. Despite adopting the transformer architecture, these models differ in pre-training strategies and results. Choosing the most appropriate model is necessary, considering the size and training data set of the model. For future work we consider optimizing the results and generating summaries by fine-tuning or prompt-tuning the models to achieve even better outputs. Collecting and classifying datasets from various social media platforms is a difficult task thus our intention is to check the performance of multiple datasets. Since this is an on-going research process we will be able to present more detailed results shortly.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The publication fees were totally covered by ELKE at the University of West Attica.

References

[1] A. Nenkova, K. McKeown, “A survey of text summarization techniques”, Springer US: 43–76, 2012, doi:10.1007/978-1-4614-3223-4_3.

[2] V. Varma, L.J. Kurisinkel, P. Radhakrishnan, “Social Media Summarization”, In A practical Guide to Sentiment Analysis, Chapter 7, 135-153. 2017.

[3] D. Suleiman, A. Awajan, "Deep Learning Based Abstractive Text Summarization: Approaches, Datasets, Evaluation Measures, and Challenges," Mathematical Problems in Engineering, **2020**, 2020, doi:10.1155/2020/9365340.

[4] V. Gupta, G.S. Lehal, "A Survey of Text Summarization Extractive techniques," in Journal of Emerging Technologies in Web Intelligence, 258–268, 2010, doi:10.4304/jetwi.2.3.258-268.

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, "Attention Is All You Need,". In 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA., June 2017.

[6] S. Haque, Z. Eberhart, A. Bansal, C. McMillan, "Semantic Similarity Metrics for Evaluating Source Code Summarization," in IEEE International Conference on Program Comprehension, IEEE Computer Society: 36–47, 2022.

[7] I.S. Blekanov, N. Tarasov, S.S. Bodrunova, "Transformer-Based Abstractive Summarization for Reddit and Twitter: Single Posts vs. Comment Pools in Three Languages," Future Internet, **14**(3), 2022, doi:10.3390/fi14030069.

[8] Q. Li, Q. Zhang, "Abstractive Event Summarization on Twitter," in The Web Conference 2020 - Companion of the World Wide Web Conference, WWW 2020, Association for Computing Machinery: 22–23, 2020, doi:10.1145/3366424.3382678.

[9] I.T. Palma, M. Mendoza, E. Milios, "Neural Abstractive Unsupervised Summarization of Online News Discussions". In: Arai, K. (eds) Intelligent Systems and Applications. IntelliSys 2021. Lecture Notes in Networks and Systems, vol 295. Springer, Cham 2021.

[10] F. Amato, F. Moscato, V. Moscato, A. Picariello, G. Sperli', "Summarizing social media content for multimedia stories creation". The 27th Italian Symposium on Advanced Database Systems (SEB 2019).

[11] F. Amato, A. Castiglione, F. Mercorio, M. Mezzanzanica, V. Moscato, A. Picariello, G. Sperlì, "Multimedia story creation on social networks," Future Generation Computer Systems, **86**, 412–420, 2018, doi:10.1016/j.future.2018.04.006.

[12] J. Bian, Y. Yang, H. Zhang, T.S. Chua, "Multimedia summarization for social events in microblog stream," IEEE Transactions on Multimedia, **17**(2), 216–228, 2015, doi:10.1109/TMM.2014.2384912.

[13] J. Kalita, B. Sharifi, M.-A. Hutton, "Summarizing Microblogs Automatically". Association for Computational Linguistics, Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL, pages 685-688, 2010.

[14] B. Sharifi, D. Inouye, J.K. Kalita, "Summarization of Twitter Microblogs". *The Computer Journal*, Volume 57, Issue 3, March 2014, Pages 378–402, https://doi.org/10.1093/comjnl/bxt109

[15] D. Chakrabarti, K. Punera, "Event Summarization Using Tweets". Proceedings of the International AAAI Conference on Web and Social Media, 5(1), 66-73. https://doi.org/10.1609/icwsm.v5i1.14138. 2011

[16] F. Chong, T. Chua, S. Asur, "Automatic Summarization of Events From Social Media". Proceedings of the International AAAI Conference on Web and Social Media, **7**(1), 81-90. https://doi.org/10.1609/icwsm.v7i1.14394, 2021.

[17] S. Gao, X. Chen, P. Li, Z. Ren, L. Bing, D. Zhao, R. Yan, "Abstractive Text Summarization by Incorporating Reader Comments". In The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19), 2019.

[18] P. Bhandarkar, K.T. Thomas, "Text Summarization Using Combination of Sequence-To-Sequence Model with Attention Approach", Springer Science and Business Media Deutschland GmbH: 283–293, 2023, doi:10.1007/978-981-19-3035-5_22.

[19] Q. Wang, J. Ren, "Summary-aware attention for social media short text abstractive summarization," Neurocomputing, **425**, 290–299, 2021, doi:10.1016/j.neucom.2020.04.136.

[20] A. Varol, "Innovative Technologies for Digital Transformation". In the 1st International Informatics and Software Engineering Conference (IISEC-2019) proceedings book : 6-7 November 2019, Ankara/Turkey.

[21] Z. Liang, J. Du, C. Li, "Abstractive social media text summarization using selective reinforced Seq2Seq attention model," Neurocomputing, **410**, 432–440, 2020, doi:10.1016/j.neucom.2020.04.137.

[22] Z. Kerui, H. Haichao, L. Yuxia, "Automatic text summarization on social media," in ACM International Conference Proceeding Series, Association for Computing Machinery, 2020, doi:10.1145/3440084.3441182.

[23] M.T. Nguyen, V.C. Nguyen, H.T. Vu, V.H. Nguyen, "Transformer-based Summarization by Exploiting Social Information," in Proceedings - 2020 12th International Conference on Knowledge and Systems Engineering, KSE 2020, Institute of Electrical and Electronics Engineers Inc.: 25–30, 2020, doi:10.1109/KSE50997.2020.9287388.

[24] M.H. Su, C.H. Wu, H.T. Cheng, "A Two-Stage Transformer-Based Approach for Variable-Length Abstractive Summarization," IEEE/ACM Transactions on Audio Speech and Language Processing, **28**, 2061–2072, 2020, doi:10.1109/TASLP.2020.3006731.

[25] D. Singhal, K. Khatter, A. Tejaswini, R. Jayashree, "Abstractive Summarization of Meeting Conversations," in 2020 IEEE International Conference for Innovation in Technology, INOCON 2020, Institute of Electrical and Electronics Engineers Inc., 2020, doi:10.1109/INOCON50539.2020.9298305.

[26] A. Gupta, D. Chugh, R. Katarya, "Automated News Summarization Using Transformers.", In Sustainable Advanced Computing, 2022, Volume 840. ISBN : 978-981-16-9011-2, 2022

[27] A. Pal, L. Fan, V. Igodifo, Text Summarization using BERT and T5. https://anjali001.github.io/Project_Report.pdf

[28] K. Pipalia, R. Bhadja, M. Shukla, "Comparative analysis of different transformer based architectures used in sentiment analysis," in Proceedings of the 2020 9th International Conference on System Modeling and Advancement in Research Trends, SMART 2020, Institute of Electrical and Electronics Engineers Inc.: 411–415, 2020, doi:10.1109/SMART50582.2020.9337081.

[29] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, A.M. Rush, "HuggingFace's Transformers: State-of-the-art Natural Language Processing," 2019.

[30] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," 2019.

[31] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P.J. Liu, "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," 2019.

[32] J. Zhang, Y. Zhao, M. Saleh, P.J. Liu, "PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization," 2019.

[33] C.-Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries". In Text Summarization Branches Out, pages 74–81, Barcelona, Spain. Association for Computational Linguistics, 2004

APPENDIX I

Table 1: Methods and Summarization Techniques on Social Media

Author	Summarization Technique	Implementation Method	Pre-neural Approaches	Deep-Learning Approaches		
				RNN/CNN	Transformer-based	Social Media
[10]	Summarization Graph with multimedia output	Influence Maximization and ABC algorithms	V	X	X	V
[11]	Summarization Graph with multimedia output	Influence Maximization and ABC algorithms	V	X	X	V
[12]	Extractive Summarization	A cross-media probabilistic model, termed Cross-Media-LDA (CMLDA),	V	X	X	V
[13] & [14]	Extractive Summarization	Phrase Reinforcement Algorithm & TF-IDF algorithm	V	X	X	V
[15]	Extractive Summarization	A modified Hidden Markov Model	V	X	X	V
[16]	Extractive Summarization	Gaussian Decay Topic Model (GDTM)	V	X	X	V
[17]	Abstractive Summarization	Reader-aware model based on CNN and Bi-RNN with Seq2Seq **(No Transformer)**	X	V	X	V
[18]	Abstractive Summarization	Sequence-to-Sequence with Attention Approach **(No Transformer)**	X	V	X	V
[21]	Abstractive Summarization	Selective Reinforced Sequence-To-Sequence Attention Model **(No Transformer)**	X	V	X	V
[19]	Abstractive Summarization	Summary-aware attention neural model **(No Transformer)**	X	V	X	V
[20]	Abstractive Summarization	Word2Vec with Recurrent Neural Network model GRU **(No Transformer)**	X	V	X	V
[8]	Abstractive Summarization	NNs with BERT model as the encoder and Transformer architecture as the decoder	X	X	V	V
[22]	Extractive Summarization	BERT, reinforcement learning and other technologies	X	X	V	V
[9]	Abstractive Summarization	NNs with BERT model as the encoder and Transformer architecture as the decoder	X	X	V	V
[7]	Abstractive Summarization	Transformer-based NNS LongFormer Large and T5 Large	X	X	V	V
[23]	Extractive Summarization	BERT with use of CNN	X	X	V	V
[24]	Extractive Summarization	Pre-trained BERT, RNN/LSTM, BERTSUM	X	X	V	X
[25]	Abstractive Summarization	Classic Transformer Model	X	X	V	X
[26]	Abstractive Summarization	Pre-trained Transformer Models BART, PEGASUS, T5	X	X	V	X
[27]	Abstractive Summarization	BERT, T5	X	X	V	X

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 8, No. 4, 28-35 (2023)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Investigating the Impression Effects of a Teacher-Type Robot Equipped a Perplexion Estimation Method on College Students

Kohei Okawa[*,1], Felix Jimenez[2], Shuichi Akizuki[3], Tomohiro Yoshikawa[4]

[1]*Graduate School of Information Science and Technology,Aichi Prefectural University, 1522-3 Ibaragabasama, Nagakute-shi, Aichi, 480-1198, Japan*

[2]*School of Information Science and Technology,Aichi Prefectural University, 1522-3 Ibaragabasama, Nagakute-shi, Aichi, 480-1198, Japan*

[3]*School of Engineering, Chukyo University, 101-2 Yagoto Honmachi, Showa-ku, Nagoya, Aichi, 466-8666, Japan*

[4]*Faculty of Medical Engineering, Suzuka University of Medical Science, 1001-1, Kishioka, Suzuka-city, Mie, 510-0293, Japan*

ARTICLE INFO

Article history:
Received: 30 March, 2023
Accepted: 08 July, 2023
Online: 25 July, 2023

Keywords:
Deep Learning
The Educational support robot
The Impression
Collaborative learning

ABSTRACT

In recent years, the adoption of ICT education has increased in educational settings. Research and development of educational support robots have garnered considerable interest as a promising approach to inspire and engage students. Conventional robots provide learning support through button operations by the learners. However, the frequent need for button operation to request support may lead to a tedious impression on the learner and lower the efficiency of the learning process. Therefore, in this study, we developed a Perplexion Estimation Method that estimates the learner's state of perplexity by analyzing their facial expressions and provides autonomous learning support. We verified the impact of a teacher-type robot (referred to as the proposed robot) that autonomously provides learning support by estimating the learners' perplexity states in joint learning with university students. The results of a subject experiment showed that the impression of the proposed robot was not different from that of the conventional robot. However, the proposed robot demonstrated the ability to provide optimal support timing compared to the conventional robot. Based on these results, it is expected that the utilization of the perplexion estimation method with teacher-type robots can create a learning environment similar to human-to-human interaction.

1 Introduction

This paper is an extension of the one presented at SCIS [1]. In this conference, we presented the results of the life quality and the timing of support given to the learner by a supervised robot equipped with the proposed puzzling estimation method. In addition to these results, this paper provides additional analytical information on the results of subject experiments.

In recent years, the introduction of ICT education has become increasingly active in the field of education. The introduction of ICT education using educational big data that collects the learning status of individual students is being promoted in order to realize "fair, individualized and optimized learning that leaves no one behind" for children who have difficulty learning with other children due to reasons such as not attending school, children with developmental disabilities, and other children who are becoming increasingly diverse [2]. By utilizing information technology, it is possible to provide individualized educational support. In this ICT education, collaborative learning with robots that have a "presence" in the real world is shown to be effective in creating a learning environment where people can learn from each other, as well as in stimulating interest in learning [3]. It has also been reported that the robot's advise is superior to that of an on-screen agent [4]. Therefore, we believe that educational support by robots is more effective than on-screen agents in educational settings. For these reasons, research and development of robots that can play an active role in educational settings (hereinafter referred to as "educational support robots") has been attracting attention in Japan and abroad [5].

Educational robots include "teacher-type robots" that instruct learners like a teacher. The role of a supervised robot in conventional research is to teach the learner how to solve a problem[6]. For example, Yoshizawa et al, proposed a supervised robot that

[*]Corresponding Author: Kohei Okawa, Aichi Prefectural University, 1522-3 Ibaragabasama, Nagakute-shi, Aichi, 480-1198, JAPAN, 0561-76-8826, im222002@cis.aichi-pu.ac.jp

switches learning support according to the number of correct answers given by the learner. Experimental results showed that the robot that switches the learning support has the potential to provide high learning effectiveness for university students.

However, conventional supervised robots have a problem in that they cannot autonomously provide learning support in response to the learner's state of being unable to solve a problem (hereafter referred to as "perplexion"). For example, Yoshizawa et al.'s teacher-type robot provides learning support by pressing a button of the learner (hereinafter referred to as a "conventional robot"). The learner presses a button every time he/she needs learning support.

Conversely, enhancing the provision of support in learning systems is a critical focus in educational psychology research. In an environment where learners can request unlimited support, excessive use of hints [7] and search for hint patterns [8] occur regardless of the need for support. The use of artificial intelligence techniques has been reported to be effective in preventing these problems [9].

Therefore, the goal of this study creates a learning environment where a teacher-type robot provides autonomous learning support without the need for learners to press buttons, using deep learning techniques. We believe that it is effective for a teacher-type robot to estimate the learner's state of perplexion and autonomously provide learning support at the most appropriate timing, instead of providing learning support at the push of a button by the learner. Furthermore, when estimating the learner's state of perplexion, it is important to ensure that no burden is placed on the learner. We believe that this will prevent the learner from becoming dependent on the assistance and will enable smooth interaction with the robot, thereby improving the effectiveness of the teacher-type robot on the learner.

Regarding the learner's state estimation, Matsui has conducted previous research on autonomous learning support through perplexity state estimation [10]. Matsui's research attempted to estimate the learner's mental state by combining biometric devices, back-and-forth facial movements, and mouse movements. By using biological signals, we can obtain a state that is closer to the learner's raw data, which allows us to accurately estimate the perplexion. On the other hand, when using measurement equipment, the burden on the learner is large, and the data obtained is likely to contain noise in the real perplexity data. In the case of back-and-forth facial movements and mouse movements, the influence of the learner's posture and thinking habits is considered to be significant. Few studies have focused on estimating perplexion solely based on facial expression, despite the existing research on combining it with biometric signals.

In this study, we focus on research on human facial expression recognition [11] to provide autonomous learning support through teacher-type robots. In particular, methods based on deep learning have been widely used in research on facial expressions, and have shown high performance in image recognition and image classification Convolutional neural networks (hereafter, this is called CNN) have been proposed [12]. However, conventional research on facial expression recognition has focused only on the seven basic emotions of anger, disgust, fear, happy, sad, surprise, and neutral (hereafter referred to as the seven basic emotions), and has not focused on perplexion.

Therefore, in this study, we constructed the proposed method by transfer learning, using the seven basic emotion estimation methods of [13] as a base model. The proposed method is a perplexion estimation method that classifies two classes of perplexion state and non-perplexion state. In this study, we first extended the seven emotion estimation method to the eight emotion estimation method, including the perplexion state (67% recall).

Then, we conducted a subjective experiment on the impression effect of a teacher-type robot equipped with the eight emotion estimation method (hereinafter referred to as "the eight emotion robot") and discovered that it elicited a similar level of the impression as a conventional robot [14]. On the other hand, the accuracy of the robot was not sufficient, as some subjects commented in a questionnaire that "the timing of support is too fast" and "the robot is noisy when it repeatedly speaks. To solve this problem, we constructed a proposed method to improve the estimation accuracy (88% recall). Nevertheless, we have been unable to make a direct comparison between the impressions and support timing of the robot equipped with the proposed method and the eight emotion robot.

In this paper, we verify the impression effect on university students and the support timing of joint learning with a teacher-type robot that provides learning support autonomously (hereinafter referred to as "the proposed robot") equipped with the proposed method specialized for estimating the perplexion on a teacher-type robot. In the experiment, we will conduct a comparison between the proposed robot and the eight emotion robot.

TThis paper begins with a description of the proposed method in Chapter 2. Then, in Chapter 3, we verify the impression that the proposed robot gives to the learner and the timing of support through experiments with participants. Chapter 4 discusses the results and Chapter 5 summarizes them.

2 Perplexioin Estimation Method

Eight emotion estimation methods are designed to estimate the seven basic emotions together with the state of perplexion. However, including the estimation of the eight emotions reduces the accuracy of the estimation. Since the primary focus during learning is to determine whether the state is perplexed or not, estimating the eight emotions becomes unnecessary. In this chapter, we have developed a method for estimating perplexity using transfer learning, specifically focusing on estimating only the perplexed and non-perplexion states.

2.1 How to collect perplexed facial expression data

Perplexed state data were collected by capturing learners' facial expressions (at a resolution of 1920 and 1080 pixels, 30 frames per second) while they interacted with software related to the Technology Passport Examination (IT Passport) or software dealing with mathematical graphics difficulties and the Computer Aptitude Battery (CAB) [15] provided by SHL Japan. The participants in this study were university students, and the learning software was specifically designed to present them with moderately difficult problems that required careful thinking, based on their university lectures and high school mathematics knowledge.

The software presented a question with a hint button underneath. Participants were informed that they could press the button as many

times as they wanted to get pedagogical help. This facilitated the annotation of facial expression data during puzzles.

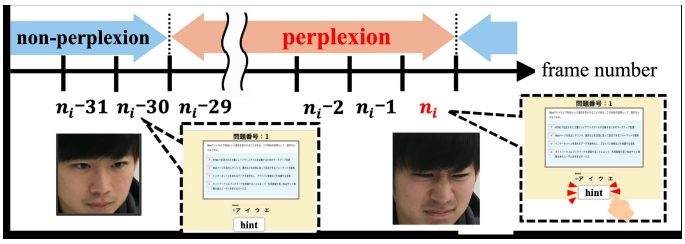


Figure 1: Definition of perplexion

In this study, facial expression data captured within 1 second (frames $n_i - 29$ at 30 frames per second) before the i-th button press (frame n_i) during learning with the learning software were categorized as perplexed state data, while other facial expression data were classified as non-perplexed state data (Fig. 1). The decision to define facial expression data up to 1 second ago as perplexed state data was based on the assumption that the expression of perplexity would be prominently evident during that period. Including facial expressions starting from 2 seconds ago may introduce variability in the strength of the puzzled state, leading to potential errors. Therefore, in this study, perplexed facial expressions are defined as the facial expressions that occurred 1 second before. The non-perplexed state is defined as the data collected from the frame immediately following the i-th button press (frame n_{i+1}) to the frame just before the onset of the perplexed state at the time of the i-th button press (frames n_{i+i} to $n_{i+i} + 30$).

2.2 Method Overview

In the field of deep learning, when the dataset available for training is small, transfer learning [16] is often utilized to leverage features learned by pre-existing models. In our study, in order to effectively capture subtle changes in facial expressions such as perplexion and non-perplexion, we employed seven basic emotion recognition models based on the FER2013 dataset as the base models (referred to as the "base model" in Fig. 2).

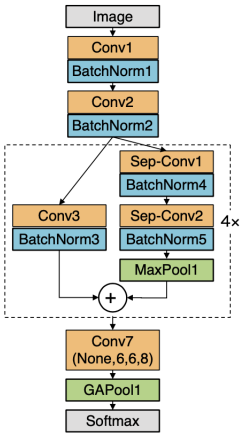


Figure 2: CNN model configuration

Specifically, we utilized the emotion recognition model using CNN developed by O. Arriaga et al [13]. By adapting the base model's features to focus on facial expressions, we anticipated an enhancement in the accuracy of perplexion estimation. Consequently, we constructed a dedicated model for perplexion estimation. In this model, the final layer generates the likelihood of two classes based on the perplexion state and the non-perplexion states. he perplexion estimation method of this paper was constructed using only facial expression data of college students. Therefore, the present method is specific to college students. To construct the perplexion estimation method, we utilized 36 face images out of the total of 52 collected in Section 2.1 as training data for perplexion/non-perplexion states. Of these, 26 images were used as training data (teacher data), while 10 images were kept aside for testing. Furthermore, since the method involves the binary classification of perplexion and non-perplexion states, we also included data from the FER2013 dataset representing other emotions into the non-perplexion data category (Table 1). Table 2 presents the number of data samples obtained from the aforementioned datasets.

The estimation of perplexion state is performed by inputting the face image to be estimated into a pre-trained model. The output of the model, denoted as y_c, represents the probability distribution over each class, with c indicating corresponding perplexion or non-perplexion state.

$$\arg\max_{1 \le c \le 2} y_c = 2 \tag{1}$$

Therefore, the state of perplexion is determined when the value of the formula (1) is "2", while the state of non-perplexion corresponds to all other cases. The process of estimating the perplexion state, referred to as "perplexion state estimation," commences after an average duration of time, denoted as x, has elapsed since the learner requests a hint. This approach prevents erroneous recognition that may occur when detecting, for example, a furrowed brow immediately upon the learner encountering a problem.

Table 1: Breakdown of the learning data.

	breakdown
Perplexion	Perplexion data collected
Non-Perplexion	Basic seven emotions from FER2013 Non-Perplexion data collected

Table 2: Configuration and number of learning data

Data Name	Data source	Number of data
anger	FER-2013	1997
disgust	FER-2013	218
fear	FER-2013	2048
happy	FER-2013	3607
sad	FER-2013	2415
surprise	FER-2013	1585
neutral	FER-2013	2482
non-perplexion	By Section 2.1	3504
perplexion	By Section 2.1	3475

To illustrate, the perplexion state estimation initiates after 30 seconds if the average time is set to be 30s.

3 participant experiment

3.1 Robot overview

For the experiment, we utilized Tabot (Figure 3), a tablet-type robot with a tablet serving as its head, capable of displaying agents and expressing different facial expressions. Tabot's body consists of 3 degrees of freedom (DOFs) in the neck, 5 DOFs in an arm, and 1 DOF in the legs, for a total of 14 DOFs that allow different body movements.

In this experiment, as shown in Figure 3, a camera was installed on the robot's head to capture facial expressions during the learning process. The camera records the student's facial expressions as they look at the tablet screen during the learning process. The camera footage is regulated and processed by a processing PC. If a state of perplexion is detected from the captured facial expression data, an instruction is sent to Tabot via Local Area Network communication. The camera has a resolution of 640 and 480 pixels and a frame rate of 30 frames per second.

3.2 Learning system

Participants study using the learning system shown in Figure 4. In the pre-training phase, subjects study, only with the learning system installed on their PCs, while in the collaborative phase, they study with the learning system displayed at the bottom of the tabot. The learning system displays the screen shown in Figure 4(a) when the subject proceeds to study. In Figure 4(a), there is a question and a button below the question that provides a hint. We informed the participants that they could press the button as many times as they wanted and that they would receive a hint by pressing the button. This makes it possible to label the facial expression data in the perplexoin. However, the hint button is installed only in the pre-training phase, and is removed in the collaborative learning phase. When an answer is given on the screen shown in Figure 4(a), the user is taken to the screen shown in Figure 4(b). The system repeats this process for the number of questions to reach the final screen (Figure 4(c)).

The learning system comprises two types of challenges that are designed to involve a puzzle-like element or spark.

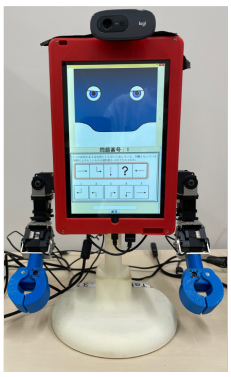


Figure 3: Tabot and camera

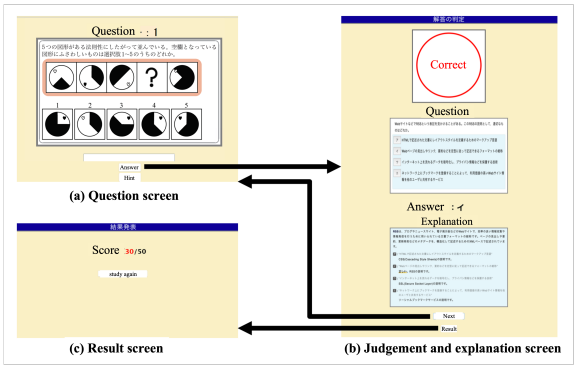


Figure 4: Learning system

These challenges consist of mathematical figure problems and Computer Aptitude Battery (CAB) problems provided by SHL Japan [15]. The mathematical figure problems can be effectively tackled by utilizing auxiliary lines in the drawing process, while the CAB problems involve identifying patterns and regularities. By presenting these challenges that demand both inspiration and puzzle-solving skills, we created an environment in which participants were consistently perplexed.

Table 3: Experimental information

Item	Contents
Property	Undergraduate and graduate students
The number of participants	34
Male:Female	20:14
Period	2020/9/14 ~1/30 2021/5/6 ~6/30 2022/2/17 ~3/17

3.3 Experimental procedure

Experiments were conducted in which undergraduate and graduate students and robots learned together. Experimental information is summarized in the table3 below.

In the experiment, we compare two groups of teacher-type robots equipped with different emotion recognition models. The groups to be compared are the eight emotion group equipped with the eight emotion estimation method and the proposed group equipped with the perplexion estimation method.

The information for the groups is shown in Table 4. For the proposed group, the estimated start time is set to 106 seconds after the first press of the hint button. This is because the average time elapsed until the first press of the hint button is 106 seconds during the experimental period up to 2021. This experiment is divided into three-time periods due to the number of participants. The proposal group was conducted in 2022, and the eight emotion group was conducted in the other dates.

The experimental procedure is shown below.

Step 1 (Pre-learning) Participants learn with a computer-based learning system. The learning time was about 60 minutes.

Table 4: Group information

	Eight emotion group	Proposed group
emotion recognition model	Eight emotion estimation method	Perplexion estimation method
hint provide condition	When estimating the perplexion	
Support frequency limitation	No limit	two times
hint button	nothing	
Estimated start time	Immediately after the problem is presented	After 106s

Step 2 To reduce the learning effect, we allowed a one-week interval. In addition, the groups were assigned so that the number of times, the participants pressed the hint button in the pre-learning period would be as equal as possible.

Step 3 (Collaborative learning) Participants learned with a robot from the group to which they were assigned. The learning time was about 60 minutes.

Step4(Survey) Immediately after the collaborative learning the participants answer a survey.

3.4 Evaluation index

We employed the ”Godspeed Questionnaire [17, 18],” a questionnaire methodology designed for the subjective evaluation of human-robot interaction. One of the questionnaire items utilized in our study was focused on assessing ”Animacy.” Additionally, we used the questionnaire to evaluate the timing of the robot's assistance. We selected these evaluation indices based on their perceived significance in the context of teacher-type robots. Previous research suggests that users tend to be more emotionally engaged and influenced by objects exhibiting animacy [19]. Therefore, a higher level of animacy in the robot may lead to a greater receptiveness to advise provided by the teacher-type robot. Furthermore, the autonomous timing of the robot's assistance is crucial for creating a more human-like learning environment. Each adjective pair in Godspeed Questionnaire was rated on a 5-point scale, and we quantified each pair on a scale of 1 to 5, with the positive adjective side receiving a higher score. We defined the average score of six animacy items as ”animacy” and compared it across different groups. Figure 5 illustrates the questionnaire pertaining to support timing.

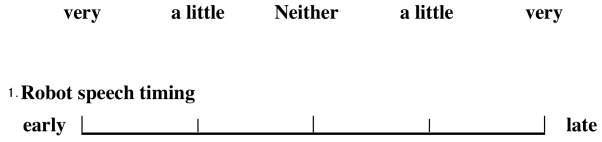


Figure 5: Questionnaire on support timing

Table 5: Scoring Procedure

survey	score
3	3
2 or 4	2
1 or 5	1

To analyze the evaluation results, we employed a Student's t-test for animacy, assuming equal population variance across the groups. Regarding the timing of support, two types of analyses were performed. The first analysis involved testing the population mean when the population variance was unknown, using ”3” for ”neither” as the optimal timing criterion. The second analysis utilized a Student's t-test with the timing questionnaire as the scoring criterion. Details of the scoring procedure can be found in Table 5. The significance level was adjusted to $p < 0.0125$ using the Bonferroni correction to account for multiple comparisons, as four tests were conducted. The tests included comparisons of animacy, the average support timing between the eight emotion group and the proposed group, and comparison of support timing scores.

3.5 Result

Figure 6, 7, and 8 show the mean values of lifeliness and support timing, the mean values when scoring, and the test results. The proposed group had higher animacy than the eight emotion group. The test results showed no significant differences. These results suggest that the animacy that participants felt toward the robot was the same in each group of robots.

The support timing results for the “eight emotion group” and the “proposed group” were both found to be early and close to the criterion value of “3”. The statistical analysis revealed a significant difference between the eight emotion group and an insignificant difference between the proposed group. This indicates that the eight emotion group was perceived as providing support too early, while the proposed group offered support at an appropriate timing. Moreover, employing the perplexion estimation method for support timing could lead to optimal timing of assistance. The results of scoring the timing of support were higher in the suggestion group than in the eight-emotion group.

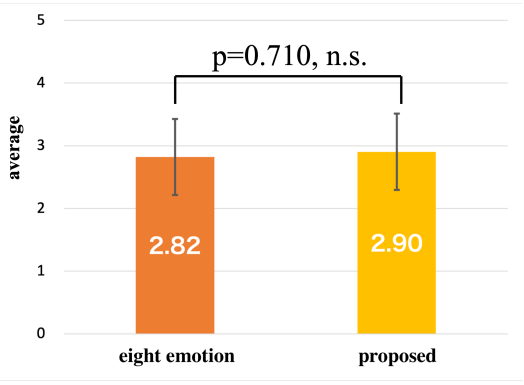


Figure 6: Result of Animacy

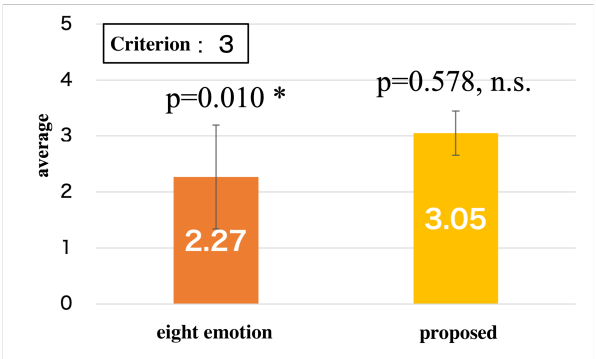


Figure 7: Result of Support timing

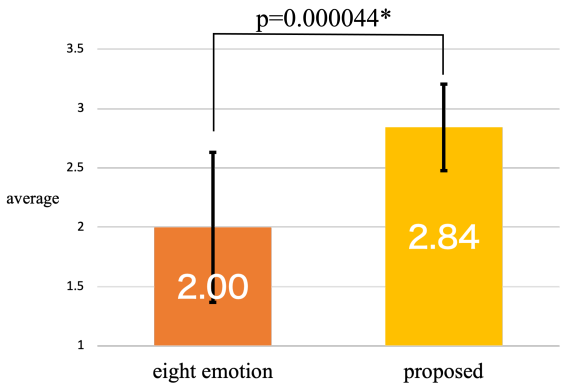


Figure 8: Result of support timing score

The results of the test showed that there was a significant difference. Therefore, the results indicated that the proposed group was better in support timing.

4 discussion

The experimental results indicated that there was no significant difference in animacy, suggesting that the learners had similar impressions regardless of the group they were in. However, a significant difference was found in the support timing between the eight emotion group and the proposed group. In the one-group test based on the criterion value of "3", the eight emotion group showed a timing that was too early for the learners, while the proposed group provided support at the optimal timing. Furthermore, a significant difference was also found in the test when the questionnaires were converted into scores. These differences will be discussed in the following sections, along with the results of each questionnaire.

Animacy was higher in the proposed group than in the eight emotion group. We believe that this may be related to the number of speech utterances during collaborative learning. To investigate this, we examined the number of speech utterances for six participants in the eight emotion group and nineteen participants in the proposed group. The average number of utterances per participant was 29 for the eight emotion group and 0.2 for the proposed group. These results suggest that too many speech utterances may impair Animacy.

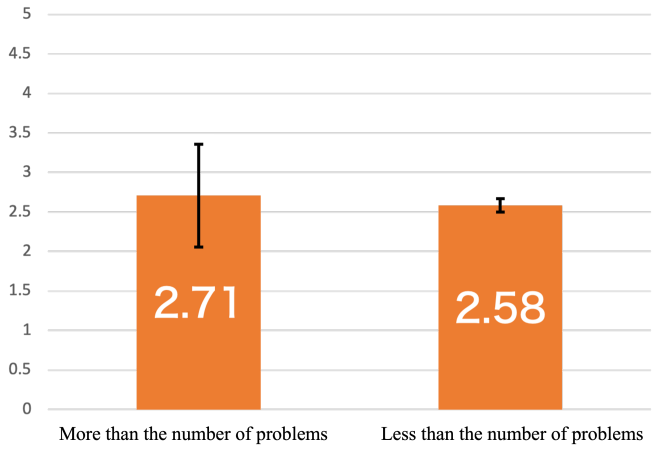


Figure 9: Animacy when divided by the number of support (The eight emotion group)

To further explore this, we investigated the Animacy scores of participants in the eight emotion group whose the number of speech utterances exceeded the number of questions. Four out of the six participants exceeded the number of questions, and we calculated the average Animacy scores of these four participants and the other two participants (Figure 9). Contrary to our expectations, Animacy was higher for those who received more support. However, due to the small number of participants, we cannot conclude with certainty that there is a clear trend. One possible explanation is that repeated utterances made the participants feel that there was a response.

In the Animacy section, there is a "responsiveness" column. The average score of the four respondents who received more support was 3.5, while that of the two respondents who received less support was 2. This result also suggests that a higher number of speech utterances may influence Animacy, as the respondents feel more responsive. However, the average response to the questionnaire regarding the timing of support was 1.75 when the number of times support was high and 2 when the number was low. Based on the open-ended responses from participants who had a high frequency of speech utterances, some of them expressed that there were "excessive number of hints" and they felt "confused by the continuous stream of advice". Therefore, we believe that an excessive number of speech utterances does not have a positive effect.

Similarly, in the proposed group, the mean score of Animacy was calculated for four of the 19 participants for whom learning support was confirmed and for 15 participants for whom it was not confirmed (see Figure 10). Animacy scores were higher for the four participants for whom learning support was confirmed. Additionally, as in the eight emotion group, we calculated the mean score for each "responsive" item. The mean score for the four participants for whom learning support was confirmed was 4, and for the 15 participants for whom learning support was not confirmed, it was 3.3. Based on the analysis of the eight emotion group, we believe that having learning support is better than not having it, but that frequent learning support, such as continuous speech, may not give a good impression.

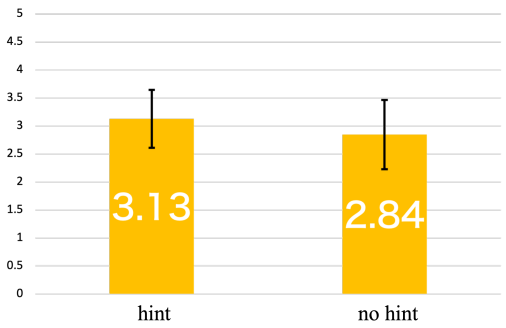


Figure 10: Animacy when divided by the number of support (The proposed group)

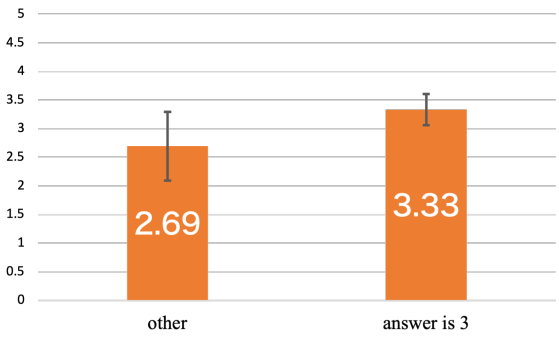


Figure 11: Animacy by support timing answers (eight emotion group)

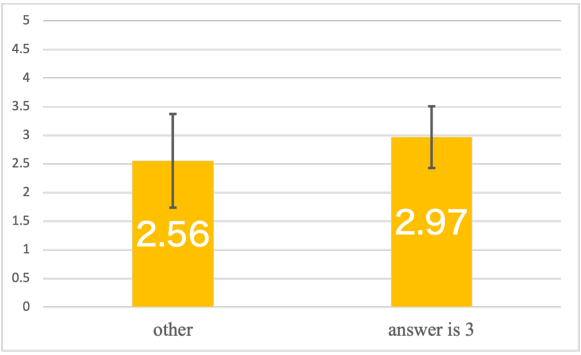


Figure 12: Animacy by support timing answers (proposed group)

We believe that the timing of support is an essential factor in autonomous support provided by a teacher-type robot. If support is given too early, learners may feel frustrated that it is offered at a time when they do not need it, whereas if it is provided too late, it may cause learners to feel frustration at not receiving support and not being able to solve problems. The experimental results showed a significant difference between the eight emotion group and the ideal timing of "3". Therefore, it is thought that the participants felt that the support was premature. However, no significant difference was observed in the proposed group. This suggests that the support timing used in the proposed group may be the optimal timing. The test of scoring the timing of support showed a significant difference, with the proposed group having a higher score than the eight emotion group, indicating that the proposed group was superior. However, as seen in the previous data, the eight emotion group tends to talk too much, resulting in a lower score. Similarly, the proposed group may not have been judged as early or late because they provided support significantly less frequently. Therefore, the results suggest that the perplexion estimation method used to solve this problem led to good results, as the evaluation of too many utterances was low.

Animacy and timing results suggest that teacher-type robots that provide appropriate timing and frequency of support can improve Animacy. Figures 11 and 12 show the average Animacy scores of participants who answered "3" in the questionnaire about support timing, and those who answered "other." Three participants in the eight emotion group and 16 participants in the proposed group answered "3" for support timing, and in both cases, those who answered "3" had higher scores. These results indicate that refining support timing is effective in improving Animacy. To achieve this, we believe that the accuracy of the perplexity estimation method needs improvement. The current method has a reproducibility of 88%, but its accuracy is 67%, which is inferior to the eight emotion estimation methods. We believe that improving accuracy is necessary to provide an environment that can estimate the learner's perplexity more accurately without missing it.

5 Conclusion

Our study developed a deep learning-based method to estimate the learner's state of perplexion. We conducted experiments to compare the performance of a robot equipped with our proposed method to a robot equipped with the conventional eight emotion estimation methods. The results of the participant experiment revealed that there was no significant difference in terms of animacy between the two groups. This indicates that our proposed method provided a similar level of animacy to the learners compared to the conventional robot. However, when it comes to support timing, the robot equipped with our proposed method demonstrated the ability to provide support at the optimal timing. Furthermore, when the support timing was scored, a significant difference was observed, indicating that the support timing of the robot equipped with the proposed method was better. In other words, the proposed method can potentially enable the realization of a collaborative learning environment between a teacher-type robot and a learner without the need for buttons. Overall, we believe that this study has demonstrated the potential for realizing a collaborative learning environment between a robot and a learner using the proposed perplexion estimation method.

In the future, our plan is to develop a robot suitable for actual learning environments for junior high school and high school students, and to evaluate the impression the proposed robot leaves on learners as well as its effectiveness in supporting learning. To achieve this, we intend to conduct participant experiments that evaluate both the impression of the robot and its impact on learning. Moreover, to create a more effective learning environment for the learners, it is essential to enhance the accuracy of the perplexion estimation method. To achieve this, we are considering two approaches. The first approach is to fine-tune the method by training it on the data collected from the planned experiments, so that it can accommodate a wide range of age groups. This will enable us to effectively support learners with youthful facial features, such

as middle and high school students, who are expected to be actual users of the system. The second approach is to treat the learning data as time series data, including changes in facial expressions. Currently, we only focus on the moments of perplexion. However, by capturing the trends of perplexion onset, we anticipate that we can improve the estimation accuracy to a greater extent.

Conflict of Interest The authors declare no conflict of interest.

References

[1] K. Okawa, F. Jimenez, S. Akizuki, T. Yoshikawa, "Effects of a Teacher-type Robot Using a Perplexion Estimation Method on College Students," in 2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS), 2022, doi:10.1109/SCISISIS55246.2022.10001988.

[2] Ministry of Education, Culture, Sports, Science and Technology, "Measure to Promote the Use of Adbanced Technology to Support Learning in the New Era (Final Summary)," 2019, doi:https://www.mext.go.jp/a_menu/other/1411332.htm.

[3] T. Kanda, "How a communication robot can contribute to education?" Journal of Japanese Society for Artificial Intelligence : AI, **23**(2), 229–236, 2008, doi:10.11517/jjsai.23.2_229.

[4] K. Shinozawa, F. Naya, J. Yamato, K. Kogure, "Differences in Effect of Robot and Screen Agent Recommendations on Human Decision-Making," International Journal of Human-Computer Studies, **62**(2), 267–279, 2005, doi:10.1016/j.ijhcs.2004.11.003.

[5] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, F. Tanaka, "Social robots for education: A review," Science Robotics, **3**, 2018, doi:https://doi.org/10.1126/scirobotics.aat5954.

[6] R. Yoshizawa, F. Jimenez, K. Murakami, "Proposal of a Behavioral Model for Robots Supporting Learning According to Learners'Learning Performance," Journal of Robotics and Mechatronics, **32**(4), 769–779, 2020, doi:https://doi.org/10.20965/jrm.2020.p0769.

[7] V. Aleven, K. R. Koedinger, "Limitations of student control:Do students know when they need help?" in Proceedings of the 5th International Conference on Intelligent Tutoring Systems, 2000, doi:10.1007/3-540-45108-0_33.

[8] J. A. Walonoski, N. T. Heffernan, "Detection and Analysis of Off-Task Gaming Behavior in Intelligent Tutoring Systems," in Proceedings of the 8th International Conference on Intelligent Tutoring Systems, 2006, doi:10.1007/11774303_38.

[9] I. Roll, V. Aleven, B. M. McLearen, K. R. Koedinger, "Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system," Learning and Instruction, **21**(2), 267–280, 2011, doi:10.1016/j.learninstruc.2010.07.004.

[10] T. Matsui, "Estimation of Learners' Physiological Information and Learners' Mental States by Machine Learning and Its Application for Learning Support," Transactions of Japanese Society for Information and Systems in Education, **36**(2), 76–83, 2019, doi:https://doi.org/10.14926/jsise.36.76.

[11] M. Wimmer, B. A. MacDonald, D. Jayamuni, A. Yadav, "Facial Expression Recognition for Human- Robot Interaction - A Prototype," in In Proceedings of the 2nd International Workshop on Robot Vision, 2008, doi:10.1007/978-3-540-78157-8_11.

[12] T. Nishime, S. Endo, K. Yamada, N. Toma, Y. Akamine, "Feature Acquisition From Facial Expression Image Using Convolutional Neural Networks," Journal of Robotics, Networking and Artificial Life, **32**(5), 1–8, 2017, doi:http://dx.doi.org/10.2991/jrnal.2016.3.1.3.

[13] O. Arriaga, P. G. Ploger, M. Valdenegro, "Real-time Convolutional Neural Networks for Emotion and Gender Classification," 2017, doi:10.48550/arXiv.1710.07557.

[14] K. Okawa, F. Jimenez, K. Murakami, S. Akizuki, T. Yoshikawa, "A Survey on Impressions of Collaborative Learning with an Educational Support Robot Estimating Learners' Perplexion," in Proceedings of the 37th Fuzzy Systems Symposium, 171–176, 2021, doi:10.14864/fss.37.0_171.

[15] SHL Japan products, 2020/12/06 accessed, doi:http://www2.shl.ne.jp/product/index.asp?view=recruit.

[16] A. Gulli, S. Pal, Chokkan Deep learning: Python×Keras de aidea o katachi ni suru reshipi, O'Reilly Japan, 2018, doi:https://www.ohmsha.co.jp/book/9784873118260/.

[17] C. Bartneck, D. Kulic, E. Croft, S. Zoghbi, "Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived, Intelligence, and Perceived Safety of Robots," International Journal of Social Robotics, **1**(1), 71–81, 2009, doi:https://psycnet.apa.org/doi/10.1007/s12369-008-0001-3.

[18] T. Nomura, "Humans' Subjective Evaluation in Human-Agent Interaction(HAI)," Journal of the Japanese Society for Artificial Intelligence, **31**(2), 224–229, 2016, doi:https://doi.org/10.11517/jjsai.31.2_224.

[19] B. J. Fogg, Persuasive Technology: Using Computers to Change what We Think and Do, Morgan Kaufmann, 2002, doi:10.1016/B978-1-55860-643-2.X5000-8.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 9, No. 2, 33-44 (2024)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Revolutionizing Robo-Advisors: Unveiling Global Financial Markets, AI-Driven Innovations, and Technological Landscapes for Enhanced Investment Decisions

Marco I. Bonelli*, [1], Jiahao Liu[2]

[1]*Dr. Vishwanath Karad MIT World Peace University, Pune, 411038, India*

[2] *Judge Business School, University of Cambridge, Cambridge, CB2 1AG, United Kingdom*

ARTICLE INFO

Article history:
Received: 05 November, 2023
Revised: 05 March, 2024
Accepted: 16 March, 2024
Online: 20 April, 2024

Keywords:
Robo-advisors
Full-scale optimization
Digital twin
Natural language processing
Artificial intelligence

ABSTRACT

Robo-advisors, fundamental to the financial services sector, have undergone substantial technological metamorphosis. Innovations in artificial intelligence, blockchain, cloud technology, augmented reality, and virtual reality have reshaped the financial industry's landscape. As automated investment solutions, robo-advisors are on the brink of further technological evolution. This comprehensive research amalgamates historical data, behavioral insights, and emerging market trends to provide technology-centric recommendations for the robo-advisory industry. Emphasizing the significance of a global perspective, the study explores the adoption of full-scale optimization in portfolio construction and the integration of digital twin capabilities. It delves into the burgeoning realm of Natural Language Processing facilitated by AI-driven chatbots in financial technology companies. These recommendations stand as pivotal pillars for steering the ongoing technological advancements of robo-advisors in the ever-evolving landscape of the financial sector.

1. Introduction

This paper extends work originally presented in 2023 at the 9th International Conference on Virtual Reality (ICVR 2023), held on May 14th, 2023, in Xianyang, China. Investors today leverage robo-advisors as powerful tools for efficient portfolio management. This is particularly beneficial for small investors navigating the complexities of crafting tailored investment strategies. Operating on sophisticated algorithms, robo-advisors autonomously oversee investment portfolios, increasing the likelihood of attaining financial goals.

Beyond portfolio management, these digital platforms extend their support to investors by providing a spectrum of financial planning services, ranging from budgeting and debt management to comprehensive retirement planning. This holistic approach empowers businesses to cultivate sustainable financial models, fortifying their overall financial health.

The emergence of robo-advisors signifies a transformative shift in investment advisory, offering a cost-effective alternative to traditional human advisors. Originating as fintech startups post the global financial crisis, robo-advisors have surged in popularity, with established financial institutions like Charles Schwab and BlackRock integrating robo-advisory services into their portfolios. Major players, including Bank of America and Wells Fargo, have also introduced their automated advisory services.

The United States dominates the robo-advisory market, hosting over 200 robo-advisors and commanding 57% of global robo-advisory investments as of 2017. With assets exceeding $400 billion in 2018 and an impressive annual growth rate of approximately 30%, projections indicate that by 2023, assets overseen by robo-advisors in the US will approach $1.5 trillion. Vanguard, Schwab Intelligent Portfolios, and Betterment stand out as major players by assets under management [1]. While the US leads, robo-advisory services are gaining traction globally. Europe hosts over 70 robo-advisors, with five managing assets exceeding €100 million. In Asia, particularly in developing markets, the proliferation of robo-advisors is propelled by a growing middle class and robust technological connectivity.

This study navigates through the industry's history, emerging trends, strengths, and weaknesses, pinpointing potential areas for growth. It posits three key predictions: a shift from Modern Portfolio Theory towards Full-Scale Optimization, the integration of Digital Twin capabilities to harness user data, and the

*Corresponding Author: Marco I. Bonelli, Mibonelli6@gmail.com

https://dx.doi.org/10.25046/aj090205

development of Natural Language Processing to elevate the user experience. These insights encapsulate the evolving landscape of robo-advisory services, positioning them for sustained global growth and innovation.

2. Literature Review

In recent years, the landscape of the financial services sector has been marked by a profound and dynamic transformation propelled by the integration of cutting-edge technologies, notably artificial intelligence (AI) [2-5]. This technological wave has seen the rise of robo-advisors as a pivotal force, fundamentally reshaping investment management and financial planning services, as highlighted by a study by Deloitte [6].

Nevertheless, while the preceding introduction offers a glimpse into this transformative landscape, it only scratches the surface of a multidimensional tableau. This study sets sail to delve into the complexities of robo-advisors on a global scale, focusing on their evolutionary journey and technological advancements. In this exploration, it aims to shed light on the disparities observed in the worldwide automated investment advisory sphere while acknowledging, for example an emerging market like India's with significantly higher growth, anticipated to reach a notable 8.5% annually. [1].

Guided by a central research question resonating across scholarly discourse—"Amid the AI surge, how can robo-advisors ascend to new frontiers, and what fault lines underlie the landscape of robo-advisory services globally?"[4], [7-11]—this intellectual exploration aspires to unearth insights that shed light on potential advancements and bridge the gaps awaiting in the realm of AI-augmented financial guidance and strategy.

2.1. Global Landscape of Robo-Advisors and AI Technology

Robo-advisors have garnered swift global recognition, primarily for their capacity to provide automated and personalized investment solutions, a facet reaffirmed in longitudinal studies [12]. These platforms, driven by intricate algorithms, consider individual risk profiles and financial objectives when creating and managing portfolios [13], [14] even extending to crypto asset allocation [15]. AI, a linchpin of robo-advisory technology, empowers comprehensive data analysis, pattern recognition, and predictive modeling [1]. AI-fueled robo-advisors, both from established institutions and startups, possess the potential to reshape the financial sector, making investment opportunities more accessible and cost-efficient [4], [6], [8], [15], [16]. Notably, robo-advisors provide the capability to incorporate global environmental constraints into stock forecasts [4]. A prevalent challenge in the adoption of advanced AI-powered robo-advisor technology is the lack of trust, a pivotal topic in customer studies globally and within regional markets [3], [8], [10], [11].

2.2. Advancing Technology Through AI

The AI explosion has ushered in new horizons for enhancing robo-advisory capabilities [4], [8]. The incorporation of AI into robo-advisors has unlocked the potential for more precise predictions, streamlined risk assessment, and optimized portfolio management [17]. Techniques such as machine learning, deep learning, and natural language processing have contributed to the refinement of investment strategies and the enhancement of the customer experience [11]. By harnessing the capabilities of AI, robo-advisors stand poised to offer more resilient and personalized solutions, ultimately elevating investor satisfaction and engagement [10].

3. Applications, Advantages, and Insights from Behavioral Studies

3.1. Efficient Functioning of Robo-Advisors

Robo-advisors adopt a personalized approach to shaping investment decisions, initiating the creation of individualized investment plans tailored to specific goals and risk preferences. Prospective clients undergo inquiries regarding their investment purpose, horizon, and readiness to endure market fluctuations, enabling robo-advisors to propose plans for various objectives, such as significant expenditures, retirement, income generation, or emergency fund building [18].

To ensure cost-effectiveness, robo-advisors employ online questionnaires to assess clients' needs, investment goals, and risk profiles. Automated algorithms utilize this information to recommend asset allocations across different types of assets, with portfolio optimization typically grounded in Modern Portfolio Theory [19]. The entire process, often completed online without human interaction, ensures efficient portfolio management.

3.2. Cost Optimization and Returns

Retail investors utilizing robo-advisors benefit from significantly lower management fees, typically ranging from 20% to 30% of fees charged by human financial advisors due to automation and scalability [20, 21]. Robo-advisors aim to reduce costs, discouraging fee-inducing practices like excessive trading. Studies indicate a substantial reduction in average retail investor expense ratios, leading to significant savings and reinvestment opportunities [22].

Utilizing passively managed ETFs, robo-advisors offer less expensive and transparent fee structures, empowering clients to determine costs easily. Research indicates that robo-advisors in the US have outperformed popular mutual funds and major equity indices in value-weighted returns between 2016 and 2019, emphasizing their cost-effectiveness and potential for better risk-adjusted returns [24].

3.3. Behavioral Impact on Retail Investors

Robo-advisors play a crucial role in mitigating behavioral biases, such as under-diversification, overtrading, and herd mentality, prevalent among retail investors [25-27]. These platforms, driven by algorithms, craft portfolios tailored to individual objectives and risk tolerance, effectively addressing concerns related to behavioral biases [28].

Research highlights that robo-advisors effectively counter under-diversification, reducing portfolio volatility and enhancing performance. The services act as a shield against behavioral biases associated with DIY investment approaches, providing a more disciplined investment strategy. While not entirely immune to biases, robo-advisors contribute to countering behavioral pitfalls such as the disposition effect, rank effect, and trend-chasing, ensuring more rational investment decisions [29]."

3.4. Current State of the Industry

Robo-advisory services are poised for a robust Compound Annual Growth Rate (CAGR) of 40% from 2021 to 2026, emphasizing the steady expansion of this sector. The industry is witnessing a shift toward advanced portfolio strategies, integrating artificial intelligence (AI) and machine learning to align with the ongoing wave of financial digitization [30, 31]. Hybrid robo-advisors, combining human and automated advice, dominate the sector, reflecting the enduring demand for human involvement in portfolio management [32].

The Covid-19 pandemic underscored the resilience of robo-advisors, with certain platforms doubling their client base during lockdowns. Platforms equipped with advanced technology showcased their ability to withstand market shocks, attracting a larger clientele [30]. Notable performers during volatile market conditions included Wealth Front, Morgan Stanley Inflation Conscious, and Personal Capital, emphasizing the appeal of robo-advisory services in navigating challenging economic landscapes.

4. Research Study

The Web-Based Research Design study revolves around gathering data from the Internet concerning global financial indicators. Information regarding financial markets was systematically collected using platforms like statista.com. Additionally, the study involved employing Google Trends to gauge the level of interest in financial technology, as outlined by Vasylchenko and Bonelli in 2023 [33]-[36]. The data collection process also encompassed obtaining comprehensive insights into the most popular global Robo-Advisors and AI-driven solutions.

4.1. Analysis Key Aspects:

During our analysis, we focused on evaluating the following key aspects:

1) Obtaining a comprehensive global perspective of the financial markets, both on a global scale.
2) Investigating the status of technology within the global landscape of Robo-Advisors.
3) Exploring opportunities for enhancing technology through the implementation of AI-driven solutions.
4) Exploring the status of technology within the global landscape of Robo-Advisors.

Statista [7], the publisher of financial data, allowed us to collect data about the global perspective of the financial market within the context of India. That includes key points of financial market development, actual financial data about market volume worldwide, and forecasts for India until 2030.

In the first stage, information about the most frequently mentioned robo-advisors online was filtered [37-45]. About 70 international Robo-Advisors (Advisors who work worldwide, in many countries) lists were formed. For each Robo-Advisor, data about user count, capitalization of assets managed by, and interest data about search requests about Advisors were collected.

The Betterment Robo-Advisor, serving 775,000 customers with assets totaling $36 billion, was chosen as the benchmark for comparing relative measures with other advisors. Operating globally, with its headquarters in the US, this data was utilized to compute the relative interest in other advisors worldwide, incorporating the average Google Trend values for Robo-Advisors [36]. The tools advisors use to help investors through Finance Advisors, Robo Advisors, and AI trading were classified into the next categories.

Table 1: Advisor tools

Tool Category	Definition
Risk Assessment and Profiling	Robo-advisors collect information about an investor's risk tolerance, investment goals, and financial situation to create personalized investment profiles.
Portfolio Creation	Based on the investor's profile, robo-advisors automatically build diversified portfolios using various asset classes like stocks, bonds, and alternative investments.
Asset Allocation	Robo-advisors use algorithms to determine the optimal distribution of assets in the portfolio based on the investor's risk profile and market conditions.
Rebalancing	Robo-advisors monitor the portfolio and periodically adjust the allocation to maintain the desired balance as market conditions change.
Tax Efficiency	Some robo-advisors optimize investment decisions for tax efficiency, including strategies like tax-loss harvesting to minimize capital gains taxes.
Goal-Based Investing	Robo-advisors help investors set specific financial goals, such as retirement or buying a home, and create investment strategies tailored to those goals.
Automated Trading	Many robo-advisors automatically execute trades based on pre-defined investment strategies and market conditions.
Cost Optimization	Robo-advisors often build portfolios using low-cost exchange-traded funds (ETFs) or index funds, aiming to minimize investment expenses.
Diversification	Robo-advisors ensure that portfolios are well-diversified across various assets to reduce risk and enhance returns.
Continuous Monitoring	They continuously monitor the portfolio's performance and provide updates and notifications to investors.
Education and Insights	Robo-advisors offer educational resources and insights to help investors understand their investments and make informed decisions.
User-Friendly Interfaces	Many robo-advisors provide easy-to-use online platforms and mobile apps for investors to track their portfolios, make changes, and access information.
Savings and Investing Automation	Some robo-advisors allow users to automate regular contributions to their investment accounts, helping investors save and invest consistently over time.

Accessibility	Robo-advisors make investing accessible by offering lower account minimums compared to traditional financial advisors.
Personalized Recommendations	They provide recommendations for additional investments or adjustments based on changing market conditions and the investor's goals.
Behavioral Finance Strategies	Certain robo-advisors incorporate behavioral finance principles to help investors stay disciplined and avoid emotional decisions.
Financial Planning	Some robo-advisors offer comprehensive financial planning services, including retirement planning, estate planning, and goal tracking.
Transparency	Many robo-advisors provide transparent fee structures, allowing investors to see exactly how much they are paying for the service.
Customer Support	Although automated, robo-advisors often offer customer support through various channels to assist with account-related queries.

4.2. Exploring opportunities for enhancing technology through the implementation of AI-driven solutions.

Advisor tools and AI Enhancement options were collected in the next table [37-50]. The next conjoint of these papers was summarized in Table 2 [36].

Table 2: AI Enhancement

AI Enhancement	Description
Advanced-Data Analysis	AI analyzes extensive financial data from various sources to identify patterns and insights for informed investment strategies.
Predictive Modeling	AI creates predictive models using historical data to estimate future market movements, enabling more accurate investment decisions.
Natural Language Processing (NLP)	NLP technology helps robo-advisors understand text-based information, such as news and social media, to assess market sentiment.
Personalized Recommendations	AI analyzes individual investor profiles, preferences, and risk tolerance to provide highly tailored investment recommendations.
Dynamic Asset Allocation	AI algorithms adjust asset allocations in real time based on changing market conditions to keep portfolios aligned with goals.
Behavioral Finance Analysis	AI studies investor behavior patterns and biases, offering strategies to avoid behavioral pitfalls and make rational decisions.
Automated Tax Optimization	AI optimizes tax-related decisions, like tax-loss harvesting, to minimize tax liabilities and enhance after-tax returns.
Risk Assessment and Management	AI performs advanced risk analysis beyond traditional metrics to provide accurate risk assessments and effective risk management.
Alternative Data Utilization	AI processes unconventional data sources to gain insights into potential investment opportunities, enhancing decision-making.
Portfolio Stress Testing	AI simulates the impact of market scenarios on portfolios, helping investors understand potential performance under different conditions.
Market Timing	AI identifies potential entry and exit points by analyzing technical indicators and historical trends to enhance investment timing.
Real-time Alerts	AI generates real-time alerts for significant market events or changes, keeping investors informed of potential adjustments.
Natural Language Interaction	AI-enabled interfaces offer conversational interactions, allowing users to engage with the platform using natural language.
Fraud Detection	AI identifies unusual patterns in account activity, detecting potentially fraudulent activities and enhancing security.
Continuous Learning	AI systems learn from interactions and continuously improve recommendations and strategies based on feedback and outcomes.
Market Sentiment Analysis	AI analyzes social media and news to gauge market sentiment, incorporating sentiment analysis into investment decisions.
Deep Learning for Complex Analysis	AI techniques like deep learning analyze intricate financial models and predict complex outcomes for more accurate decisions.
Cognitive Computing	AI processes unstructured data, making sense of diverse information sources and enhancing decision-making capabilities.
Enhanced User Experience	AI-driven interfaces offer personalized dashboards, interactive visualizations, and user-centric features, enhancing the investor experience.

Google Trends used search engine data to determine interest in finding interesting information during the last year. Developer reports were used to determine the actual number of users and assets under management. Sometimes, if user data was not accessible, Google Play market data were used to determine the current count of users for applications.

These are the worldwide robo-advisor list used for this study: Capital One Investing, Schwab Intelligent Portfolios, Etoro, Marcus Invest, TD Ameritrade Essential Portfolios, XM Group,

Ally Invest, Fidelity Go, SoFi Invest, E*TRADE Core Portfolios, Acorns, Charles Schwab Intelligent Portfolios, DBS, FxPro, Morgan Stanley Access Investing, Betterment, Interactive Brokers, Wealthfront, M1 Finance, Fundrise, Plus500, Nutmeg, Ellevest, Syfe, StashAway, OCBCBank, Merrill Guided Investing, Kristal.AI, Titan Invest, SigFig, CitiWealthBuilder, Modern Advisor, BMO Smartfolio, MSCI, FutureAdvisor, FP Markets, Qtrade Guided Portfolios, Empower/Personal Capital, Endowus, CI Direct, and Axos Invest [36].

4.3. Research Methodology and Evaluation

In our pursuit of understanding the global robo-advisor landscape, we used Google Trends and developers' reports to assess user interest and metrics. The study, anchored in a worldwide robo-advisor list featuring major players like Capital One Investing and Schwab Intelligent Portfolios, formed the basis of our approach. Our analysis covered 94.7% of the global robo-advisor market, excluding smaller Indian players. We curated robo-advisor lists, explored AI roles, and utilized Google Trends for global and Indian insights on popularity and enhancements [36].

4.4. Interest Measures

To complete our quantitative analysis, we took advantage of the power of Google Trends, recognizing its dominance as the predominant search engine globally, capturing approximately 96-98% of search requests. This expansive coverage allowed us to gain insights into the worldwide interest trends related to robo-advisors, offering a perspective on the evolving landscape of financial technology.

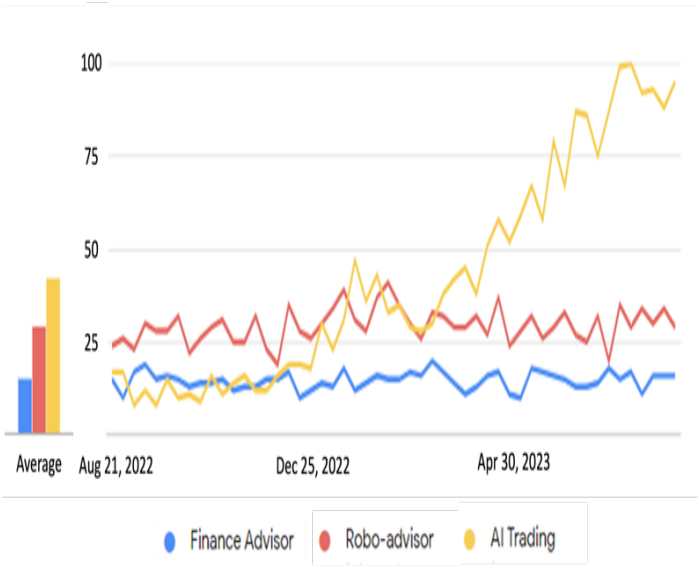


Figure 1: Interest Over Time

Figure 1, titled "Interest Over Time," illustrates the frequency of searches for key terms such as "Finance Advisor," "Robo Advisor," and "AI Trading" over the past year.

The data derived from Google Trends serves as a valuable tool for monitoring and comparing the popularity of specific search queries, enabling a dynamic understanding of user preferences and interests in the financial technology domain [36].

5. Findings

5.1. Global perspective of the financial market

In Figure 2, we present a comprehensive overview of the global financial services sector, revealing a projected market value of around $31 billion based on insights from a report by BLinC Invest [51]. This figure encapsulates the expansive scope and economic significance of the financial services industry on a global scale. The forecasted trajectory of the market points towards a notable growth trend, with an anticipated annual growth rate of 6.8%. Notable is the growth in emerging markets like India, which is over double the rate compared to the average overall growth rate [51]. This projection underscores the sector's resilience and its capacity to evolve, signaling opportunities for expansion and innovation within the financial services domain. Analyzing these projected values provides crucial insights into the dynamic nature of the global financial landscape, setting the stage for strategic considerations and potential advancements in the industry.

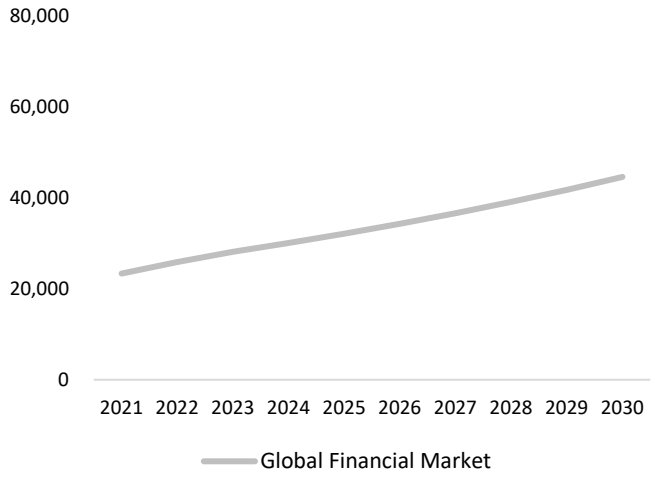


Figure 2: Global Financial Markets: Statista (2023)

5.2. Investigating the status of technology within the global landscape of Robo-Advisors

Our investigation reveals a global landscape of robo-advisors, where the most influential players collectively command 94.7% of the world's assets. To discern their prominence, we utilized assets under management as the benchmark for ranking these advisors. This criterion was selected to accommodate the diverse nature of advisors catering to a myriad of customer segments.

Figure 3, as depicted in our research [36], visually underscores the dominance of these major global advisors and serves as a key reference point for understanding their significance in the industry.

The diverse clientele of these Robo-Advisors spans a wide spectrum of needs and preferences, encompassing various categories such as Individual Investors, Quick Traders, International Funds, Banks and Institutions, Retirement Planners, Novice Investors, Ethical and Socially Responsible Investors, High-Net-Worth Individuals, Young Investors, Passive Investors, Digital Natives, Small Businesses, Educational Savings, Real Estate Investors, Customized Portfolios, Tax-Efficient Investors, Conservative Investors, Alternative Investments.

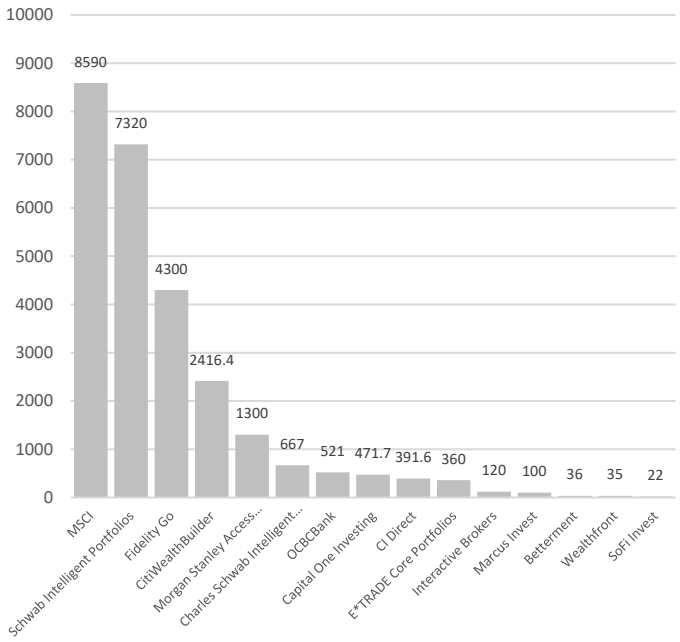


Figure 3: Highlights the dominance of the most influential global advisors [36].

5.3. Exploring Opportunities for Enhancing Technology through AI-Driven Solutions

In our quest to advance technology within the financial sector, we conducted a thorough examination of AI-driven solutions, shedding light on significant trends and user preferences. Figure 4 visually represents the evolving landscape of AI trading and its intersection with user interests, serving as a compelling testament to the comprehensive analysis [36].

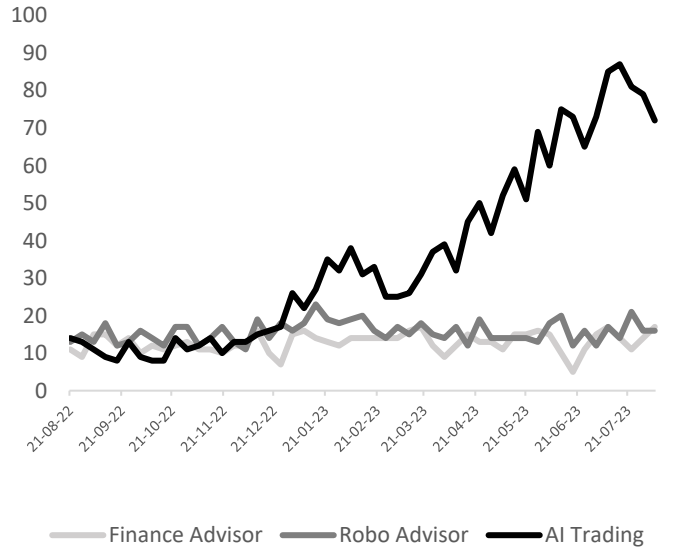


Figure 4: Illustrates the trend in the interest surrounding AI trading [36].

Figure 4 offers valuable insights into the dynamic preferences surrounding AI trading and financial advisory services. The depicted global trends showcase a consistent pattern; however, a closer examination reveals intriguing nuances.

The data indicates a noticeable divergence in the level of interest between Finance Advisors and Robo-Advisors. Finance Advisors command a notable interest score of 12.8 points, while Robo-Advisors register at 15.6 points. This suggests a preference for the automated and technologically driven approach of Robo-Advisors, even when faced with the typically higher costs associated with traditional Finance Advisors[36].

A noteworthy disparity emerges when specifically examining the interest in AI trading. Globally, AI trading garners a considerable interest score of 9.3 points, surpassing the interest in Robo-Advisors at 3 points. This signifies a distinct global enthusiasm for AI-driven trading solutions, potentially fueled by the allure of sophisticated algorithms and data-driven decision-making processes.

In summary, the trends illustrated in Figure 4 underscore the increasing prominence of AI trading in the financial technology landscape. The global preference for Robo-Advisors over traditional Finance Advisors, combined with the growing interest in AI trading, indicates a notable shift toward technologically advanced solutions in the financial advisory realm [36]. These insights not only mirror current user preferences but also provide valuable foresight into the trajectory of AI-driven financial services in the future (see Fig. 4 for illustration)

6. Development of Robo-Advisory Services

6.1. Full-Scale Optimization

The initial research focus revolves around the investment strategies employed by robo-advisors to serve their clients. Given the saturation within this sub-sector, the remaining robo-advisors face the imperative of distinguishing themselves from their competitors to both attract assets and maintain customer loyalty, thus heralding the anticipation of evolving investment methodologies [29].

Modern Portfolio Theory (MPT), initially embraced within this subsector, offered a well-established, Nobel Prize-backed approach known for its safety and appeal to investors seeking automated investment solutions. MPT is grounded in mean-variance analysis, striving to maximize expected returns while minimizing associated risks [52].

At its core, MPT employs mean-variance analysis, an approach that aims to strike a balance between maximizing expected returns and minimizing associated risks. Mathematically, it is represented as follows:

$$CapRp = \sum i = 1nwi * Ri \quad (1)$$

$$\sigma p^2 = \sum i = 1n \sum j = 1n\, wi * wj * \sigma i * \sigma j * \rho ij \quad (2)$$

Here, Rp represents the expected portfolio return, wi signifies the weight of each asset in the portfolio, Ri is the expected return of each asset, σi denotes the standard deviation of each asset, ρij reflects the correlation between assets i and j, and σp is the standard deviation of the portfolio.

While MPT has been a valuable tool, the robo-advisory landscape has evolved, revealing its limitations. These limitations include an overreliance on data inputs, an undue concentration in

specific asset classes, and the assumption of normal asset return distributions, which can be problematic in the presence of extreme returns.

Rather than shifting towards entirely new strategies, numerous firms are instead choosing to enhance the existing MPT investment framework, thus giving rise to the phenomenon known as the "Multidimensional Improvement of Modern Portfolio Theory" [29].

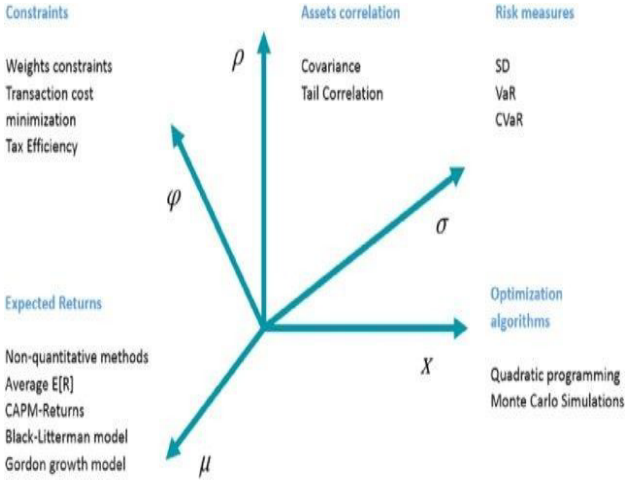


Figure 5: Many robo-advisory firms are adopting the Multidimensional Improvement of Modern Portfolio Theory, which incorporates additional factors, inputs, methods, and considerations in the newer portfolio models [53].

Figure 5 offers a summary of the multidimensional improvement of Modern Portfolio Theory (MPT), although it does not encompass all the methodologies currently being implemented by robo-advisors [29].

To overcome the limitations inherent in Modern Portfolio Theory (MPT), an advanced approach known as the "Multidimensional Improvement of Modern Portfolio Theory" (MPT+) has emerged. MPT+ goes beyond the traditional MPT framework by considering a broader array of factors, inputs, methods, and considerations in portfolio models. This extended framework is designed to overcome some of the constraints associated with MPT [53].

The MPT+ introduces an alternative formula for portfolio optimization:

$$Rp = \sum i = 1nwi * Ri - \lambda * \sum i = 1nwi * Ai \quad (3)$$

$$\sigma p^2 = \sum i = 1n \sum j = 1nwi * wj * \sigma ij \quad (4)$$

In this formula, Ai represents alternative risk measures, and λ is a risk aversion parameter. By incorporating these additional risk measures, MPT+ offers a more nuanced approach to portfolio optimization, effectively addressing the limitations of traditional MPT [53]

As the robo-advisory landscape evolves, a growing disparity emerges between conventional investment strategies and newer, academically validated approaches. Modern Portfolio Theory (MPT), while widely adopted, reveals limitations such as data reliance, estimation errors, and over-concentration in specific asset classes within portfolios. It assumes normal asset return distributions, neglecting "fat tails" observed in empirical evidence, and overlooks crucial factors like time horizon and diverse financial circumstances [54-55].

MPT's reliance on quadratic utility presents further challenges, assuming uniform emotions towards market volatility and an inaccurate representation of investor utility changes with wealth accumulation [4]. Alternative utility functions have been proposed, acknowledging that wealth benefits increase with growth, investors prefer upward market volatility, and relative risk aversion remains constant. Despite advantages, some argue these functions may oversimplify preferences, suggesting S-shaped or kinked utility functions for a more accurate representation [54-55].

Addressing these limitations, advanced strategies like Full-Scale Optimization, Risk Parity, Scenario Optimization, and Risk Parity with Skewness Risk have been introduced. Schwab Intelligent Portfolios has already implemented Full-Scale Optimization, showcasing practical advancements in the robo-advisory sector. Research suggests firms adopting these methodologies attract higher client investment levels [56].

As the robo-advisory sector evolves, these advanced strategies are poised to play a crucial role in meeting diverse investor preferences and optimizing portfolio performance. Particularly, the Multidimensional Improvement of Modern Portfolio Theory (MPT+) has gained traction, providing a more comprehensive and adaptable approach compared to traditional MPT, incorporating various risk measures, and accommodating different asset return distributions [57].

6.2. Digital Twin

Within the evolving landscape of robo-advisory, a new frontier emerges with the exploration of Digital Twin capabilities (Figure 6). This concept introduces the prospect of creating a virtual replica of the user's environment, a digital twin that serves as the foundation for personalized advice delivered by robo-advisors [58]. It should be underscored that this approach remains speculative, warranting thorough regulatory examination to establish trust in the responsible use of consumer data.

The envisioned integration of Digital Twin capabilities into robo-advisor software marks a theoretical leap. Data extracted within the virtual environment shapes a digital replica. This conceptual duplicate becomes a tool for robo-advisors to tailor advice, utilizing insights gleaned from the user's simulated surroundings [42].

The process can be summarized as:

$$Digital\ Twin\ Creation: DT = f(Data)$$

Where DT represents the digital twin, and Data signifies the data collected from the user's environment.

Digital twins hold promise in delivering more personalized recommendations compared to current robo-advisors. These replicas dynamically adapt guidance based on evolving client circumstances. For instance, if a client publicly shares on social media that they've become a parent, the robo-advisor could promptly suggest a college savings plan. As the software accumulates data, the digital twin's understanding of the client's financial situation becomes more comprehensive, enabling

adjustments to their financial strategies [58]. The recommendations can be quantified as:

$$Recommendations = g(Client\ Data))$$

where recommendations denote advice from the digital twin, and Client Data represents data from the client's virtual environment.

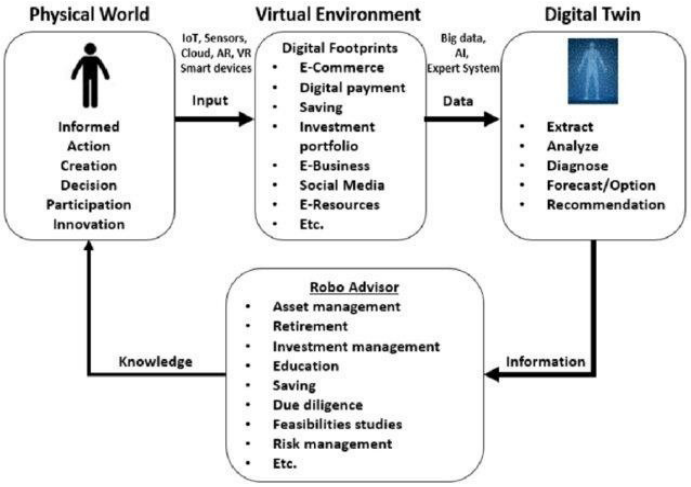


Figure 6: The integration of Digital Twin capabilities into robo-advisor software is a theoretical approach that involves creating a digital replica of the user's environment. The data collected from the virtual environment would be used to create the digital twin, which the robo-advisor could then use to provide advice to the real user [58].

Leveraging insights from the digital twin, the robo-advisor can offer a spectrum of recommendations, including adjusting the client's portfolio, providing guidance on insurance, suggesting new accounts, and delivering insights on significant purchases. Additionally, the robo-advisor can anticipate both near-term and long-term scenarios and recommend appropriate actions.

The digital nature of robo-advisors provides higher accessibility for clients with internet connectivity, a notable advantage over traditional human advisors. However, integrating digital twins into robo-advisory services poses challenges and risks, requiring robust data security measures to handle substantial volumes of sensitive client information and build trust in the technology[58].

6.3. Natural Language Processing

Natural Language Processing (NLP), a cornerstone of Artificial Intelligence (AI), is a crucial force bridging human language and machine comprehension. Rooted in Computational Linguistics, NLP enables machines to understand and manipulate human language, evolving beyond traditional predefined instructions [59].

The rise of Machine Learning and AI signifies a transformative era, empowering machines to make nuanced decisions based on real-time contextual information. This shift is evident in technologies like voice-activated virtual assistants (e.g., Alexa, Google Home), showcasing NLP's role in enhancing human-machine interactions [59].

Figure 7 illustrates NLP's profound impact on the financial landscape. Sourced from Analytics Vidhya, it exemplifies NLP's transformative role in refining communication between users and robo-advisors, shaping the financial services landscape.

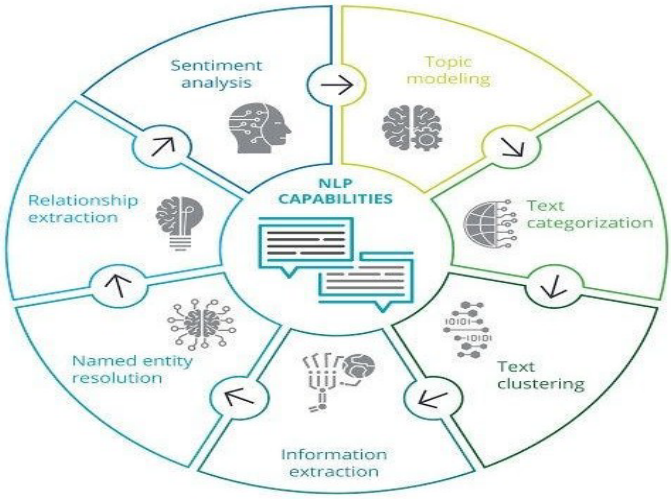


Figure 7: The impact of NLP within the financial industry and robo advisers is palpable and far-reaching, as exemplified in the above figure scenarios (Source: Analytics Vidhya)

In the financial realm, grappling with time-consuming methods for document retrieval and extracting insights from vast, unstructured data has been a persistent challenge for decades. The intricate nature of unstructured data makes analysis and deciphering a formidable task. However, the emergence of Natural Language Processing (NLP) revolutionized this landscape, delivering unprecedented efficiency and accuracy [60].

NLP, prized for its innate ability to discern patterns within unstructured data, provides a profound advantage to financial institutions and, more recently, robo-advisers. Its applications span text classification, information retrieval, natural language generation, and natural language understanding. These capabilities collectively yield streamlined and powerful results, establishing NLP as an indispensable tool for automating numerous processes within the financial sector, particularly benefiting robo-advisers [61].

Robo-advisors in finance leverage Natural Language Processing (NLP) for significant enhancements:

1. **Personalized Customer Care:** NLP-powered chatbots serve as 24/7 personal clerks for clients. They efficiently manage repetitive tasks and provide financial advice, aiding clients in saving, on-time payments, and overdraft prevention.
2. **Automated Trading Platforms:** NLP is vital for high-frequency trading in robo-advisors. It enables the analysis of data, including news and press releases, to identify trading patterns. These tools operate almost in real-time, aiding brokers in validating trading decisions.
3. **Monitor Investor's Sentiment:** NLP tools track mentions of companies and assess sentiment in news coverage. Understanding the tone of news articles provides valuable insights into investor perceptions.
4. **Upgrade Quality of Analyst Reporting:** NLP enhances analyst reporting by swiftly accessing relevant information. This enables analysts to create more detailed reports and offer superior advice.

5. **Easier Data Retrieval:** Robo-advisors use NLP to implement conversational interfaces, allowing effortless searches within vast internal databases, even for complex queries.

These advancements in robo-advisors are powered by the capabilities of NLP, contributing to efficiency and effectiveness in various financial processes. That can be quantified as follows:

Let:

- ***PC:*** *Personalized Customer Care efficiency.*
- ***ATP****: Efficiency in Automated Trading Platforms.*
- ***MCS****: Monitoring Company Sentiment.*
- ***UQAR****: Upgrade in Quality of Analyst Reporting.*
- ***EDR****: Easier Data Retrieval.*

And:

- ***Total Benefits (TB)*** *= PC + ATP + MCS + UQAR + EDR*

Quantitative Representation:

$$TB = PC + ATP + MCS + UQAR + EDRbg$$

Where:

- **PC** is measured as the percentage increase in the efficiency of customer care and financial advice provided.
- **ATP** is quantified as the percentage improvement in trading efficiency.
- **MCS** is evaluated based on the accuracy of sentiment analysis and its impact on investment decisions.
- **UQAR** is expressed as the percentage increase in the quality of analyst reporting.
- **EDR** is assessed as the percentage reduction in the time required for data retrieval and its impact on decision-making processes.

By assigning values to each of these components, the overall benefits of implementing NLP in their robo-advisory services can be accurately quantified.

In [62], the authors underscores the potential of reciprocal conversations between human and robo advisors through AI chatbots to foster trust and enhance the perceived relationship between clients and RA firms. By investing in NLP technology on their platforms, RA companies can significantly enhance the user experience and cultivate customer loyalty. AI chatbots bring a conversational element that can potentially increase the likelihood of clients adhering to the portfolio guidance provided by robo-advisors, improve investors' perceptions of the investment firm, and create a more positive and engaging experience for customers [62].

Furthermore, AI chatbots enable investors to obtain more personalized portfolio recommendations based on customer data. However, the preference for conversationalist RAs over non-conversational ones does raise potential ethical implications, as investors might be inclined to follow suboptimal investment decisions [62]. In [56], the athors delves into the advantages of AI-powered consultants for enhancing decision-making in businesses. It highlights the introduction of psychological barriers and emphasizes the potential risks associated with imbuing robo-counselors with social design elements to make them appear more human-like.

7. Regulating Robo-Advisors: Navigating Legal and Regulatory Frontiers

The rise of robo-advisors, internet-based advisory services driven by algorithms, has revolutionized the financial landscape, offering cost-effective solutions compared to traditional advisors. However, the unique client-machine interaction raises intricate legal questions, challenging regulators globally. This section explores the regulatory aspects surrounding robo-advisors, emphasizing their distinctive nature and the responses of major regulatory bodies [63].

Regulators, especially in major jurisdictions, grapple with adapting existing frameworks to address the peculiarities of robo-advisory services. While the European Union's financial intermediary framework is seen as broadly applicable, the challenge lies in interpreting these rules without stifling innovation. Globally, regulators are cautious, often opting for a conservative stance. For instance, the U.S. Securities and Exchange Commission (SEC) issued guidance cautioning investors against sole reliance on robo-advisors, highlighting the need for individual research. Meanwhile, the Financial Industry Regulatory Authority (FINRA) in the U.S. expressed concerns about the adequacy of robo-advisor advice, emphasizing the importance of human intervention [64].

7.1. Elements of Robo-Advisory Regulation:

a. **Ongoing Duties, Compliance, and Authorization:** The regulation of robo-advisors involves authorization by national regulators, aligning with the traditional framework for financial intermediaries. However, the challenge lies in ensuring that robo-advisors comply with ongoing duties and standards of care, particularly concerning communication with clients. Regulatory bodies globally, including the U.S. SEC and the Australian Securities and Investments Commission (ASIC), have issued guidance emphasizing the need for disclosure and investor protection [64].

b. **Collective Fallibility and Market Failure:** Concerns surrounding market stability and systemic risks due to the collective decision-making processes of robo-advisors have been raised. While robo-advisors are yet to achieve a market share that poses a threat to systemic stability, regulators must be vigilant. The potential for instantaneous reactions during market crises and the risk of market distortions due to uniform advice patterns are areas where future regulatory focus may be warranted.

This section addresses concerns about the effectiveness of communication between robo-advisors and clients and delves into the need for a standard of investor protection equivalent to that for human advisors. It emphasizes that while robo-advisors need not be flawless, they should provide a similar level of investor protection as their human counterparts. The importance of recourse options for clients in case of flawed advice and the role of regulatory mechanisms in preventing or responding to such instances are underscored.

As robo-advisors continue to gain prominence, the regulatory landscape must evolve to accommodate their unique characteristics. While current frameworks offer a foundation, thoughtful interpretation, and adjustments are crucial to ensure a level playing field. Balancing innovation with investor protection remains a key challenge for regulators globally, urging ongoing dialogue and adaptability in the face of a dynamic financial landscape [63].

In conclusion, the regulatory framework for robo-advisors requires careful calibration to meet the evolving dynamics of the financial landscape. This section advocates for an investor protection standard akin to that applied to human advisors, recognizing the importance of addressing communication concerns and ensuring recourse options for clients. As the prominence of robo-advisors continues to grow, a nuanced and adaptable regulatory approach is paramount. Striking a balance between fostering innovation and safeguarding investor interests stands as an ongoing challenge, urging regulators to engage in continuous dialogue and adaptability. This proactive stance will contribute to the sustainable development of robo-advisory services in a rapidly changing financial ecosystem.

8. Pros and Cons of Robo-advisors

8.1. Pros

1. **Lower Costs** Robo-advisors tend to have lower costs than traditional financial advisors because they operate more efficiently, reducing the workload of human advisors.
2. **Automated Investing** Robo-advisors can automatically manage the investment portfolios of businesses, streamlining the investment process for enterprises and saving time.
3. **Efficient Financial Planning** Robo-advisors can offer a more efficient financial planning process, aiding businesses in achieving their financial goals.

8.2. Cons

1. **Lack of Personalization** Robo-advisors may not offer services as personalized as human consultants, potentially falling short of fully meeting the specific needs of businesses.
2. **Risk Management** Robo- advisors may not provide sufficient expertise in risk management strategies for enterprises, necessitating additional support in this domain.
3. **Lack of Emotional Connection** Robo- advisors may struggle to establish the same level of emotional connection as human consultants, which may result in a diminished human element in the consulting services offered to businesses.

As a result, robo-advisors can undoubtedly assist investors in constructing a more efficient business model, particularly in areas such as financial planning, investment, and risk management. However, they may encounter limitations when it comes to addressing their specific needs, and the absence of the human factor can potentially reduce the emotional connection in the consulting services provided.

9. Conclusive Remarks

In conclusion, this comprehensive analysis, rooted in the exploration of technological advancements within the robo-advisory sector, integrates insights from academia, industry reports, and emerging trends. The primary goal is to offer a technically grounded perspective that informs practical decision-making in a competitive landscape.

The recommendations emerging from this study herald a new era in robo-advisory technology. First and foremost, the adoption of advanced portfolio management methodologies, like Full-Scale Optimization (FSO), is pivotal for robo-advisory firms to gain a distinctive edge. Second, the incorporation of digital twin capabilities offers a future where ultra-personalized services are tailored to each unique client. Lastly, the infusion of Natural Language Processing (NLP) into robo-advisor software promises substantial enhancements in customer satisfaction, portfolio guidance adherence, and client-advisor relationships.

These implications reverberate across the spectrum of industry stakeholders, guiding firms, regulators, and investors. Firms can employ these technical recommendations to initiate conversations and explore the potential of integrating these technologies. Regulators responsible for oversight, must vigilantly monitor the development of these trends as robo-advisors evolve. For investors, this research provides a valuable framework to assess prospective platforms.

In the realm of research, our web-based design incorporated data gathered from platforms like statista.com, Google Trends, and developer reports. This study delved into key aspects, encompassing a global perspective of financial markets, an investigation into the status of technology within the global landscape of Robo-Advisors, and an exploration of opportunities for enhancing technology through the implementation of AI-driven solutions.

The pathway for future research extends towards delving into alternative investment strategies like Risk Parity, Scenario Optimization, and Risk Parity with Skewness Risk, as well as their adaptability within the robo-advisory sub-sector. Further studies can delve into the enduring impact of AI chatbots on the intricacies of the client-robo-advisor relationship, particularly as the younger generation inherits wealth and navigates its utilization.

While this study aims to offer technical insights, it encounters certain limitations. The availability of pertinent data points and the relatively nascent nature of the robo-advisory sector may constrain the breadth of current academic research. Nevertheless, this research sets the stage for technical innovation and paves the way for future explorations and advancements within the robo-advisory sub-sector.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Statista, Finance - Value and Forecasts, 2022. https://www.statista.com/outlook/dmo/app/finance/india [on 05.08.2023]

[2] G. Bayramoglu, An Overview of the Artificial Intelligence Applications in Fintech and Regtech, 291–298, 2021, doi:10.1007/978-981-33-6811-8 15.

[3] D. Belanche, L. V. Casal, C. Flavi´an, "Artificial Intelligence in FinTech: understanding robo-advisors adoption among customers," Industrial Management Data Systems, 119, 1411–1430, 2019, doi:10.1108/IMDS-08-2018-0368.

[4] M. I. Bonelli, E. S. Dongul, "Robo-Advisors in the Financial Services Industry: Recommendations for Full-Scale Optimization, Digital Twin Integration and Leveraging Natural Language Processing Trends," 268–275, IEEE, 2023, doi:10.1109/ICVR57957.2023.10169615.

[5] L. D. Bernardis, ROBO ADVISORY Artificial intelligence applied to finance: An industry-specific analysis, comparing incumbents to startups: the Money-Farm vs YellowAdvice, LAP LAMBERT Academic Publishing, 2021.

[6] Deloitte, "2021 investment management outlook," https://www2.deloitte.com/jp/en/pages/financial-services/articles/im/2021-investment-management-outlook.html, 1–26, 2021.

[7] K. Wahabi, Lakhchini, Rachid Wahabi, "Artificial Intelligence Machine Learning in Finance: A literature review," International Journal of Accounting, Finance, Auditing, Management and Economics, **3**, 437–455, 2022, doi: 10.5281/zenodo.7454232.

[8] A. Bhatia, A. Chandani, R. Atiq, M. Mehta, R. Divekar, "Artificial intelligence in financial services: a qualitative research to discover robo-advisory services," Qualitative Research in Financial Markets, **13**, 632–654, 2021, doi:10.1108/QRFM-10-2020-0199.

[9] V. Joshi, Dr. Manjari Sharma, "Artificial Intelligence and Financial Decision Making – Review of efficacy in Usage within Financial Organizations in India," Turkish Online Journal of Qualitative Inquiry (TOJQI), **12**, 12138–12160, 2021.

[10] R. Manrai, K. P. Gupta, "Investor's perceptions on artificial intelligence (AI) technology adoption in investment services in India," Journal of Financial Services Marketing, **28**, 1–14, 2023, doi:10.1057/s41264-021-00134-9.

[11] Nain, S. Rajan, "Algorithms for better decision-making: a qualitative study exploring the landscape of robo-advisors in India," Managerial Finance, **49**, 1750–1761, 2023, doi:10.1108/MF-01-2023-0055.

[12] M. Shanmuganathan, "Behavioural finance in an era of artificial intelligence: Longitudinal case study of robo-advisors in investment decisions," Journal of Behavioral and Experimental Finance, **27**, 100297, 2020, doi:10.1016/j.jbef.2020.100297.

[13] M. Sarea, A. H. Elsayed, S. A. Bin-Nashwan, Artificial Intelligence and Islamic Finance, Routledge, 2021, doi:10.4324/9781003171638.

[14] W. L. Shih, C. Wei, H. Lin, P. Chang, "RETRACTED: Selecting and balancing market portfolio using artificial intelligence and fuzzy Mult objective decision-making model," Concurrent Engineering, **30**, 382–398, 2022, doi:10.1177/1063293X221115768.

[15] G. Babaei, P. Giudici, E. Raffinetti, "Explainable artificial intelligence for crypto asset allocation," Finance Research Letters, **47**, 102941, 2022, doi: 10.1016/j.frl.2022.102941.

[16] S. Deo, "The Under-appreciated Regulatory Challenges posed by Algorithms in FinTech. Understanding interactions among users, firms, algorithm decision systems regulators," 1–299, 2022, doi:10.48462/opus4-4288.

[17] Burnmark, "Digital Wealth," 2017. https://www.burnmark.com/uploads/reports/Burnmark_Report_Apr17_ D igital_Wealth.pdf (accessed on 1/16/2023).

[18] W. Lam, "Robo-Advisors: A Portfolio Management Perspective," 1–105, 2016.: https://economics.yale.edu/sites/default/files/2023-01/Jonathan_Lam_Senior%20Essay%20Revised.pdf

[19] Bjerknes, A. Vukovic, A. A. Gaivoronski, "Automated advice: A portfolio management perspective on robo-advisors," 2017.

[20] Backend Benchmarking, "The Robo Report and Robo Ranking. The Robo Report." 2022. https://www.theroboreport.com/ (accessed on 01/18/ 2023)

[21] S. Utkus, A. Rossi, "Who Benefits from Robo-advising? Evidence from Machine Learning," 2020.: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3552671.

[22] D. Jung, V. Dorner, F. Glaser, S. Morana, "Robo-Advisory," Business Information Systems Engineering, **60**, 81–86, 2018, doi:10.1007/s12599-018-0521-9.

[23] R. Tao, C. Su, Y. Xiao, K. Dai, F. Khalid, "Robo advisors, algorithmic trading and investment management: Wonders of fourth industrial revolution in financial markets," Technological Forecasting and Social Change, **163**, 120421, 2021, doi:10.1016/j.techfore.2020.120421.

[24] B. Barber, Y. Lee, Y. Liu, T. Odean, "Just How Much Do Individual Investors Lose by Trading?" Review of Financial Studies, **22**, 609–632, 2009, doi:10.1093/rfs/hhn046.

[25] W. N. Goetzmann, A. Kumar, "Equity Portfolio Diversification*," Review of Finance, **12**, 433–463, 2008, doi:10.1093/rof/rfn005.

[26] FROOT, D. Scharfstein, J. C. Stein, "Herd on the Street: Informational Inefficiencies in a Market with Short-Term Speculation," The Journal of Finance, **47**, 1461–1484, 1992, doi:10.1111/j.1540-6261.1992.tb04665.x.

[27] S. Chandler, "Robo-advisors offer automated investing services at a low cost —here's how to tell if they're right for you," 2020: https://www.businessinsider.com/personal-finance/what-is-a-robo- advisor (accessed on 1/17/2023).

[28] F. D'Acunto, N. Prabhala, A. G. Rossi, "The Promises and Pitfalls of Robo-Advising," The Review of Financial Studies, **32**, 1983–2020, 2019, doi:10.1093/rfs/hhz014.

[29] Research, Markets, "Robo-advisory Services Market - Growth, Trends, COVID-19 Impact, and Forecasts (2022 - 2027)," 2022.: https://www.researchandmarkets.com/reports/5120011/roboadvisory-services-market-growth-trends (accessed on 01/18/ 2023).

[30] R. Dive, "The Use of Digital Technology and Automation in Wealth Management are creating Growth Opportunities for the Robo Advisory Market Globally," 2022.

[31] J. P. Costa, "Vanguard research Quantifying the investor's view on the value of human and robo-advice," 2022. :https://advisors.vanguard.com/iwe/pdf/ISGHVD.pdf (accessed on 02/18/ 2023).

[32] B. Benchmarking, "The Robo Report and Robo Ranking." 2022. : https://www.theroboreport.com/ (accessed on 02/20/ 2023)

[33] Beketov, K. Lehmann, M. Wittke, "Robo Advisors: quantitative methods inside the robots," Journal of Asset Management, **19**, 363–370, 2018, doi:10.1057/s41260-018-0092-9.

[34] Anshari, M. N. Almunawar, M. Masri, "Digital Twin: Financial Technology's Next Frontier of Robo-Advisor," Journal of Risk and Financial Management, **15**, 163, 2022, doi:10.3390/jrfm15040163.

[35] M. B. A. Vasylchenko, "Web research and Google Trend data for Robo-Advisors study," 2023.

[36] M. L. Sette, "Best Robo-Advisors. Investopedia. Reviewed by S. Silberstein. Fact checked by M. M. Klammer," 2023. https://www.investopedia.com/best-robo-advisors-4693125

[37] Benson, "12 Best Robo-Advisors of August 2023," 2023.

[38] T. Tepper, "5 Best Robo-Advisors Of August 2023. Forbes Advisor," 2023. https://www.nerdwallet.com/best/investing/robo-advisors

[39] M. Frankel, "Best Robo-Advisors for August 2023,". https://www.fool.com/the-ascent/buying-stocks/best-robo-advisors/

[40] E. Gravier, "The 5 Best Robo-Advisors When You Want to Be Hands Off with Your Investments. CNBC," 2023. https://www.cnbc.com/select/best-robo-advisors/

[41] D. Sherarrow, "Best Robo Advisors in the U.S. August 2023," 2023. https://www.investing.com/brokers/robo-advisors/

[42] I. Pino, "The Top Robo-Advisor Offerings for 2023." 2023. https://fortune.com/recommends/investing/best-robo-advisors/

[43] M. C. Zinn, "Best Robo-Advisors for Investing for August 2023," 2023. https://www.cnet.com/personal-finance/investing/advice/best-robo-advisor/

[44] J. Fell, "Best Robo Advisory: Top 6 Robo Advisors (2023)," 2023. https://www.entrepreneur.com/guide/investing/robo-advisors

[45] F. Allen, G. Xian, "A survey of fintech research and policy discussion," 2021, doi:10.21799/frbp.wp.2020.21.

[46] M. I. Bonelli, "Macroeconomic Factors and SENSEX Performance in India: Unveiling the Post-Liberalization Era (1980-2020)," Indian Journal of Research in Capital Markets, Forthcoming, 2023.

[47] L. D. Bernardis, "Artificial intelligence applied to finance: Robo advisory: An industry-specific analysis, comparing incumbents to startups in business model sustainability: MoneyFarm vs YellowAdvice," 2020.

[48] Deloitte, "2021 investment management outlook,": https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/financial-services/2021-investment-management-outlook.pdf.

[49] X. Cheng, F. Guo, J. Chen, K. Li, Y. Zhang, P. Gao, "Exploring the Trust Influencing Mechanism of Robo-Advisor Service: A Mixed Method Approach," Sustainability, **11**, 4917, 2019, doi:10.3390/su11184917.

[50] BLNC, "BLnC Invest,". https://www.blincinvest.com

[51] T. Idzorek, A step-by-step guide to the Black-Litterman model, 17–38, Elsevier, 2007, doi:10.1016/B978-075068321-0.50003-0.

[52] R. O. Michaud, "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?" Financial Analysts Journal, **45**, 31–42, 1989, doi:10.2469/faj.v45.n1.31.

[53] M. Best, R. Grauer, "On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results," Review of Financial Studies, **4**, 315–342, 1991, doi:10.1093/rfs/4.2.315.

[54] T. Adler, M. Kritzman, "Mean-variance versus full-scale optimization: In and out of sample," Journal of Asset Management, **7**, 302–311, 2007, doi: 10.1057/palgrave.jam.2250042.

[55] C. Back, S. Morana, M. Spann, "When do robo-advisors make us better investors? The impact of social design elements on investor behavior," Journal of Behavioral and Experimental Economics, **103**, 101984, 2023, doi:10.1016/j.socec.2023.101984.

[56] M. Beketov, K. Lehmann, M. Wittke, "Robo Advisors: quantitative methods inside the robots," Journal of Asset Management, **19**, 363–370, 2018, doi:10.1057/s41260-018-0092-9.

[57] M. Anshari, M. N. Almunawar, M. Masri, "Digital Twin: Financial Technology's Next Frontier of Robo-Advisor," Journal of Risk and Financial Management, **15**, 163, 2022, doi:10.3390/jrfm15040163.

[58] B. Mishra R. Agrawal, "Natural Language Processing in Artificial Intelligence," Apple Academic Press, 2019.

[59] T. Rouzbeh A. Shirvani, Y. Keneshloo, "Natural Language Processing Advancements By Deep Learning: A Survey," https://arxiv.org/, 2021, doi:10.48550/arXiv.2003.01200.

[60] R. Kliman, B. Arinze, Cognitive Computing: Impacts on Financial Advice in Wealth Management, 11–23, 2019, doi:10.1007/978-3-319-93299-6 2.

[61] C. Hildebrand, A. Bergner, "Conversational robo advisors as surrogates of trust: onboarding experience, firm perception, and consumer financial decision making," Journal of the Academy of Marketing Science, **49**, 659–676, 2021, doi:10.1007/s11747-020-00753-z.

[62] Maume, "Regulating Robo-Advisory," SSRN Electronic Journal, 2018, doi:10.2139/ssrn.3167137.

[63] M. L. Fein, Regulation of Robo-Advisers in the United States, 105–131, 2021, doi:10.1007/978-3-030-40818-3 7.

[64] R. Steennot, "Robo-advisory services and investor protection," Law and Financial Markets Review, **15**, 262–277, 2021, doi:10.1080/17521440.2022.2153610.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 9, No. 1, 33-41 (2024)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Strengthening LoRaWAN Servers: A Comprehensive Update with AES Encryption and Grafana Mapping Solutions

Sheikh Tareq Ahmed*, Annamalai Annamalai, Mohamed Chouikha

Electrical & Computer Engineering, Prairie View A&M University, Texas, 77445, USA

ARTICLE INFO

Article history:
Received: 15 November, 2023
Revised: 06 January, 2024
Accepted: 06 January, 2024
Online: 20 January, 2024

Keywords:
TTN
IoT
LoRaWAN
Grafana
AES
RC4

ABSTRACT

This work enhances the LoRaWAN server framework, focusing on an innovative approach for robust security and dynamic data visualization in network management. Migrating from RVC4 to AES encryption, it fortifies the network's defense against cyber threats, a crucial advancement in IoT security. Furthermore, the integration with Grafana's mapping plugin capitalizes on geolocation data, a strategic element for network oversight and IoT data analysis. In-depth configuration and application of this plugin are explored, revealing substantial benefits for network administrators and end-users. Expanded discussions, backed by new experimental data, illustrate the real-world efficacy of these technological improvements. The research substantially enriches the understanding of LoRaWAN's technological evolution, addressing vital aspects of IoT security and geolocation integration. The outcomes are expected to resonate significantly within academic circles and practical domains, particularly in reinforcing IoT and network security. This work marks a significant stride in the progression of LoRaWAN technologies, with implications that extend well into the broader landscape of network management and security in the digital age.

1. Introduction

"In the evolving realm of the Internet of Things (IoT), LoRaWAN (Long Range Wide Area Network) has become a crucial technology, enabling applications from smart city infrastructure to industrial automation. Building on our prior work on LoRaWAN server implementations and data visualization tools [1], this paper addresses the crucial aspects of security and data management in LoRaWAN's expanding network. We transition from RVC4 to the more robust AES (Advanced Encryption Standard) for enhanced security [1, 2]. AES, while computationally heavier, offers superior encryption and resistance against cyber-attacks, a necessary upgrade in the IoT's dynamic threat landscape [3, 4].

Besides security enhancements, we integrate Grafana's map plugin to advance geolocation tracking in LoRaWAN, enriching network management and user experience. This integration allows for nuanced visualization of geolocation data, essential in diverse IoT applications like asset tracking and urban planning [5, 6]. The implementation involves configuring Grafana with LoRaWAN servers, mapping geolocation data, and developing custom dashboards for real-time monitoring, thus aiding in decision-making processes and supporting advanced analytics [7, 8]. The proliferation of mobile communication and the Internet of Things (IoT) has significantly improved daily life but also introduces substantial security challenges [3, 9, 10]. A body of research dedicated to fortifying IoT security [3, 11, 12] has emerged. In [11], the authors developed a security architecture blending cyber, physical, and social elements to protect diverse IoT environments. In [12], the author highlighted the need for comprehensive security measures, including user authentication and robust access control, particularly vital in government and industry applications. In [13], the researcher outlined key IoT security challenges, such as key management and privacy protection, and suggested blockchain as a viable solution. Studies [14] have identified specific vulnerabilities in LoRaWAN security, especially in servers utilizing RC4 encryption, limited by the low computational power of LoRa devices. Addressing these vulnerabilities, the implementation of low-power AES encryption in LoRaWAN servers emerges as a crucial solution. This approach not only aligns with the power constraints of IoT devices but also significantly enhances the security framework, mitigating the identified threats and providing a more robust defense mechanism for the IoT infrastructure.

However, challenges such as ensuring geolocation data accuracy and managing large data volumes are addressed through robust data management strategies and scalable cloud solutions [8]. In conclusion, the shift to AES and Grafana's map integration mark significant advancements in LoRaWAN and IoT, enhancing network

*Corresponding Author: Sheikh Tareq Ahmed, Email: sahmed13@pvamu.edu

security and operational efficiency. Future research will aim to optimize these integrations and expand IoT capabilities to meet the digital world's growing demands [15].

2. Overview

2.1. LoRaWAN Overview

LoRaWAN, a cornerstone technology in the Internet of Things (IoT) landscape, has significantly evolved since its inception, primarily addressing the need for long-range, low-power communication solutions in various IoT applications. This protocol, part of the Low Power Wide Area Network (LPWAN) family, was developed to enable efficient, long-distance wireless communication with minimal energy consumption, making it ideal for remote sensors and devices in smart cities, agriculture, and industrial monitoring. Its evolution mirrors the growing demands of IoT for connectivity that balances range, power efficiency, and bandwidth, revolutionizing how devices communicate over vast distances without the need for extensive power resources [16].

LoRaWAN (Long Range Wide Area Network) has emerged as a frontrunner in the Internet of Things (IoT) landscape, revolutionizing the way devices communicate and data is collected. Embracing a spread spectrum modulation technique known as Chirp Spread Spectrum (CSS), LoRaWAN enables long-range, low-power data transmission. It operates in the unlicensed Industrial, Scientific, and Medical (ISM) bands, offering superior penetration capabilities amidst challenging environments [15].

LoRaWAN's architecture comprises three key elements:

1. LoRaWAN End Devices: These battery-powered devices gather and transmit data to LoRaWAN gateways.
2. LoRaWAN Gateways: Acting as intermediaries, LoRaWAN gateways receive data from end devices and relay it to the network server.
3. LoRaWAN Network Server: The network server manages the network, routing data from gateways to application servers and handling device registration and authentication.

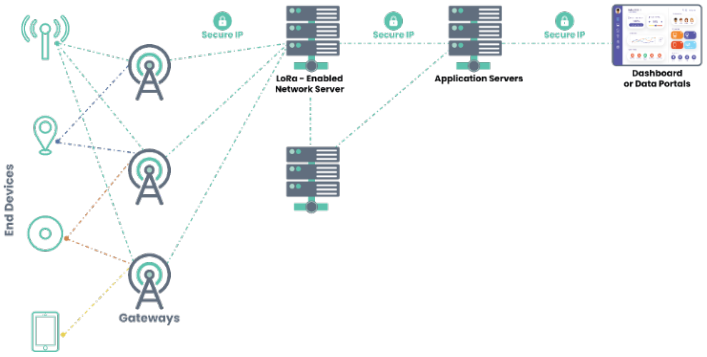


Figure 1: LoRaWAN Network Diagram

3. The Things Stack V3

The Things Stack V3 emerges as a pivotal player in the LoRaWAN ecosystem, providing a comprehensive and scalable platform for managing LoRaWAN deployments. Its API-driven approach simplifies integration with various applications and backend systems [16].

The Things Stack V3's architecture encompasses four primary components:

1. Join Server: The Join Server securely manages device credentials and facilitates device activation, ensuring only authorized devices can join the network.
2. Network Server: The Network Server orchestrates data routing, relaying messages between gateways and application servers. It also handles device management tasks, such as device registration and status updates.
3. Application Server: The Application Server acts as the interface between the LoRaWAN network and user applications. It processes incoming data, parses it into meaningful information, and triggers predefined actions.
4. Console: The Console serves as the primary user interface, providing a centralized platform for managing the LoRaWAN network, including creating and configuring organizations, gateways, devices, and applications.

As a critical component of the LoRaWAN infrastructure, The Things Stack V3 stands out for its robust and flexible framework, offering an API-centric approach that seamlessly integrates with a variety of applications and backend systems [17].

3.1. RC4 Stream Cipher in Join Server

The RC4 stream cipher is used in the Join Server component of The Things Stack V3 to generate security keys for new devices joining the network. The RC4 stream cipher is a fast and efficient algorithm that is well-suited for this application [18].

Specifically, the RC4 stream cipher is used to generate a random number that is then used as the encryption key for the device's AES encryption key. The AES encryption key is used to encrypt the device's credentials, which are then sent to the network server. The network server uses the same RC4 stream cipher to decrypt the device's credentials and validate them [19].

The security concern with using RC4 in the Join Server component of The Things Stack V3 is that it is a stream cipher that is not considered to be secure anymore. RC4 has been shown to be vulnerable to a number of attacks, including the Fluhrer, Mantin, and Rivest (FMR) attack.

To address this security concern, The Things Stack V3 should use a more secure cipher, such as AES, to generate security keys for new devices joining the network. AES is a block cipher that is considered to be much more secure than RC4. It is also more efficient, so it will not have a significant impact on the performance of the Join Server [19].

Using AES in the Join Server component of The Things Stack V3 will help to ensure the confidentiality of the device's credentials and protect the network from attack.

3.2. AES Encryption

AES Encryption:

- Plaintext Input: The process starts with a plaintext block of 128 bits.
- Key Expansion (Key Schedule): The AES key is expanded into several round keys using Rijndael's key schedule.
- Initial Round:
  - AddRoundKey: Each byte of the state is combined with the round key using bitwise XOR.
- Main Rounds (9, 11, or 13 rounds depending on key size):
  - SubBytes: A non-linear substitution step where each byte is replaced with another according to a lookup table (S-box).
  - ShiftRows: A transposition step where each row of the state is shifted cyclically a certain number of steps.
  - MixColumns: A mixing operation that operates on the columns of the state, combining the four bytes in each column.
  - AddRoundKey: The round key is added to the state.
- Final Round (no MixColumns):
  - SubBytes
  - ShiftRows
  - AddRoundKey
- Ciphertext Output: The result is a block of 128-bit ciphertext.

AES Decryption: Decryption in AES is not merely the encryption process in reverse order; instead, each step has an inverse operation.

- Ciphertext Input: Start with a block of 128-bit ciphertext.
- Inverse Key Expansion: Similar to encryption, the keys are derived for each round.
- Initial Round:
  - AddRoundKey
- Inverse Main Rounds:
  - InvShiftRows: The inverse of the ShiftRows step.
  - InvSubBytes: The inverse of the SubBytes step using the inverse S-box.
  - AddRoundKey
  - InvMixColumns: The inverse of the MixColumns step.
- Final Round (no InvMixColumns):
  - InvShiftRows
  - InvSubBytes
  - AddRoundKey
- Plaintext Output: The result is the decrypted plaintext.

AES Example:

- Plaintext: "HELLO WORLD AES"
- Key: "SECUREKEY123456" (128-bit key)

Note: In practice, both plaintext and key would be in binary format. Our example uses text for clarity. AES Encryption:

- Plaintext Input: "HELLO WORLD AES" (Converted to 128-bit binary format)
- Key Expansion:
  - Original Key: "SECUREKEY123456"
  - Expanded into several round keys.
- Initial Round:
  - AddRoundKey: Combine plaintext with the first round key.
- Main Rounds (9 rounds for 128-bit key):
  - SubBytes: Substitute bytes based on S-box.
  - ShiftRows: Shift rows of the matrix.
  - MixColumns: Mix columns using a special mathematical function.
  - AddRoundKey: Add the next round key to the matrix.
- Final Round:
  - SubBytes
  - ShiftRows
  - AddRoundKey
- Ciphertext Output: A 128-bit binary block (represented as text for the example, e.g., "GHIJK ZYXWV UTSA").

AES Decryption: Using the ciphertext "GHIJK ZYXWV UTSA" and the same key "SECUREKEY123456":

- Ciphertext Input: "GHIJK ZYXWV UTSA"
- Inverse Key Expansion: Use the same key expansion process as in encryption.
- Initial Round:
  - AddRoundKey
- Inverse Main Rounds:
  - InvShiftRows
  - InvSubBytes
  - AddRoundKey
  - InvMixColumns
- Final Round:
  - InvShiftRows
  - InvSubBytes
  - AddRoundKey
- Plaintext Output: "HELLO WORLD AES"

3.3. Grafana in IoT

Grafana stands out as a dynamic open-source platform for data visualization and monitoring, especially in the realm of the Internet of Things (IoT). Its primary strength lies in enabling users to craft interactive dashboards and graphical representations, offering a real-time window into data from diverse sources, including IoT devices and systems [7].

3.3.1. Key Features and Capabilities

1. Support for Various Data Sources: One of Grafana's most notable features is its extensive compatibility with numerous data sources, encompassing time series databases, metrics aggregators, and IoT platforms. This wide-ranging support facilitates seamless integration of Grafana into IoT ecosystems, allowing for effective visualization and real-time monitoring of data generated by IoT devices [6].

2. Real-Time Data Visualization: Grafana empowers users to visualize IoT data through customizable dashboards that provide instant insights. These dashboards can be tailored to display data in various formats such as graphs, charts, and maps, making complex data sets understandable and actionable [6].

3. Alerting and Notification System: Beyond visualization, Grafana offers robust alerting functionalities. Users can define alerts based on specific data thresholds or conditions. When these criteria are met, Grafana triggers notifications, aiding in prompt detection and response to anomalies or critical situations in IoT environments [5].

4. Insights into System Performance: By visualizing data from IoT systems, Grafana aids users in comprehending the performance and behavior of these systems. This understanding is crucial for making data-driven decisions and optimizing IoT operations [6].

3.3.2. Grafana Availability Options

Grafana's versatility is further highlighted by its availability in different formats, catering to varied user needs:

1. Grafana Cloud: This is a subscription-based service offering of Grafana, charged based on usage. The free tier of Grafana Cloud has certain limitations, particularly regarding the number of active users and data storage capacity.

2. Grafana OSS (Open Source Software): Reflecting its open-source roots, Grafana OSS allows users to install the platform on their servers or personal computers. It offers functionalities akin to Grafana Cloud, making it a viable option for those preferring local deployment.

3.3.3. Advanced Features

A) Metrics, Logs, and Traces Exploration: Grafana excels in providing tools for exploring metrics from databases, crucial for developing insightful dashboards and troubleshooting database queries. It presents data visualizations in various formats like lines or bars and enables users to inspect JSON data directly fetched from databases.

B) Customization and Flexibility: Grafana's interface and features are highly customizable, allowing users to tailor their monitoring experience to their specific needs. This flexibility is vital in adapting to the diverse requirements of IoT data monitoring and analysis.

C) Community and Plugin Ecosystem: With a robust community and a growing repository of plugins, Grafana continually evolves, integrating new data sources, visualization tools, and features. This community-driven development ensures Grafana remains at the forefront of data visualization technologies.

4. Experimental Setup

In our innovative experiment, we will be employing a combination of tools and hardware meticulously selected for their efficiency in data handling and visualization in IoT systems. Grafana, a sophisticated data visualization tool, InfluxDB as our chosen time-series database, and Node-Red for intuitive workflow management will comprise our software suite. The LoRaWAN server duties will be managed by ChirpStack.

For our hardware infrastructure, we are utilizing the Raspberry Pi 3, a compact and cost-effective single-board computer produced by the Raspberry Pi Foundation. This third-generation model, released in February 2016, is equipped with a 1.2 GHz 64-bit quad-core ARM Cortex-A53 processor, 1 GB of RAM, and onboard 802.11n Wi-Fi and Bluetooth 4.1, making it a versatile and powerful choice for a variety of applications [17].

Our selection for the LoRaWAN gateway is the Dragino LPS8, known for its low-power consumption and compact design, suitable for both outdoor and industrial settings. It can operate independently with an external power source or a solar panel, ideal for remote or autonomous installations. The LPS8 is versatile, supporting both Class A and Class C devices, and provides connectivity options via Ethernet or Wi-Fi. It also includes a GPS module for location services and features a user-friendly web-based interface for simple setup and configuration of the LoRaWAN network [17].

For endpoint device communication, we've chosen the ESP-32, a system-on-chip microcontroller with integrated Wi-Fi and Bluetooth, offering a blend of low-cost and energy efficiency for applications like home automation and wearable technology [17].

Furthermore, our sensor nodes will connect to the LoRaWAN gateway using the Hope-RF RF-95W, a transceiver module by Hope-RF designed for long-range IoT and M2M applications. Based on the Semtech SX1276 chip, it supports LoRaWAN protocols, providing up to 6 km range in open fields and data transmission speeds up to 300 kbps. The module's integrated antenna and compatibility with several development platforms, including Arduino, make it an excellent choice for our IoT projects [9].Incorporating into our experimental setup, we utilized a 6mm GPS module for precise location tracking, enhancing our array of IoT devices. This compact module is essential for applications requiring accurate geolocation data, allowing for the monitoring and deployment of IoT solutions

that rely on spatial information. Its integration into our system aligns with the increasing demand for location-aware IoT applications, providing an additional layer of functionality to our robust experimental framework. The GPS module's small form factor ensures it can be embedded seamlessly with our other components without imposing significant power or space requirements, which is critical for maintaining the efficiency and scalability of IoT devices.

A Grafana dashboard offers a tailored, interactive interface accessible via web, which aggregates and presents data drawn from one or multiple data sources. The construction of a Grafana dashboard is modular, consisting of rows and individual panels that can each portray various data visualizations such as graphs, tables, or singular metrics [6]. Employed for the surveillance and analytical assessment of diverse data streams, Grafana dashboards excel in presenting server statistics, application efficacy, and commercial indicators. Their versatility is showcased in their ability to display information in numerous formats, whether as instantaneous graphs or cumulative tables. These dashboards are collaborative and integrative, designed to be shared among users or incorporated into other digital services. The inception of a Grafana dashboard can be from the ground up or can be expedited through pre-existing templates. Panels can be appended and tailored to manifest data from chosen sources, allowing for personalization in the presentation and configuration of the dashboard. For the intermediary data collection from peripheral devices, Node-Red operates using MQTT protocol, which then forwards this data to an InfluxDB instance active on a Raspberry Pi-4. The query functionalities of InfluxDB, facilitated by InfluxQL, are employed to channel data into the database. This assemblage of data handling and visualization employs a quartet of tools: an MQTT broker, Node-Red, InfluxDB, and Grafana. Grafana is typically initiated by executing the Grafana-server application through the command line. For those needing to operate Grafana as a persistent Windows service, NSSM can be downloaded and used to manage Grafana in this capacity [20].

Creating a Grafana dashboard involves a series of steps that enable the visualization of data collected from various sources, including IoT devices. This process is complemented by the use of middleware such as Node-Red to facilitate data collection and InfluxDB to store this data efficiently. Below is a step-by-step guide to developing a Grafana dashboard with an emphasis on incorporating a location tracking plugin, such as the World Map plugin:

4.1. Developing a Grafana Dashboard

I. Installation of Grafana

- Begin by downloading the latest version of Grafana from the official website.
- Follow the provided installation instructions to set up Grafana on your system.

II. Configuring a Data Source:

- After installation, navigate to the "Configuration" menu in Grafana.
- Select "Data Sources" and add the required data source (e.g., InfluxDB).

III. Dashboard Creation:

- Click the "+" icon on the left-hand menu and choose "Create Dashboard."
- Start with a new empty dashboard where you will add your visualizations.

IV. Adding Panels:

- Click on "Add Panel" to choose the type of visualization (graphs, tables, maps, etc.).
- Configure each panel using the editor to represent your data effectively.

V. Customizing Your Dashboard:

- Adjust the dashboard settings to modify layout, theme, and other visual elements.
- Organize your panels for a coherent flow of information.

VI. Publishing the Dashboard:

- Save your progress by clicking the "Save" button.
- To share the dashboard with others, select "Save and Publish."

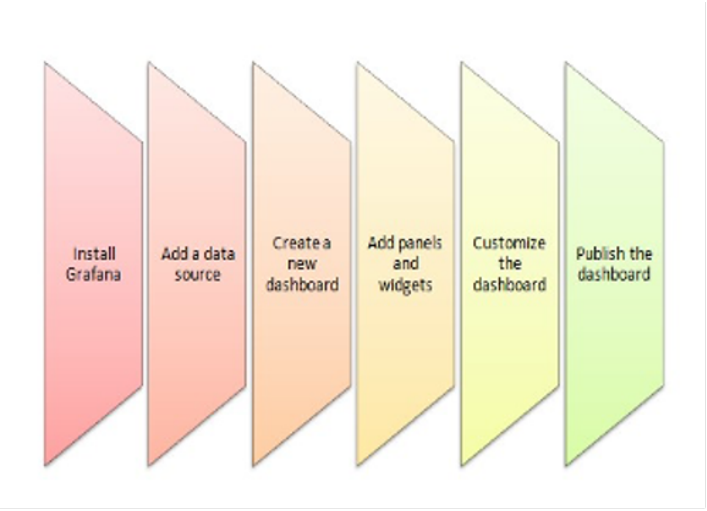


Figure 2: Grafana dashboard Installation[1]

4.2. Integrating Middleware and Data Collection

- Utilize Node-Red as an intermediary to collect data from edge devices via MQTT.
- Data is then forwarded to InfluxDB, which is hosted locally on a Raspberry Pi-4.

4.3. *Installing Grafana as a Windows Service (if applicable)*

If running Grafana on Windows and you wish to run it as a service, you can use NSSM (Non-Sucking Service Manager).

Download NSSM and follow the instructions to set up Grafana to run as a service in the background.

Grafana Dashboard

4.4. *Installing the Worldmap Plugin for Location Tracking*

VII. Install the Worldmap Plugin:

- Within Grafana, navigate to the ”Plugins” section from the sidebar.
- Search for the ”Worldmap Panel” plugin and click ”Install.”

VIII. Configuring the Worldmap Plugin:

- Once installed, add a new panel and select the Worldmap plugin as the visualization type.
- Link the panel to your geolocation data source, such as GPS coordinates from IoT devices.

IX. Customizing the Worldmap Panel:

- Use the Worldmap settings to define how data points are displayed, set up map visuals, and tailor marker colors and sizes.
- Configure queries to fetch the correct geospatial data from your InfluxDB.

X. Finalizing the Worldmap Setup:

After configuring, ensure that the map displays the data points correctly.Save your dashboard to retain the Worldmap panel configuration.

To implement AES in The Things Stack V3 instead of RC4, you will need to make some changes to the Join Server configuration. Here are the steps involved:

1. Generate an AES encryption key: This key will be used to encrypt the device's credentials. You can generate a random AES key using a secure random number generator.
2. Configure the Join Server to use AES: You will need to configure the Join Server to use the AES encryption key that you generated in step 1. This can be done by editing the Join Server configuration file.
3. Update the device firmware: You will need to update the device firmware to use AES. This can be done by compiling the device firmware with the AES encryption library.
4. Test the new code: Once you have made all of the necessary changes, you will need to test the new code to make sure that it is working correctly. You can do this by joining a device to the network and verifying that the device's credentials are being encrypted correctly.

Here is an example of how to configure the Join Server to use AES:

[join_server]

encryption_key = <your_aes_encryption_key>

Once you have made the necessary changes to the Join Server configuration, you will need to update the device firmware to use AES. This can be done by compiling the device firmware with the AES encryption library.

To compile the device firmware with the AES encryption library:

make USE_AES=1

Once we have compiled the device firmware with the AES encryption library, we will need to flash the new firmware onto the device.

Once we have flashed the new firmware onto the device, we will need to test the new code to make sure that it is working correctly. WE can do this by joining a device to the network and verifying that the device's credentials are being encrypted correctly.

To join a device to the network:

ttn-lw154f-join <dev_eui> <app_eui> <app_key>

Once we have joined the device to the network, we can verify that the device's credentials are being encrypted correctly by using the ttn-ctl tool.

Use the ttn-ctl tool to verify that the device's credentials are being encrypted correctly:

ttn-ctl devices devices get <dev_eui>

The output of the ttn-ctl command ‘ttn-ctl ‘ shows that the device's credentials are being encrypted with AES.

5. Result

In this experimental setup, we were able to successfully send data from our LoRa based sensor nodes to LoRaWAN gateway. We were able to receive and store data in our time-series database InfluxDB and later visualize it in Grafana.

For Grafana visualization we used temperature, humidity, barometric pressure sensor and air-quality sensors. We have also able to use the geolocation plugin to enable tracking facilities.

Figure 3: Grafana Dashboard[1]

Figure 4: World map plugin enabled

Packet sniffers are tools used to capture and analyze network traffic. When applied to TTN, packet sniffers first capture the LoRaWAN communication, which typically includes join requests, device data transmissions, and network server responses. This data is crucial for identifying potential vulnerabilities within the network's communication protocol. For instance, by analyzing join request packets, one can simulate replay attacks, where an attacker resends a join request to gain network access. Similarly, sniffing data packets can reveal patterns or encryption weaknesses, aiding in crafting sophisticated MITM or eavesdropping attacks that intercept and decode message payloads.

In our experiment within The Things Network (TTN) version 3, also known as The Things Stack V3, we implemented packet sniffing to analyze the network traffic between end devices and the TTN server. This process was essential for understanding the communication patterns and identifying potential vulnerabilities. To accomplish this, we accessed the gateway traffic through the TTN console. After logging in, we navigated to our specific gateway's console page, where we could monitor the live data traffic including uplink and downlink packets, as well as their associated metadata such as RSSI and SNR. For a more in-depth analysis, we utilized external tools, including Wireshark, by redirecting the gateway traffic to a local server. This setup allowed us to capture and examine the packet exchanges in real time, providing valuable insights for our security assessment of LoRaWAN implementations in TTN V3.

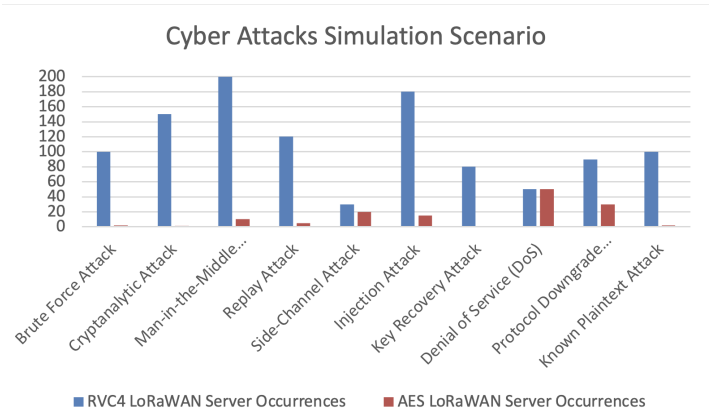


Figure 5: Result Summary

In our study, packet sniffing was employed on both AES and RC4-configured TTN servers to analyze vulnerabilities for various attack types. This involved capturing and scrutinizing network traffic to simulate attacks like Brute Force, MITM, Replay, and DoS, aiming to evaluate the servers' resilience. For advanced attacks such as Cryptanalytic, Injection, and Known Plaintext, packet sniffing facilitated the understanding of encryption patterns and potential weaknesses. The insights gained were instrumental in developing and refining attack simulations, providing a comprehensive security assessment of LoRaWAN implementations under TTN. After implementing the AES we have made 1000 attack attempts on both RVC4 and AES-based LoRaWAN servers, the occurrence of successful attacks we found out as follows:

Table 1: Simulation Of different attacks and Number of occurrences on both RVC4 and AES LoRAWAN server

Attack Type	RVC4	AES
Brute Force Attack	100	2
Cryptanalytic Attack	150	1
Man-in-the-Middle (MITM)	200	10
Replay Attack	120	5
Side-Channel Attack	30	20
Injection Attack	180	15
Key Recovery Attack	80	0
Denial of Service (DoS)	50	50
Protocol Downgrade Attack	90	30
Known Plaintext Attack	100	2

Notes: The dataset indicates various attack types and their occurrences in RVC4 and AES. AES shows a strong resistance to brute force attacks due to its key size and complexity, making it superior in this regard compared to RVC4. In the case of cryptanalytic attacks, AES is assumed to be more robust than RVC4. The data also suggests that AES includes measures to prevent Man-in-the-Middle (MITM) attacks, which might not be as robust in RVC4. For Replay Attacks, AES's secure nonce and timestamping are noted to be more effective. Both cryptographic systems show vulnerabilities to Side-Channel Attacks, although implementations can mitigate risks. AES's structured protocol and packet validation are observed to reduce the risk of Injection Attacks. In terms of

Key Recovery Attacks, AES's key schedule and algorithm design provide a high level of resistance. Denial of Service (DoS) attacks are seen to exploit network vulnerabilities rather than cryptographic weaknesses, affecting both systems equally. AES enforces higher security standards in protocol use, reducing the risk of Protocol Downgrade Attacks. Lastly, AES is not susceptible to Known Plaintext Attacks, which could be a concern for RVC4.

In our experiment on The Things Network (TTN) V3, we systematically generated and analyzed various attack types to assess network security. Here's how we approached each attack type:

Brute Force Attack: We attempted to access the network by systematically trying numerous password combinations, aiming to exploit weak credentials.

Cryptanalytic Attack: We analyzed the encrypted packets captured by our packet sniffer for vulnerabilities in the encryption algorithms used by the network.

Man-in-the-Middle (MITM): We intercepted communication between devices and the network to eavesdrop and potentially alter the data being transmitted.

Replay Attack: We captured valid data packets and retransmitted them to the network to test if the network would accept repeated or delayed packets.

Side-Channel Attack: We monitored indirect information from the network, like power consumption and emission patterns, to extract sensitive data.

Injection Attack: We introduced malicious data or commands into the network to observe how the system would react to unauthorized inputs.

Key Recovery Attack: We attempted to reverse-engineer the encryption keys used in the network by analyzing the captured encrypted traffic.

Denial of Service (DoS): We flooded the network with excessive traffic to test its ability to handle high load and potential service disruptions.

Protocol Downgrade Attack: We forced the network to revert to older, less secure protocols to exploit known vulnerabilities in those older systems.

Known Plaintext Attack: We used known plaintext data and its corresponding ciphertext to attempt decryption of other messages.

6. Conclusion

In conclusion, this paper has presented a thorough analysis of server implementations, revealing the need for a more secure architecture in LoRaWAN networks. By transitioning from RVC4 to AES encryption, we have demonstrated a significant strengthening of the network's security framework. The experimental results substantiate the superiority of AES in resisting various types of attacks, offering a more resilient and robust defense against the escalating threats in the IoT landscape. Moreover, the integration of advanced data visualization techniques using Grafana has provided a sophisticated solution for representing data in LoRaWAN networks. The implementation of the Worldmap plugin enhances the network management capabilities by utilizing geolocation data, which is instrumental in a wide range of IoT applications.

The experiments conducted have shown a marked improvement in security with the adoption of AES, reflecting in the reduced occurrences of successful cyber attacks. Additionally, the ability to visualize and monitor environmental conditions in real-time through Grafana dashboards has augmented the analytical capabilities of network administrators.

Future work will continue to explore optimization strategies for AES integration and expand the analytical functionalities of IoT networks. The evolving digital landscape demands continuous advancements in security protocols and data visualization tools to cater to the burgeoning needs of smart, connected environments. This paper's contributions lay a foundational step towards achieving more secure and efficient IoT networks, driving forward the potential of LoRaWAN in the IoT ecosystem.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment This research work is supported in part by the National Science Foundation (NSF) under grants # 2219611, 1910868 and 2200377. Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect NSF's views.

References

[1] S. T. Ahmed and A. Annamalai, "On Private Server Implementations and Data Visualization for LoRaWAN," in *2023 IEEE 13th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Penang, Malaysia, 2023, 342-347, doi: 10.1109/ISCAIE57739.2023.10165109.

[2] "IEEE Draft Standard for Low-Rate Wireless Networks Amendment Defining Support for Advanced Encryption Standard (AES)-256 Encryption and Security Extensions," in *IEEE P802.15.4y/D2*, October 2020, 1-20, 16 Dec. 2020.

[3] K. -L. Tsai, F. -Y. Leu, I. You, S. -W. Chang, S. -J. Hu and H. Park, "Low-Power AES Data Encryption Architecture for a LoRaWAN," in *IEEE Access*, **7**, 146348-146357, 2019, doi: 10.1109/ACCESS.2019.2941972.

[4] S. T. Ahmed and A. Annamalai, "Improving Geo-Location Performance of LoRa with Adaptive Spreading Factor," in *2023 IEEE 13th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Penang, Malaysia, 2023, 386-391, doi: 10.1109/ISCAIE57739.2023.10165296.

[5] Grafana Worldmap Panel, [Online]. Available: https://grafana.com/grafana/plugins/grafana-worldmap-panel/

[6] Grafana, [Online]. Available: https://grafana.com/grafana/

[7] J. Lee, Y. Souryal, K. Dongsoo, K. Kyubyung, and D. Koo, "Building a Private LoRaWAN Platform," 2019.

[8] A. A. Ahmed and S. Ahmed, "A Real-Time Car Towing Management System Using ML-Powered Automatic Number Plate Recognition," *Algorithms*, **14**, 317, 2021. [Online]. Available: https://doi.org/10.3390/a14110317

[9] J. Zhou, Z. Cao, X. Dong, and A. V. Vasilakos, "Security and privacy for cloud-based IoT: Challenges, countermeasures, and future directions," *IEEE Commun. Mag.*, **55**(1), 26-33, Jan. 2017.

[10] G. Choudhary, J. Kim, and V. Sharma, "Security of 5G-mobile backhaul networks: A survey," *J. Wireless Mobile Netw. Ubiquitous Comput. Dependable Appl.*, **9**(4), 41-70, Dec. 2018.

[11] H. Ning and H. Liu, "Cyber-physical-social based security architecture for future Internet of Things," *Adv. Internet Things*, **2**(1), 1-7, Jan. 2012.

[12] S. Li, T. Tryfonas, and H. Li, "The Internet of Things: A security point of view," *Internet Res.*, **26**(2), 337-359, Apr. 2016.

[13] H. Hui, X. An, H. Wang, W. Ju, H. Yang, H. Gao, et al., "Survey on blockchain for Internet of Things," *J. Internet Services Inf. Secur.*, **9**(2), 1-30, May 2019.

[14] P. I. R. Grammatikis, P. G. Sarigiannidis, and I. D. Moscholios, "Securing the Internet of Things: Challenges, threats, and solutions," *Internet Things*, **5**, 41-70, Mar. 2019.

[15] LoRa Technical Overview, [Online]. Available: https://lora-developers.semtech.com/documentation/tech-papers-and-guides/lora-and-lorawan/

[16] The Things Industries, [Online]. Available: https://www.thethingsindustries.com/docs/the-things-stack/

[17] ThingStack Installation, [Online]. Available: https://www.thethingsindustries.com/docs/getting-started/installation/

[18] T. D. B. Weerasinghe, "An effective RC4 stream cipher," in *2013 IEEE 8th International Conference on Industrial and Information Systems*, Peradeniya, Sri Lanka, 2013, 69-74, doi: 10.1109/ICIInfS.2013.6731957.

[19] N. Atikah, M. R. Ashila, D. R. Ignatius Moses Setiadi, E. H. Rachmawanto, and C. A. Sari, "AES-RC4 Encryption Technique to Improve File Security," in *2019 Fourth International Conference on Informatics and Computing (ICIC)*, Semarang, Indonesia, 2019, 1-5, doi: 10.1109/ICIC47613.2019.8985825.

[20] S. K. Dhar, S. Das, and N. Baidya, "Range enhanced IR-UWB remote biosensor system with integrated communication," in *2013 International Conference on Informatics, Electronics and Vision (ICIEV)*, Dhaka, Bangladesh, 2013, 1-5, doi: 10.1109/ICIEV.2013.6572633.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 5, 35-54 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Condition Assessment of Medium Voltage Assets: A Review

Eilin Gómez Mesino[*,1], Joaquín Caicedo[2], Miguel Mamaní[1], David Romero Quete[3], Andrés Cerón Piamba[4], Diego García Gómez[4], Guillermo Aponte Mayor[4], José Caicedo Erazo[4], Wilmar Moreno López[5], Edward Jay[5], Andrés Romero Quete[1]

[1]*Instituto de Energía Eléctrica, Universidad Nacional de San Juan–CONICET, San Juan, Argentina*

[2]*Facultad de Ingeniería, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia*

[3]*Departamento de Ingeniería Eléctrica y Electrónica, Universidad Nacional de Colombia, Bogotá, Colombia*

[4]*Grupo de investigación GRALTA, Escuela de Ingeniería Eléctrica y Electrónica, Universidad del Valle, Cali, Colombia*

[5]*Empresa de Energía del Archipiélago de San Andrés, Providencia y Santa Catalina, EEDAS S.A. E.S.P., San Andrés Islas, Colombia*

ARTICLE INFO

Article history:
Received:29 July, 2023
Accepted:25 September, 2023
Online: 30 October, 2023

Keywords:
Asset management
Condition assessment
Condition monitoring
Distribution network
Literature review
Medium-Voltage

ABSTRACT

Condition assessment of medium voltage assets is essential to ensure reliability and cost-effective operation of power distribution networks. This article presents a literature review of condition assessment of medium voltage assets related to a distribution system in a non-interconnected zone in Colombia, namely, power transformers, photovoltaic systems, switchgear, lines and cables, and instrument transformers. Advanced search rules are formulated to obtain bibliographic records of relevant academic literature from the database Scopus. The retrieved data are analyzed quantitatively to provide insights on the current state of research on the topic. Next, the most relevant academic papers for each medium voltage asset are selected and analyzed in a critical review along with more diverse literature including standards, technical reports, and white papers obtained through complementary searches. The results of the review show that several approaches have been formulated for condition assessment of medium voltage assets, ranging from traditional diagnostic methods to advanced artificial intelligence-based approaches. Moreover, research on some assets is already mature including power transformers, and photovoltaic systems, whereas other assets have been incipiently studied such as distribution and instrument transformers. Therefore, the need for deeper condition assessment research for these critical assets is highlighted. Research gaps are identified in the standardization and integration of condition assessment tools for distribution system operators.

1. Introduction

This paper extends a work presented in IEEE Biennial Congress of Argentina, ARGENCON 2022, [1]. This extended review covers other important Medium-Voltage (MV) assets operating in distribution networks. The review is a key component of a government-funded research project in Colombia aimed at implementing a digital platform for asset management of components at the archipelago San Andrés, Providencia, and Santa Catalina distribution network, in agreement with ISO 55000 [2].

In the islands, the distribution network consists of distributed generators (diesel and solar) which supply power to MV feeders at 34,5 kV and 13,2 kV, and Low-Voltage (LV) end customers at 208/120 V. The main assets in the distribution network are:

- Power Transformers (PT)
- Photovoltaic (PV) systems
- Distribution Transformers (DT)
- MV switchgear
- MV lines and cables
- Instrument Transformers (IT)

[*]Corresponding Author: Eilin Gómez Mesino, Universidad Nacional de San Juan, Av. Libertador Gral. San Martin 1109, egomez@iee.unsj.edu.ar

https://dx.doi.org/10.25046/aj080505

In this context, this paper offers a review of the main technical aspects related to condition assessment of distribution network assets. The study focuses on a non-interconnected grid.

The paper is structured as follows. The next subsections describe the methodology adopted for the literature review and the taxonomy to structure the analysis. Section 2 introduces the review of condition assessment of PT. Similarly, Section 3 reviews condition assessment of PV systems, followed by the review of DT, switchgear, lines and cables, and IT, in Sections 4 to 7. Section 8 outlines the discussion of results. Section 9 explores prospects for future research. Section 10 states the conclusions of the paper.

1.1. Methodology for the literature review

The literature review methodology is based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [3]. The advanced search rules illustrated in Figure 1, are applied in titles, abstracts, and keywords of the citation database Scopus to identify the most relevant literature related to condition assessment of MV assets. On the one hand, concepts related to condition assessment are included in each search rule and synonyms and associated concepts are also used according to Table 1. On the other hand, seven individual search rules are formulated to consider the most relevant assets of the distribution network in San Andrés Islands, i.e., (1) PT, (2) PV systems, (3) DT, (4) switchgear, (5) lines and cables, and (6) IT. In this case, synonym keywords and compound search formulas are defined according to Table 2.

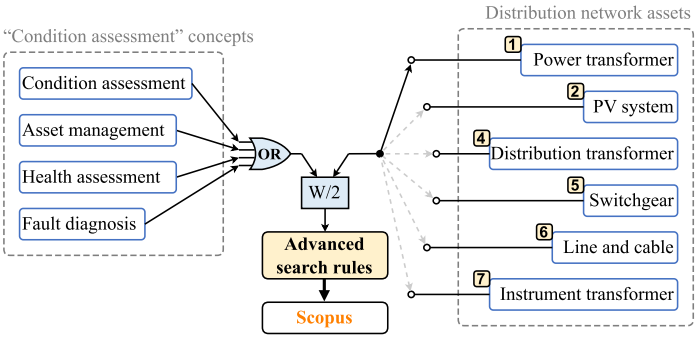


Figure 1: Advanced search rules applied in Scopus on April 1st, 2023.

Table 1: Set of synonym keywords related to condition assessment concepts.

Concept	Set of synonym keywords
Condition assessment	"condition assessment" OR "condition monitoring" OR "remaining useful life" OR "performance index"
Asset management	"asset management"
Health assessment	"health monitoring" OR "health index"
Fault diagnosis	"failure detection" OR "failure diagnosis" OR "fault detection" OR "fault diagnosis"

Table 2: Set of synonym keywords and compound search formulas for MV assets.

Asset	Set of synonym keywords / compound searches
PT	"power transformer"
PV system	"photovoltaic generat*" OR "pv generat*" OR "photovoltaic system" OR "pv system"
DT	"distribution transformer"
Switchgear	switchgear OR "power circuit breaker" OR disconnector OR recloser OR ("circuit breaker" W/2 (distribution OR OR medium-voltage OR mv))
Line / cable	(distribution OR medium-voltage OR mv) W/2 (line OR cable)
IT	"instrument transformer" OR "current transformer" OR "voltage transformer"

Figure 1 shows that the concepts related to "condition assessment" are merge with an "OR" operator, i.e., the terms are considered synonyms and any of them is useful to retrieve the bibliographic records. Moreover, "condition assessment" concepts and "distribution network assets" are associated with the operator "W/2", i.e., one of the terms can be found within a distance of two words from the other. For instance, if "condition assessment" (or "asset management", or "health assessment", or "fault diagnosis") appears in the tittle, abstract or keywords of a bibliographic record, and "power transformer" is found in the same record within a distance of two words, then the record is retrieved.

The number of records retrieved using the search rules on April 1st, 2023, are reported in Table 3. Results show extensive research related to strategic assets such as PT and PV systems, where several recently published literature reviews are found. However, limited research has been performed on condition assessment of other relevant assets such as DT and IT. In this context, this paper offers an overview of condition assessment for both widely and incipiently studied assets. It analyzes previously less explored aspects in detail to provide a comprehensive review of key technical considerations.

Table 3: Number of records retrieved with the proposed search rules.

Search rule	Jour. articles	Conf. papers	Reviews	Others	Total
PT	349	592	29	6	976
PV system	124	91	15	7	237
DT	12	37	2	0	51
Switchgear	26	53	3	0	82
Line / cable	42	68	2	1	113
IT	6	20	0	0	26

Figure 2 presents the evolution from 2003 to 2022 of the number of publications retrieved using each search rule. The figure shows an increasing trend across nearly all topics, with a particular emphasis on PV systems, which has seen a significant surge in publications over the last five years. A significant increase occurred in PT condition assessment from 2006 to 2008, and then the trend is slightly positive. Other moderately studied topics show an increasing interest such as lines and cables, switchgear, and DT. Finally, research is very incipient in IT condition assessment.

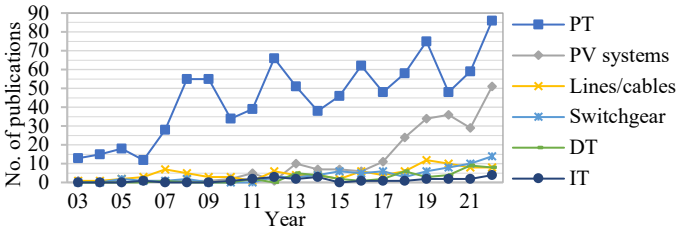


Figure 2: Publications related to condition assessment of MV assets over time.

A set of relevant publications are selected from the retrieved bibliographic records to perform the critical review of each asset. In the case of extensively studied topics, i.e., PT, and PV systems, the analysis is based on recently published reviews and a limited selection of additional relevant publications to observe aspects that required further analysis. The review of the remainder assets, i.e., DT, switchgear, lines and cables, and IT, is based on relevant publications selected by reading and assessing titles, abstracts, and conclusions of the retrieved bibliographic records. Moreover, references are complemented with related standards, technical reports, and white papers.

1.2. Taxonomy for condition assessment in distribution networks

The following structure is adopted to analyze each asset:

- A short description of the main asset's components, subsystems and features is given.
- Fault diagnosis and condition assessment approaches for each asset are described and the references are classified according to the type of test and diagnostic technique.
- Indices to monitor the condition of assets are analyzed.

The following subsections provide a broader description of the aspects covered in fault diagnosis and condition assessment of the MV assets, along with a taxonomy of test types and diagnostic techniques. Aspects related to indices are also introduced.

a) Fault diagnosis and condition assessment

Fault diagnostic and condition monitoring techniques for power assets is a wide and very dynamic research and innovative field. The variety of tests and techniques ranges from traditional and widely accepted methods, to advanced and not completely validated approaches. The main differences are based on the possibility that offers the technique to obtain a diagnosis with the asset in-service or off-service, i.e., the asset operating status. The diagnostic test nature differs according to the parameter to be analyzed, e.g., gas chromatography, which is a method from analytical chemistry commonly used to assess the insulating oil condition. In addition to chemical tests, mechanical or electrical tests can also be applied to analyze other oil features.

Once the tests or inspections have been carried out on the asset, the acquired data must be analyzed for diagnosis. This analysis can be based on expert judgment, who, like doctors in medicine, examine the data and offer a conclusion on the health status or presence of failures in the asset. This type of expert knowledge-based approaches also includes computational methods based on rules such as expert systems, and fuzzy logic.

Alternatively, analytical methods are applied by inferring a set of mathematical equations to model the phenomena involved in the asset condition as close to reality as possible. Thus, analytical models offer very accurate diagnostics, but require comprehensive understanding of the asset, in terms of specific components and operating modes, as well as the status of asset parameters.

Finally, the data acquired from tests can also be analyzed automatically with Artificial Intelligence (AI)-based techniques including Support Vector Machines (SVM), decision trees, and Artificial Neural Networks (ANN), among others. AI-based approaches provide flexible and efficient performance, which is of especial interest in large and complex systems with thousands of assets. However, these AI-based methods, also referred to as data-driven approaches, need large amounts of training and validation data, and are prone to underfitting and overfitting.

Figure 3 summarizes the process described above for fault diagnosis and condition assessment, i.e., 1) data from tests or inspections are acquired, and 2) diagnostic techniques are applied for data analysis. The figure also depicts the taxonomy of tests and diagnostic techniques discussed earlier. This taxonomy is used throughout the comprehensive review to classify tests and methods reported in the literature.

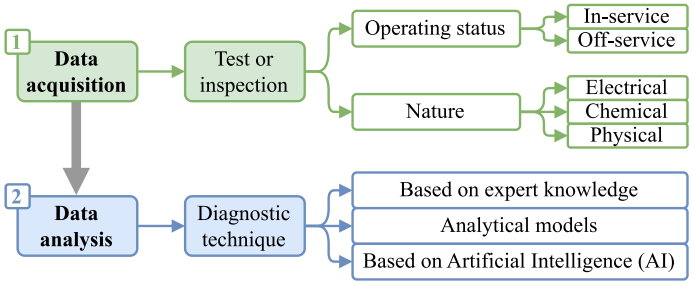


Figure 3: Taxonomy for fault diagnosis and condition assessment techniques.

b) Indices for condition assessment

In the international standard ISO 55000 [2], the main indices related to condition assessment of power assets are the Health Index (HI), consumed life and remaining life indices, performance indices, among others. Therefore, these indices are identified and described in each asset review.

2. Power Transformers (PT)

PT play a crucial role in power systems due to their strategic significance and higher costs compared to other power assets. In distribution networks, PT are usually referred to as primary DT, which connect MV networks to distributed generators, or, in interconnected systems, with High-Voltage (HV) transmission networks. Hence, the PT failure represents the impossibility of providing many customers with electrical energy. In brief, a PT comprises five subsystems. The main subsystem is the Active Part (AP) made up of windings, core, insulating paper, clamp structures, yoke, oil, and a tank containing the mentioned elements. The remainder four subsystems are (i) Bushings (BH) and terminals, (ii) On Load Tap Changer (OLTC), (iii) Conservator Tank (CTk) and breather, and (iv) the Cooling System (CS) made up of fans, pumps, controls, and radiators.

There is great interest in researching techniques for PT condition assessment. According to Table 3, a significant number of 976 bibliographic records were obtained using the search rule reported in Section 1.1. Among those publications, comprehensive and recently published literature reviews are found, e.g., [4–10]. Therefore, the following analysis summarizes the key findings of those reviews. Also, the review is complemented with standards [11–15], and selected papers [16–48].

2.1. Fault diagnosis and condition assessment of PT

Condition monitoring is an essential aspect of PT management. It involves the use of sensors and other monitoring devices to measure several parameters, such as temperature, oil level, and Partial Discharges (PD), which allows for detecting faults or changes in the asset performance. Advanced methods such as Dissolved Gas Analysis (DGA) are also employed to track the condition of transformer insulation systems. These methods allow early detection of faults and diagnosis of the PT status to take proper maintenance actions, thus reducing service interruptions and economic losses associated with its unavailability. Data sources, in both fault diagnosis and condition assessment, are acquired from the PT (primary source), power system, and environment. For this purpose, the main tests for monitoring PT parameters by subsystem and type are reported in Table 4.

Table 4: Classification of tests on PT and diagnostic techniques.

Data acquisition								Data analysis					Refs.
Test or inspection	Subsystem	Parameter	Type of test					Diagnostic technique	Purpose	Type of analysis			
			In-service	Off-service	Electrical	Chemical	Physical			Exp. know.	Analytical	AI	
DP	AP	Number of β-glucose monomers		✓		✓	✓	Viscometric method and Mark-Houwink-Sakurada equation	Direct measurement of DP value		✓		[12]
Furan analysis	AP	Furanic compound in oil (2-FAL, etc.)	✓			✓		Empirical relationships obtained in laboratory	Indirect estimation of DP value	✓			[4,9,20,39]
Load history	AP	Power (MVA)	✓		✓			Numerical models for hot-spot temperature calculation; analytical expression for pyrolysis, hydrolysis, and oxidation aging processes	Indirect estimation of DP value		✓	✓	[25,28,35]
Ambient temperature	AP, CS	Temperature (ºC)	✓				✓						
Winding/top oil temp.	AP	Temperature (ºC)	✓				✓						
Terminal voltages	AP, BH	Voltage (V)	✓		✓			Standards, guidelines, and manufacturer recommendations to assess the PT electrical performance	Detect deviations from normal, recommended, or standardized values	✓			[6,7,9,11,14]
Core to ground current	AP	Current (mA)	✓		✓								
Dissipation, power factor, capacitance	AP, BH, OLTC	Tan δ, power factor, capacitance (F)	✓	✓	✓								
Excitation current	AP	Current (A)	✓		✓								
Leakage current	AP, BH	Current (A)	✓										
Bushing voltage	BH	Voltage (V)	✓		✓								
Winding resistance	AP	Resistance (mΩ)		✓	✓								
Turns ratio	AP	Ratio		✓	✓								
Recovery voltage	AP	HV impulse (kV)		✓	✓								
Polarization index	AP	Insulation res. ratio		✓	✓								
Contact resistance	OLTC	Resistance (mΩ)		✓	✓								
DGA	AP, BH, OLTC	Gases dissolved in oil (ppm)	✓			✓		Dornenburg ratio; Key gas; Rogers ratio; trend analysis; Duval triangle; Duval pentagon; fuzzy logic	Early detection of electrical faults, insulation status assessment	✓		✓	[7,15,19,37,38]
FRA	AP	Admittance (S), impedance (Ω), transfer function	✓	✓	✓			Direct interpretation of the curve; comparison with previous measurements in the same or similar PT; analysis of harmonic components; wavelet; expert system; ANN	Detection of winding deformation and core movement	✓	✓	✓	[13,16,23,29]
DFR	AP	Transfer function (10 Hz to 1 MHz)		✓	✓			Time- and frequency-domain analysis; dielectric spectroscopy; principal component analysis; fitting models; ANN	Assessment of insulating materials frequency response	✓		✓	[17,21,27,32,36]
Oil characteristics	AP, OLTC	Moisture, acidity, color, breakdown voltage	✓			✓	✓	Direct interpretation and comparison of results from acidity, interfacial tension, dielectric, strength, and viscosity tests, Karl Fischer titration, etc., with standardized values	Evaluation of insulating oil status	✓		✓	[6,9,11,14,33]
Acoustic signals	AP, OLTC	Acoustic emission (dB)	✓				✓	Spectral analysis; time waveform analysis; pattern recognition; acoustic imaging	Health assessment of core and winding structures		✓	✓	[18,26,34,40]
PD	AP, BH	Electric charge (pC)		✓	✓			Analysis of direct PD measurements; acoustic PD measurement for source location; optical detection of the light emitted by PD	Determine the location, magnitude, and type of PD in insulation materials	✓	✓	✓	[22–24,31]
Ultra-High Frequency (UHF) PD detection	AP, BH	Electric charge (pC)		✓	✓								
Infrared thermography	AP, BH, OLTC, CTk, CS	Thermal Image (ºC)	✓				✓	The temperature variations observed in the infrared image are interpreted in the context of the PT operation and maintenance history	Identify temperature rise, loose connections, insulation failure	✓			[30,41]
Visual inspection	AP, OLTC, BH, CTk, CS	Oil levels, gas pressure, fan and pump condition, corrosion	✓				✓	The inspector records and classify observations of the components, including any signs of damage, wear, corrosion, etc. The severity of the damage is compared to standards	Identify observable anomalies	✓			[11,14]

From the tests in Table 4, it is worth noting that in-service condition monitoring allows continuous evaluation of the operation and offers the possibility of assessing the states that may impact the lifespan of the PT, whereas off-service tests require disconnecting the unit from the power grid, limiting its use to maintenance scenarios, planned inspections, or when there is a direct suspicion of failure.

Regardless the type of applied tests mentioned above, resulting data must be processed to generate information about the asset status. Consequently, diagnostic techniques are applied to assess the condition of PT and find any defects or malfunctions. For instance, these techniques include Frequency Response Analysis (FRA), which can detect winding deformation and core movement, or the analysis of tan δ, capacitance, and dielectric response measurements, among others, which can identify insulation degradation. In this regard, several diagnostic techniques have been proposed and the most relevant are also reported in Table 4 following their related tests.

Furthermore, analysis of data from traditional tests to diagnose the PT electrical performance, for instance, excitation current, power factor, and winding resistance, has rarely been addressed in recent research because of its straightforward interpretation and broad understanding and acceptance by PT managers. In contrast, most of the current research on PT condition assessment is focused on approaches or methodologies to improve the interpretation of the results, especially from FRA, Dielectric Frequency Response (DFR), DGA, acoustic signals, PD detection, and indirect Degree of Polymerization (DP) estimation. These recent approaches and

methodologies are aimed at overcoming the difficulties related to accuracy, uncertainty, interpretation, ambiguity, and discrepancies related to traditional and standard diagnosis.

For instance, in FRA and DFR, an option consists of comparing the magnitude and phase angle of the frequency response on a polar plot, which can then be analyzed by an expert. Alternatively, numerical models are used to simulate the behavior of transformers under different conditions. For instance, the effects of winding deformation, core movement, moisture content in oil and other defects on the transformer frequency response can be considered with these models. The simulated and measured frequency response are then compared to detect the presence of defects. A third alternative for analyzing the results of FRA and DFR is the use of machine learning techniques including decision trees, and ANN to identify patterns and predict the likelihood of defects or failures.

Similarly, DGA, which aims at identifying electrical faults, can be conducted with traditional analysis using Rogers ratio, key gas, Dornenburg ratio, Duval triangle and pentagon. Also based on expert knowledge, fuzzy logic allows the diagnosis of multiple failures, a more realistic representation, and quantification of the probability/severity of each damage. AI-based techniques are also applied to DGA interpretation. For instance, the use of ANN has the benefit of not requiring experts for the interpretation of DGA Approaches based on AI offer new possibilities in the application of DGA due to their automatic learning and data analysis capabilities. These approaches recognize different patterns of gas concentrations, which indicate the type of failure or problems within the transformer through the processing of large volumes of historical DGA data. These data are automatically identified and classified in real time or in certain time intervals, allowing for improvement and effectiveness in the early detection and diagnosis of failures. AI can recognize complex patterns and sudden deviations in gas concentration that are not easily noticed by experts. It is crucial to continue improving AI-based techniques to incorporate new research findings, such as the C_2O/CO ratio, for identifying failure patterns in the insulating paper.

The discussion above demonstrates that research on condition assessment of PT is a very important and fertile area of work, especially with the emergence of powerful AI-based tools to analyze large amounts of data that this type of asset generates.

2.2. Indices for condition assessment of PT

A crucial indicator in asset management of PT is the Health Index (HI) because it is required to perform fault diagnosis, life prediction and risk assessment. A PT-HI is a quantitative indicator that considers the results of asset performance observation, on-site inspections, and laboratory tests, to efficiently manage the asset. In 2017, a complete analysis of the main PT-HI calculated using mathematical equations/algorithms or expert judgment is reported in [10]. Such a paper analyzes the period 2008-2017, including the first approach published in 2008 [42]. Most of the expressions are based on the weighted summation of multiple input parameters, i.e., multi-criteria methods. In general, ranges of input parameters (collected through tests discussed above) are based on limits recommended by IEC and IEEE standards, and expert knowledge from utilities. The number of input parameters and weights varies from approach to approach. A complete list of the most used parameters is reported in [10].

Advanced techniques for PT-HI calculation are based on fuzzy logic [43–46], and ANN [47,48]. From a practical point of view, utilities prefer weighted summation approaches because of their simplicity and higher flexibility compared with advanced methods that require specialized knowledge to implement modifications.

3. Photovoltaic (PV) systems

PV systems are increasingly penetrating distribution networks as an alternative means of producing electricity motivated by the necessity of more sustainable energy sources than fossil fuels. PV systems exhibit rapid expansion within the realm of renewable energy technologies, primarily attributed to the enduring and readily accessible source of solar energy at no cost [49]. According to the International Energy Agency, the global PV capacity grew from 50 MW in 1990 to 946 GW in 2021 [49,50].

Depending on usage, PV systems can operate stand-alone used for remote power or backup applications, or grid-connected to complement the power provided to an installation from the utility grid [51]. Both stand-alone and grid-connected PV systems are mainly made up of (i) an array of solar PV panels which convert sunlight into electricity, (ii) inverters to transform DC into AC power, (iii) protection and switching equipment, (iv) supporting structure and mounting racks, (v) combiner boxes, conductors and grounding, (vi) energy meters and, optionally, (vii) an energy storage system and charge controller [49,51]. These components are exposed to deterioration and failures that make essential the use of condition assessment methods to guarantee sustainable techno-economic operation of PV systems. In the following sections, the aspects above are discussed based on the retrieved literature.

The results obtained with the search rule formulated in Section 1.1 for PV systems show that there is extensive research on the topic, i.e., 237 bibliographic records were retrieved including 15 literature reviews (see Table 3). Moreover, Figure 2 demonstrates an apparent increasing interest on the topic in the last few years. Therefore, emphasis is placed on the analysis of recent literature reviews [52–68]. In addition, the literature is complemented with standards [69–71], selected papers [72–76], and reports [77].

3.1. Fault diagnosis and condition assessment of PV systems

Condition assessment of PV systems is achieved by monitoring performance parameters of the components described above. For instance, these parameters include the operating voltage, current, and power output, which mainly depend on the proper operation of the array of solar PV panels, inverters, and the energy storage system. Integrity of combiner boxes, grounding, and insulation resistance must be also verified for condition assessment. These parameters are measured using instruments such as PV testers, solar power meters, insulation testers, battery analyzers, among other devices. In addition, testing procedures must be applied during installation, i.e., at commissioning, and regularly during the lifespan of the PV system. Standard and widely used and validated testing procedures for PV systems, as well as indication on suitable measurement instruments, are outlined by IEC 62446-1:2016 [69], and IEEE Std 1547-2018 [70], for grid-connected PV systems, and IEEE Std 1526-2020 [71], for stand-alone PV systems. Table 5 summarizes the tests applied on PV systems found in the literature,

Table 5: Classification of tests on PV systems and diagnostic techniques.

Data acquisition								Data analysis					Refs.
Test or inspection	Subsystem	Parameter	Type of test					Diagnostic technique	Purpose	Type of analysis			
			In-service	Off-service	Electrical	Chemical	Physical			Exp. know.	Analytical	AI	
Continuity	Conductors	Impedance (Ω)		✓	✓			Standards, guidelines, and manufacturer recommendations to assess electrical performance of PV systems	Detect deviations regarding normal, recommended, or standardized values	✓			[69–71]
Current (short, operating)	PV array	Current (A)	✓	✓	✓								
Open circuit voltage	PV array	Voltage (V)		✓	✓								
Voltage to ground	Grounding	Voltage (V)		✓	✓								
Insulation resistance	PV array	Impedance (Ω)	✓		✓								
Blocking diode	PV array	Voltage (V)		✓	✓								
Wet insulation resistance	PV array	Impedance (Ω)	✓		✓								
Shade evaluation	PV array	Various	✓	✓			✓	Standards, guidelines, and manufacturer recommendations to assess mechanical (physical) performance of PV systems	Verify initial and operating environmental (shading) and mechanical (connections, supports) conditions	✓			[54,69, 71]
Visual inspection	PV system	Various	✓	✓			✓						
Polarity	Comb. boxes, conductors	Voltage (V)		✓	✓		✓	Standards, guidelines, and manufacturer recommendations for electromechanical (electrical/physical) assessment of PV systems	Verify initial condition of conductors and boxes polarity, and proper operation of protections and inverters	✓			[69–71]
Functional	PV system	Various		✓	✓		✓						
Usable battery capacity	Battery	Capacity (Ah)	✓		✓			Standards, guidelines, and manufacturer recommendations to assess electrical performance of energy storage in stand-alone PV systems	Check the correct operation and requirements of the battery and charger controller	✓	✓		[71]
Recovery	Energy storage	Capacity (Ah)	✓		✓								
Battery charge	Energy storage	Autonomy (h)	✓		✓								
I-V curve	PV array	Voltage (V), current (A)	✓		✓			Handcrafted diagnosis; I-V curve tracer; real-time difference; PV and solar emulator	Identify faults or partial shading in PV arrays, and reduced efficiency in PV systems	✓			[53,57, 69,71,7 3,74]
Infrared thermography	PV array	Thermal image (°C)	✓				✓	Image processing; expert system; discriminant analysis; machine learning (SVM); deep learning (CNN, LSTM)	Detect hot spots and failures in PV panels	✓		✓	[56,62, 64,69]
Flash	PV array	Power (W)	✓		✓			Voltage, current, and power output; machine learning (SVM, GAN, transfer learning); deep learning (CNN)	Detect defects and reduced efficiency in PV panels	✓		✓	[59,60, 72]
Electroluminescence	PV array	Various	✓	✓			✓						
X-ray	PV array	Various	✓	✓			✓						
Model parameter identification	PV system	Various		✓	✓		✓	Analytical approach; machine learning (ANN); metaheuristics (particle swarm optimization, genetic algorithms, differential evolution)	Validate parameters for simulation; optimization of PV system performance; energy production forecasting		✓	✓	[55,65]
Physical fault (panel damage, defects)	PV array	Various	✓				✓	Signal processing (FFT, WT); statistics; model-based and real-time difference; machine learning (ANN, SVM, decision trees); deep learning (LSTM, CNN, DBN); fuzzy logic; handcrafted diagnosis; failures modes and effects; PV and solar emulator	Perform predictive maintenance of PV systems; early detection of faults; real-time detection and classification; monitoring; fault forecasting	✓	✓	✓	[52–54,56,5 8,61,63, 66,73,7 4]
Environmental fault (partial shading)	PV array	Irradiance (W/m^2)	✓				✓						
Electrical fault (line-line, line-ground, open circuit, blocking diode, arc)	PV array, inverters	Current (A)	✓		✓								

shows the corresponding diagnostic techniques, and indicates the PV subsystem involved in each test and the measured parameters.

Testing procedures described in standards include continuity verification of conductors and grounding, polarity check in combiner boxes and conductors, open circuit voltage, and short circuit and operational currents in the PV array, which are relevant parameters to monitor the correct operation of PV systems [69–71]. Tests are also conducted on the energy storage system such as usable battery capacity, recovery, and battery charge tests [71].

Additional testing procedures with demanding requirements for result analysis are described in the literature such as I-V curve-based approaches to study PV system performance from its I-V characteristic [53,57,73,74]. Other investigations use infrared thermography to detect hot spots in PV systems [56,62,64,69]. Moreover, flash, electroluminescence, and X-ray tests are employed to find defects in PV arrays [59,60,72]. It is worth noting that the I-V characteristic can be analyzed indirectly using thermography, flash, electroluminescence, and X-ray because there is a relation between hot spots and defects found with those tests and the voltage and current performance of the PV system. Tests and simulations are also performed for model parameter identification which is mainly intended for simulation, optimization of PV system performance, and energy production forecasting [55,65]. Moreover, modeling, simulation, and testing of faults have been widely studied. For instance, [52] presents an extensive overview of models to analyze DC faults in PV systems. The work in [53,63] extend the overview of electrical faults on the PV array, inverters, and the AC side of PV systems. In addition, [54,66] analyze not only electrical faults, but also physical, e.g., damage, cracks, and deterioration in PV panels, and environmental faults, e.g., temporary, and permanent partial shading. As an alternative to detailed modeling and simulation, PV and solar emulators are also available to perform a variety of tests [73,74]. This system type enables the emulation of faults within PV systems, as well as the evaluation of performance by means of analyzing the I-V curve.

After PV systems are tested, different diagnostic techniques are applied to analyze results. These techniques can be classified

according to the categories identified in Section 1.2 (see Figure 3), i.e., expert knowledge, analytical models, and based on AI. Table 5 also includes the classification of tests and diagnostic techniques according to the taxonomy defined in Figure 3.

A first set of approaches based on expert knowledge are standards, and manufacturer recommendations to assess electrical, mechanical, and electromechanical performance of PV systems [69–71]. In these methods, expert knowledge is required to apply tests using measurement instruments and provide a performance assessment based on the parameters established by standards. Furthermore, I-V curve tests are also analyzed through expert knowledge using handcrafted diagnosis [57], monitoring real-time difference with respect to a reference I-V curve [54,57], and using PV and solar emulators [73,74]. Also, expert systems used for interpretation of infrared thermography [62], and fuzzy logic for fault diagnosis in PV systems require expert knowledge [65].

Analytical models proposed for PV system analysis include, for instance, equations formulated to compute usable battery capacity, recovery, and battery charge in stand-alone applications [71]. Other examples include the analytical approach for model parameter identification [55,65], and the model-based difference for fault diagnosis [54], where monitored parameters of PV systems are compared with a modeled and theoretical reference to identify deviations from the correct operation.

Approaches based on AI techniques have been extensively used in the last few years. For instance, machine learning tools such as SVM, ANN, Generative Adversarial Networks (GAN), transfer learning, and decision trees are used for automatic analysis of images in infrared thermography [56,62,64,69], flash, electroluminescence, and X-ray [59,60,72], and in fault diagnosis applications [52–54,56,58,61,63,66,73,74]. Similarly, deep learning tools such as Deep Belief Networks (DBF), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN), have been also applied recently because of their extended capabilities for image analysis. All these AI techniques mentioned above usually require a previous step of feature extraction where statistics and signal processing techniques, e.g., Wavelet Transform (WT), and Fast Fourier Transform (FFT), are applied.

3.2. Indices for condition assessment of PV systems

In the literature, performance indices are offered for health and condition monitoring of PV systems. For instance, the world-wide association SolarPower Europe made up of hundreds of companies and policymakers from the solar sector, proposed best practice guidelines for asset management of PV projects [77]. Within the scope of these guidelines, Key Performance Indicators (KPI) are suggested to achieve technical asset management of PV power plants, e.g., specific and reference yield, performance ratio, energy performance index, among others. Moreover, equipment KPI are also indicated in the guidelines such as inverter specific energy losses, Mean Time Between Failures (MTBF), etc. Likewise, energy metrics are proposed in [76] to assess the behavior of PV systems. These metrics comprise daily and monthly energy output, array energy yields, PV array and system efficiency, capacity utilization factor, and life cycle conversion efficiency.

Furthermore, indices to analyze condition from the degradation perspective are also found in the literature. In this context, [67,68] offer a comprehensive review of degradation rates of PV systems around the world, and factors and parameters involved in the degradation process such as corrosion, encapsulant discoloration, and light and elevated temperature induced degradation. Similarly, [75] studies a health monitoring method based on the study of I-V curves and degradation modes of the PV array.

4. Distribution Transformers (DT)

DT are static, efficient, and reliable electric machines that, based on electromagnetic induction principle, supply electrical energy to end LV costumers. Distribution networks comprise thousands of DT, which are lower cost assets and more available into stock for replacement compared with PT. DT are also characterized by their rated powers that range between a few kVA to 2,5 MVA for three-phase transformers, or up to 833 kVA for single-phase transformers [78,79]. Moreover, according to the type of assembly, DT can be pad-mounted, i.e., on the ground, or pole-mounted, i.e., at the level of overhead lines.

Despite the differences, both PT and DT have many physical and constructive similarities, thus, DT consist of the following subsystems: (i) the AP including core, winding, tank, oil, and insulation paper (also referred to as insulation system), (ii) the BH, (iii) the CS, which can be dry or mineral oil immersed, (iv) the CTk, and (vi) OLTC. Moreover, the life span of their insulation systems is affected by the same factors, where DT useful life is usually around 25 years [80,81].

According to literature, over 60 % of in-service DT are aged [80]. Aged DT is a problem for electric utilities because the assets are more prone to failures and large-scale replacement of aged units is prohibitive due to high costs. Furthermore, the widespread incorporation of modern electrical loads and the unforeseen growth of energy demand result in unplanned overloading from the usual operation conditions [82], increasing probability of failures and operational costs [83].

4.1. Fault diagnosis and condition assessment of DT

According to the context described above, asset management and condition assessment methodologies are required for in-service DT to minimize failure risks, increase reliability, and coordinate maintenance actions. In this regard, Table 6 summarizes research related to maintenance and condition assessment approaches on DT that has been found in the literature.

In the current practice, condition assessment of DT is mainly based on preventive and corrective maintenance [84]. However, new DT maintenance trends include condition monitoring using sensors and measuring devices to detect incipient faults and continuously assess its operational performance. Currently, condition monitoring is not used on a large-scale by utilities on their in-service DT due to high implementation costs. However, technical benefits and improvement of DT efficiency using condition monitoring are demonstrated in the literature by several case studies on specific in-service units.

To achieve proper condition monitoring of DT, several parameters must be estimated with tests or inspections. For instance, chemical tests are used to measure moisture and oxygen content in the insulation system, dissolved gases, dielectric strength, power factor, and acidity, to estimate the remnant life of

Table 6: Classification of tests on DT and diagnostic techniques.

Data acquisition								Data analysis					Refs.
Test or inspection	Subsystem	Parameter	Type of test					Diagnostic technique	Purpose	Type of analysis			
			In-service	Off-service	Electrical	Chemical	Physical			Exp. know.	Analytical	AI	
Load History	AP	Power (kVA)	✓		✓			Analytical models from standards for hot spot temperature estimation; direct temperature measurements	Indirect estimation of DT loss of life; load balancing		✓	✓	[78–80,83,88–97]
Ambient temperature	AP, CS	Temperature (ºC)	✓				✓						
Winding/top oil temp.	AP	Temperature (ºC)	✓				✓						
Terminal voltages	AP, BH	Voltage (V)	✓		✓			Direct measurement and interpretation of DT parameters; standards and guidelines	Detect deviations from normal, recommended, or standardized values	✓		✓	[81,84,94,98,99]
Dissipation, power factor, capacitance	AP, BH, OLTC	Tan δ, power factor, capacitance (F)	✓	✓	✓								
Bushing inspection	BH	Voltage (V)	✓				✓						
Recovery voltage	AP	HV impulse (kV)		✓	✓								
DGA	AP, BH, OLTC	Gases in mineral oil (ppm), Gases in ester oil (ppm)	✓			✓		Dornenburg ratio; Key gas; Rogers ratio; data mining; Duval triangle; fuzzy logic; direct measurements	Early detection of electrical faults; insulation status assessment	✓		✓	[85,88,98–101]
FRA	AP	Impedance (Ω), transfer function	✓	✓	✓			Direct interpretation of the measured curve; comparison measurements in the same DT; expert system; ANN	Detection of winding displacement and core movement	✓	✓	✓	[85,87,102,103]
Oil characteristics	AP, OLTC	Moisture, acidity, color, breakdown voltage, dissipation factor (%), oil level	✓			✓	✓	kNN; direct measurement; direct interpretation and comparison of results from acidity, interfacial tension, and dielectric strength tests with standardized values	Assessment of insulating oil status	✓		✓	[92,96,98,99,104–106]
Acoustic signals	AP, OLTC	Acoustic emission (dB), freq. (kHz)	✓				✓	Spectral analysis; time waveform analysis; pattern recognition; acoustic imaging	Health assessment of core and winding structures		✓	✓	[85,94,96,107]
PD	AP, BH	Electric charge (pC)		✓	✓			Analysis of direct PD measurements; acoustic PD measurement to identify the source and location of PD activity; optical detection of the light emitted by PD	Determine the location, magnitude, and type of PD in insulation materials	✓	✓	✓	[85,92,93,98,99,101]
Infrared thermography	AP, BH, OLTC, CTk, CS	Thermal image (ºC)	✓				✓	The temperature variations observed in the infrared image are interpreted in the context of the DT operation	Identify temperature rise, loose connections, insulation failure	✓			[97]
Visual inspection	AP, OLTC, BH, CTk, CS	Oil level, gas pressure, fan and pump condition, corrosion	✓				✓	The inspector records and classify component observations, including any signs of damage, wear, corrosion, etc. The severity of the damage is compared to standards	Identify observable anomalies	✓			[84,94]

DT [85]. The most widely performed test for the analysis of dissolved gases is DGA, which is used for detection of incipient faults due to oil and solid insulating material degradation. DGA is conducted through in-service gas monitoring and considers the discharging rate and ratios of specific dissolved gas concentrations to determine fault types and severity in DT. To analyze results of DGA in DT, several diagnostic techniques are applied including Rogers ratio, key gas, Total Dissolved Combustible Gas (TDCG), Dornenburg ratio, and Duval triangle. In addition, AI-based tools are applied to DGA when more accurate fault diagnosis is required. Finally, DGA can be achieved using online gas monitoring, but it is expensive especially for fleets of extensive DT.

Further inspections in DT include physical and chemical tests of oil quality. These tests involve lower costs than online monitoring of gases and can be used to estimate health status of DT oil by moisture, oxygen, interfacial tension, acidity, dielectric strength, and power factor tests [86]. Results of the tests can be analyzed with expert knowledge or AI-based techniques.

Likewise, electrical tests are used to estimate deterioration of the AP. For instance, FRA is a non-invasive method in the frequency domain that detects mechanical and electrical damage such as winding deformations and displacements inside the DT. FRA is an effective test for DT monitoring, however, it can be affected by network conditions during online application [83]. Therefore, interpretation of FRA results still requires further research because of its complexity, including new computational methods such as AI-based diagnostic techniques, e.g., [87].

PD and short circuit impedance measurement are two additional electrical tests. PD measurement is a non-destructive testing procedure used to assess and monitor the condition of BH insulation and HV, LV and inter-turn winding insulation. PD can be measured off- or in-service [108]. Moreover, the short circuit impedance test is used to detect imperfect magnetic coupling, estimate short circuit parameters of DT, and comparing these values to factory test results.

The review in [85] divided miscellaneous diagnostic techniques applied to DT in Signal-Based (SB) and Data-Based (DB). SB approaches rely on processing of signals from mechanical vibration, sound, thermal or optical parameters, which is suitable for condition monitoring of in-service DT. Thus, SB methods can be classified into vibration analysis, optical fiber, acoustic emission, and thermography interpretation. For DB methods, data of DT is essential to develop statistics, standards, numerical and mathematical modeling. In general, DB approaches offer very reliable results with several sophisticated and validated models to infer DT aging.

AI methodologies integrate computational intelligence with sophisticated sensing, experience, learning, and training to enhance the precision of fault diagnosis and condition evaluation in DT. AI-based approaches applied to DT have been found in the

literature such as Bayesian networks, genetic algorithms, ANN, Adaptative Neuro Fuzzy Inference System (ANFIS), among others. The AI-based techniques mentioned above are used for HI calculation, FRA interpretation, hot spot detection, DGA and oil analysis, incipient fault detection, health condition estimation, insulation diagnosis, among other analyses applied to DT.

4.2. Indices for condition assessment of DT

The HI is an essential indicator to determine health status and a useful tool to perform proper management of DT in their life span. In this regard, different approaches to compute the DT-HI are found in literature, especially DB and AI-based techniques. DB approaches are developed using analytic methods based on specific measured data of DT parameters, usually recommended by IEEE and IEC standards, the Common Network Asset Indices Methodology (CNAIM) [109], or specific procedures proposed by researchers [94]. In such cases, the HI is calculated as the weighted sum of DT condition influencing factors, which are scored with weights according to importance [101]. Usually, estimation of weights is based on expert knowledge.

Because of high implementation costs of conventional DT condition assessment methods such as those based on DGA, FRA, and PD, novel advanced AI-based techniques have been developed. For instance, the DT-HI can be obtained using data mining [100], genetic algorithms [81], k-Nearest Neighbour (kNN) [104], Internet of Things (IoT) [105], and fuzzy logic [110].

5. Switchgear

MV switchgear are essential assets for power distribution utilities because they include protection, regulation, control, and measurement devices that allow safe operation of electrical networks in accordance with regulations. MV switchgear are also considered critical assets because their malfunction can put other assets at risk and affect the availability and quality of the electric power supply, causing financial losses to the utility. In extreme cases, this malfunction can affect safety of people, resulting in legal and financial sanctions and the deterioration of the utility's image. Thus, proper status assessment of MV switchgear is of utmost importance for early identification of possible failures and, consequently, for mitigating the negative impacts of failures on the utility. However, advances in condition monitoring of MV switchgear are incipient because of the relatively low cost of each of these assets compared with other critical assets [111], e.g., PT.

MV switchgear are available in various topologies primarily determined by their rated current (e.g., 630 A – 3150 A), and voltage (e.g., 7.2 kV – 36 kV). Moreover, MV switchgear can be classified based on their insulating medium in Air-Insulated Switchgear (AIS), Gas-Insulated Switchgear (GIS), or Oil-Insulated Switchgear (OIS). MV switchgear comprises subsystems such as dielectric, electronic, mechanical, and fluid components. Some of the sub-assets that make up an MV switchgear include (i) the primary circuit (e.g., bushings, busbars, cables), (ii) the interrupting chamber (e.g., breaker, isolating medium), (iii) mechanical linkages, (iv) the mechanism (e.g., spring, hydraulic, motor), and (v) control and auxiliary circuits (e.g., IT, protective relays, control, and monitoring systems) [112].

The breaker is widely regarded as the most critical sub-asset from both technical and economic perspectives because the protection of the entire system depends on its proper operation. Some studies into failure statistics of electrical components in MV networks have revealed that breakers represent the predominant components susceptible to failure in MV switchgear. Around 90% of these breaker issues are attributed to mechanical factors, impacting both the breaker motor and its operational mechanisms [113,114]. Consequently, this section mainly focuses on the diagnosis of the breaker, while also addressing some relevant aspects of the remainder components of the switchgear.

5.1. Fault diagnosis and condition assessment of switchgear

Under normal conditions and during the interruption of short-circuit currents, a breaker is subjected to mechanical, thermal, and dielectric stresses [115]. Over time, these efforts deteriorate the breaker's components, such as contacts, means of interruption, and operating mechanism. Deterioration can be accelerated by factors including manufacturing defects, frequency of operation, network development and evolution of the network failure rate [116].

According to [117], the main causes of breaker failures in MV networks in the United Kingdom are related to mechanical problems (30 %), partial discharges (26 %), and problems in the insulation (11 %). These statistics highlight the significance of diagnostic tests in assessing the breaker's health status.

In the report [118], the International Council on Large Electric Systems (CIGRE) provided a comprehensive review of the most widely accepted diagnostic methods for assessing the condition of MV switchgear. Some of these methods have reached a certain level of technological maturity and are therefore applicable for evaluating the health status of MV breakers. Table 7 presents a classification of some of these diagnostic tests.

Once diagnostic tests are applied, several diagnostic techniques can be applied for applications such as maintenance planning, and condition monitoring. Applying predictive maintenance in MV switchgear presents two main challenges [119]: determining the suitable sensors capable of consistently and resiliently measuring the critical physical parameters for MV switchgear management throughout its useful life; and the lack or small amount of data. For instance, continuous monitoring of temperature of MV switchgear is uncommon or nonexistent during their useful life and monitoring of the number of operations is carried out only few times per year, generally coinciding with periodic and scheduled maintenance.

This limited availability of data makes it difficult to implement AI algorithms. However, some studies have made certain advances in applying methods such as ANN for the diagnosis of MV switchgear, as reported in Table 7.

5.2. Indices for condition assessment of switchgear

a) Health Index (HI)

The HI of MV switchgear (HI_{SG}) is obtained by calculating the weighted sum of HI associated with each diagnostic test included in the model ($HI_{SG,n}$). A weight ($W_{SG,n}$) is assigned to each procedure based on its influence on the overall health status of the asset. The formula used for this weighted sum is expressed in (1).

$$HI_{SG} = \frac{\sum_{n=1}^{N} W_{SG,n} \cdot HI_{SG,n}}{\sum_{n=1}^{N} W_{SG,n}} \quad (1)$$

Table 7: Classification of tests on switchgear and diagnostic techniques.

Data acquisition								Data analysis					Refs.
Test or inspection	Subsystem	Parameter	Type of test					Diagnostic technique	Purpose	Type of analysis			
			In-service	Off-service	Electrical	Chemical	Physical			Exp. know.	Analytical	AI	
Dynamic and Static Resistance of Contacts	Breaker	Resistance (μΩ)		✓	✓			Gaussian process regression; ANN	Automatically discern pertinent predictor variables for a remaining useful life model of unmonitored, low-value power network assets, and contact temperature calculation	✓		✓	[121–124]
Timing of Operation	Breaker	ms / cycles	✓	✓			✓	Gaussian process regression	Automatically discern pertinent predictor variables for a remaining useful life model of unmonitored, low-value assets within power networks.	✓			[123,125,126]
PD (UHF, TEV, and ultrasound)	Primary circuit, breaker, IT	pC (directly / correlated)	✓		✓			Adaptive particle swarm optimization and ANN; extreme learning machine; automatic clustering and linguistic rules; CNN; scalogram representations and convolutional autoencoder	Partial discharge fault diagnosis and arcing fault detection	✓		✓	[111,124,127–134]
Insulation, dielectric integrity	Primary circuit, breaker, IT	Various		✓	✓			Optimal grade clustering method	Switchgear Insulation Condition Assessment			✓	[126,135]
Acoustic and vibrations (Accelerometers)	Mech. linkages, mechanism	Various	✓	✓			✓	Multilayer perceptron (ANN)	Fault diagnosis based on vibration signal			✓	[136–139]
Visual (as defined in IEC 62271-1)	Primary circuit, int. chamber; mech. linkages, mechanism, control and aux. circuits	Various	✓				✓	The inspector records and classify visual observations of the switchgear components, including any signs of damage, wear, corrosion, or other issues. The severity of the damage can be compared to established standards or guidelines for switchgear	Identify observable anomalies	✓			[122]
Temperature (IR cameras, Fiber guided sensors)	Primary circuit, int. chamber, mechanism	Temperature (°C, °F)	✓				✓	The temperature variations observed in the infrared image are interpreted in the context of the PT operation and maintenance history	Identify areas of abnormal temperature rise, overheating, loose connections, or insulation failure	✓			[140–142]
Monitoring number of switch operations	Breaker; Primary circuit	Various	✓				✓	Standards, guidelines, and manufacturer recommendations to assess the switchgear electrical performance	Detect deviations regarding normal, recommended, or standardized values	✓			[118,125,126]
Gas/Oil/Vacuum Quality (Mobile magnetron, Dielectric Vacuum Test)	Interrupting chamber	Various		✓		✓	✓						
Analysis of recordings from digital relays	Control and aux. circuits, int. chamber	Various	✓			✓							
Power factor	Prim. circuit, IT	Various	✓		✓								

Definition of weights $W_{SG,n}$ should consider limits and recommendations provided by manufacturers and knowledge of experts in the operation and maintenance of MV switchgear. For instance, indicators such as the number of mechanism operations and contact resistance are of great importance for estimating the health status of switchgear [116–118], so they should pose significant weight for computing HI_{SG}. However, the health status obtained through those tests ($HI_{SG,n}$) also depends on the intrinsic operating characteristics of the asset, defined by the manufacturer.

b) Remaining life

A methodology for computing the remaining life of MV switchgear based on failure mode analysis and performance analysis of the switchgear's components using existing data sources is presented in [120]. The approach suggests considering the components and subsystems of MV switchgear and estimating the overall HI as a weighted sum. The specific components and subsystems considered will depend on the type of switchgear, its technology, and its nominal values.

6. Overhead lines and underground cables

Lines and cables are highly susceptible to failures due to their long length. Given the great importance of lines and cables in the distribution network, extending their useful life, and making timely decisions to prevent failures and reduce maintenance and operational costs are crucial objectives. Lines and cables maintenance requires considering parameters such as the structural and environmental conditions, useful life, and operating environment. Despite overhead lines and underground cables have the same purpose, i.e., to transport electrical power, in practice, they are very different assets because of constructive aspects. Therefore, they are analyzed separately in the following.

6.1. Overhead lines

A MV overhead line is usually composed by (i) conductors, where the main types are Aluminum Conductor Steel Reinforced (ACSR), All Aluminum Alloy Conductor (AAAC), All Aluminum Conductor (AAC), Aluminum Conductor Composite

Core (ACCC) and Aluminum Conductor Steel Supported (ACSS), (ii) porcelain, composite, or glass insulators, (iii) fittings, mainly made of galvanized steel, and including suspension, tension, dead end, splice, and cross-arm fittings, (iv) poles made of wood, concrete, steel or, most recently, composite, i.e., a combination of fiberglass or carbon fiber reinforced polymer materials (PFRV), (v) cross-arms made of immunized wood, galvanized steel or PFRV, with rod or angle iron diagonals, (vi) sectioning and protection equipment including disconnectors, circuit breakers, and fuses, and (vii) guard lightning protection cables, communication cables and grounding systems.

To assess the MV overhead line condition, it is recommended to evaluate their components separately because the variety of materials mentioned above requires different assessment criteria. For instance, failure and aging processes in each material need to be known for each type of component.

6.2. Underground cables

Underground cables are in some cases preferred over overhead lines because of their lower visual impact and higher reliability. However, constructive requirements of underground cables make them more expensive than overhead lines. The main challenges that underground lines must overcome are to provide enough insulation so that the cables can be installed underground, and to dissipate the heat produced during operation.

The commonly used underground cable types are Cross-linked Polyethylene (XLPE), Self-Contained liquid Filled (SCFF), High Pressure, liquid Filled Pipe (HPFF), High pressure, tube Filled with Gas (HPGF), and Solid Cable. Solid cables demand lower maintenance, although monitoring and detecting insulation failures can be challenging. The size of XLPE cables increases with voltage. Typically, each circuit employs three distinct cables, which are enclosed within a buried or side-by-side conduit.

6.3. Methods for diagnosis and condition assessment of lines and cables

Tests and diagnoses are carried out to determine the MV lines and cables health with the objective of prioritizing decision-making in maintenance and to prevent service interruption due to damages. Among the main tests carried out on lines and cables are visual inspection, thermography, ultrasound, conductor analysis, grounding checking, structure verification, etc. [143–146]. Table 8 summarizes and classifies these main tests according to the taxonomy in Figure 3. Furthermore, Table 8 provides information on the corresponding diagnostic techniques, which range from traditional standard guidelines based on expert knowledge to advanced AI-based techniques such as decision trees.

6.4. Indices for condition assessment of lines and cables

Planning of maintenance, renewal or change of MV distribution lines and cables can be determined from asset management, using tools such as the HI calculation. There are some proposals that deal with this issue, for instance, [146] presents a method to compute an HI for overhead lines using visual inspection findings. An example is shown for foundations and poles, knowing that the aging of reinforced concrete is assessed by monitoring the gradual reduction in structural strength over time. Poles are evaluated within a condition range on a scale of 0 to 5, where a rating of 5 indicates a poor technical condition with an operational lifetime in its final stage, and a rating of 0 represents a state of good condition. Within this scale, the presence of longitudinal cracks with widths ranging from 0.3 to 0.6 mm corresponds to a score of 4. A table with the proposed score for different state conditions is detailed in [146].

Finally, the general HI, considering all the components of the overhead line, is computed using the weighted summation in (2).

$$HI_L = \sum_{i=1}^{M} HI_{L,i} \cdot W_{L,i} \tag{2}$$

where $HI_{L,i}$ is the index for each line component i, and $W_{L,i}$ its weighting factor, which measures the component importance and is usually defined by utility experts. It is advised to use an exponential scale to convert $HI_{L,i}$ values in (2) to achieve a suitable comparison for extreme values [146].

7. Instrument Transformers (IT)

The signals acquired by IT and subsequently processed and stored by measurement devices, are decisive state variables to control and protect power systems. Two main subsets of IT are Voltage Transformers (VT) and Current Transformers (CT). A VT produces a secondary voltage that is directly proportional to the primary voltage, while a CT generates a secondary current that is proportionate to the primary current. Thereby, VT and CT offer a high-fidelity measurement of voltage and current, respectively, while providing isolation from high voltage and current levels.

On the one hand, there is a variety of VT such as capacitive, resistive-capacitive, and inductive [147]. Inductive VT, for instance, are widely used for measurement and protection in MV networks. On the other hand, there are three types of CT including electromagnetic, capacitive, and optical. Optical CT based on the magneto-optic Faraday effect, are used for measurement and smart grid protection. They have high immunity to electromagnetic disturbances and are suitable for control and protection systems featuring an extensive measurement range [148].

For any type of IT, reliable measurements are very important. Input signals for protection relays are channeled via IT to execute the protective functions. It is important to highlight that network reliability may be compromised in the event of IT failure, particularly if the input signals for protection and control prove to be imprecise. In this context, it is relevant to identify tests and assessment approaches for IT management.

7.1. Fault diagnosis and condition assessment of IT

The tests to diagnose IT can be carried out with the asset in-service or off-service, as specified by the corresponding standards. These tests allow identifying anomalies or failures and provide some indication of the expected reliability and loss of life of the asset. It should be noted that a single electrical test may not be enough to diagnose the asset, but several tests may be necessary to identify a problem. In addition, the acceptance criteria are specific for each manufacturer.

There are three types of tests including routine tests, carried out on a sample of the equipment, e.g., short circuit withstand tests in CT and VT, and overvoltage test in CT; type tests, applied to one

Table 8: Classification of tests on underground cables and overhead lines and diagnostic techniques.

Data acquisition								Data analysis					Refs.
Test or inspection	Subsystem	Parameter	Type of test					Diagnostic technique	Purpose	Type of analysis			
			In-service	Off-service	Electrical	Chemical	Physical			Exp. know.	Analytical	AI	
Visual inspection	Conductors, insulators, fittings, cross-arms, poles	Corrosion, physical damage, loose strands, vegetation growth, etc.	✓	✓			✓	Check list; HI calculation based on predefined assessment criteria for each type of MV line component	Detect and document anomalies for further treatment; estimate the overall HI to identify critical lines; detect defects for MV line technical condition assessment; supplement onsite measurements, laboratory measurements or full-scale tests	✓		✓	[146]
Dielectric spectroscopy (over a low frequency range, i.e., few hertz)	Underground cables	Relative permittivity, dissipation factor, power factor, capacitance		✓	✓			Comparisons of measured relative permittivity with values of reference	Estimation of the breakdown voltage reduction; estimation of stage of water tree ageing (i.e., formation of microscopic tree-like structures within the insulation due to the presence of water or moisture over an extended period)	✓			[149]
PD	Underground cables	Electric charge (pC)		✓	✓			Decision tree theory	Time-domain and frequency-domain measurement and analysis of PD data			✓	[150,151]
Corona effect	Underground cables	Various	✓				✓	Method of noise reduction in the UV channel	Estimation of fault severity	✓	✓		[145]
Tests on structures	Structure	Various	✓		✓		✓	Post breaking strength	Determination of weight tolerance to support	✓			[146]
Thermography	Und. cables, poles, insulators, structure	Temperature (°C)	✓				✓	Standards, guidelines, and manufacturer recommendations to assess the lines and cables electrical performance	Detect deviations regarding normal, recommended, or standardized values	✓			[143]
Cable mechanical	Underground cables	Various		✓			✓						[152–154]
Cantilever, load (poles)	Underground cables	Various	✓	✓			✓						[152,155, 156]
Grounding	Structure	Resistance (Ω)	✓		✓								[144]
Electromechanical and mechanical failing load	Underground cables, insulator, structure	Various		✓			✓						[157,158]
Puncture voltage	Insulator	Voltage (kV)		✓	✓		✓						[157,158]
Ultrasound	Und. cables, poles, insulators, structure	Various	✓				✓						[143]
Tracking effect	Underground cables	Various	✓										[143]
Dry lightning impulse withstand voltage	Underground cables, insulator, structure	Various		✓	✓								[157,158]
Flashover	Underground cables, insulator, structure	Various		✓	✓								[157,158]
Wet power-frequency withstand voltage	Underground cables, insulator, structure	Various		✓	✓								[157,158]

or two identical or similar types of equipment to verify a specific characteristic, e.g., heating test, nominal and safety factor; and maintenance tests, carried out on equipment in-service, to verify its health status after a certain period of operation. A summary of diagnostic tests for IT based on the review is reported in Table 9.

Table 9 also reports diagnostic techniques applied on the results of tests for condition assessment of IT. It is observed that research on condition assessment of IT is still incipient (see also in Table 3 that only 26 records were retrieved with the formulated search rule). Diagnostic techniques found in the literature use, for instance, statistical models such as Weibull, or AI-based techniques including ANN and SVM, as shown in Table 9.

7.2. Indices for condition assessment of IT

Inadequate maintenance practices, environmental factors, and the transformer's current condition can lead to failures that have an adverse impact on the transformer's expected lifespan. In this context, the adoption of a predictive approach for asset management and future condition forecasting would allow avoiding failures through early detection of technical problems that may affect the transformer [159]. Research on indices for condition assessment of IT is also very incipient, however, some indicators have been formulated.

a) Loss of life

Presently, endeavors are in progress to establish guidelines for estimating the reduction in lifespan of a CT due to short-term overloading. In a related study [160], a thermal accelerated aging method is introduced and implemented on a 69 kV oil-immersed CT subjected to overloading. The laboratory-acquired life test outcomes are validated against analytical assessments involving DP and the analysis of furanic compounds in the transformer oil. The standard IEEE C57.91 is used as a reference to formulate the aging acceleration factor (F_{AA}) as expressed in (3).

$$F_{AA} = \exp\left(\frac{15000}{368} - \frac{15000}{\theta_H + 273}\right) \quad (3)$$

Likewise, the percentage of loss of life (*LOL*) is calculated resulting from the equivalent acceleration factor (F_{EQA}), the normal life of the insulation (DP = 200) and the duration of the stress (t), according to (4)-(5).

Table 9: Classification of tests on IT and diagnostic techniques.

Data acquisition								Data analysis					Refs.
Test or inspection	Subsystem	Parameter	Type of test					Diagnostic technique	Purpose	Type of analysis			
			In-service	Off-service	Electrical	Chemical	Physical			Exp. know.	Analytical	AI	
Power-frequency withstand tests on windings and between sections	CT, VT	Voltage (kV)		✓	✓		✓	Standards, guidelines, and manufacturer recommendations to assess the IT electrical performance	Detect deviations regarding normal, recommended, or standardized values	✓			[161–163]
Measurement of the radio interference voltage (RIV)	CT, VT	Voltage (mV)		✓	✓								
Mechanical test (Use loads due to wind and ice)	CT, VT	Incline degrees vs. gravity units		✓			✓						
Determination of errors (Accuracy class, current error, phase displacement)	CT, VT	Phasor degrees, error (%)	✓		✓								
Verification of terminal marking and polarity	CT, VT	Various		✓			✓						
Measurement of capacitance and dielectric dissipation factor	CT, VT	Tan δ, power factor, capacitance (F)		✓	✓								
Resistance measurements	CT, VT	Resistance (mΩ)		✓	✓								
Thermography (Infrared inspection)	CT, VT	Thermal Image (ºC)	✓				✓	ANN from the Least-Mean-Square (LMS)	Predicting the max. temperature in the low and high voltage winding			✓	[164]
Visual inspection	CT, VT	Various	✓	✓			✓	The inspector records and classify visual observations of the IT components, including any signs of damage, wear, corrosion, or other issues.	Identify observable anomalies	✓			[165], [164]
DGA and liquid content	CT, VT	Gases dissolved in oil (ppm)		✓		✓		Multi-stage classifier based on SVM	Direct measurement of DGA value		✓		[166–168]
Partial discharge measurement	CT, VT	Electric charge (pC)		✓	✓		✓	Severity Index (SI)	Indirect estimation of DP value		✓		[161–163,169]
Frequency Domain Dielectric Spectroscopy (FDS)	CT	Complex capacitances and Tan δ		✓	✓			Harverlik-Negami dielectric response function	The model was used to fit FDS	✓	✓		[170]
Flashover and inclined plane tracking test	CT, VT	Voltage (kV)						The Weibull distribution plot	The predicting the flashover and degradation initiation		✓		[171]
FRA	CT, VT	Admittance (S)/ impedance (Ω)	✓	✓	✓			Total Vector Error (TVE)	Evaluated with different input voltages to identify failures		✓		[172]
Synchrophasor measurement	CT, VT	Phasor measurement unit (PMU)	✓				✓	3Φ-SLSE (Three-Phase Current and voltage State Estimator + The Largest Normalized Residual (LNR))	Model based real-time application to identify the presence of incorrect measurements	✓		✓	[173,174]

$$F_{EQA} = \frac{\sum_{n=1}^{N} F_{AA,n} \cdot \Delta t_n}{\sum_{n=1}^{N} \Delta t_n} \quad (4)$$

$$LOL = \frac{F_{EQA} \cdot t}{normal\ insulation\ life} \times 100\% \quad (5)$$

b) Lifetime

Significant thermal aging of paper insulation stands out as a pivotal constraint on the operational lifespan of IT. In [175], research encompasses simulation studies and a comprehensive review of literature pertaining to insulation systems, moisture dynamics, and statistical modeling of IT lifetime data. The primary objective of this research was to identify potential triggers and factors that impact IT failures. Results showed the relationship between failures and age, environmental factors, and operating stress level.

c) Health Index (HI)

The HI serves as a metric for quantifying the enduring deterioration of a transformer. It combines intricate condition-related information into a singular value, offering a relative assessment of the overall state of a transformer. Most HI methodologies rely on either laboratory or on-site test data, with limited incorporation of actual operational time [159,176].

For instance, the calculation of a CT HI is formulated in [159], where oil tests (furans, acetylene, etc.) are considered, as well as thermal aging factors and other parameters including maintenance, environment, location, failure history, and observed external and internal condition. A life expectancy model of the upgraded CT is developed based on the calculated HI and the HI aging factor. The calculation is given by (6).

$$HI_{IT} = HI_{IT-0} \cdot e^{B \cdot (T_2 - T_1)} \quad (6)$$

where HI_{IT-0} is the initial HI considered as 0.05, B is the aging factor, T_1 is the year corresponding to HI_0 (T_{1-0} = expected life) and T_2 is the year under study ($T_{2-0} = 0$).

8. Discussion

A first insight offered by the quantitative review developed in this paper is that only research on condition assessment of PT shows high level of maturity in all the analyzed aspects including tests, diagnostic techniques, and indices. Moreover, condition assessment on PV systems is a topic of increasing interest because of the current relevance of renewable energy resources. Therefore, significant efforts have been put in the topic during the last five years, however, further work should be carried out on indices for condition assessment of PV systems. Regarding condition assessment of DT, switchgear, lines and cables, and IT, research is mature in terms of tests, but still incipient on diagnostic techniques and indices.

Different types of tests, mostly based on widely accepted practices and standards, are applied on the selected MV assets, and a taxonomy was formulated in this paper accordingly. In this sense, according to the operating status of the asset, tests are classified as in-service and off-service. Results of the review summarized in Tables 4-9 demonstrate that a similar amount of in-service, and off-service tests are available for the MV assets dealt within this review. Moreover, according to nature, most of those tests are electrical (as expected because of the electrical nature of the assets), e.g., short-circuit withstand tests, voltage and current measurements, FRA (in PT and DT), followed by physical tests, e.g., visual inspection (in all the analyzed assets), mechanical tests (of particular importance for switchgear), thermography (in all the analyzed assets), and chemical tests (mostly applied to PT and DT), e.g., DP, furan analysis, DGA. No chemical tests were identified in the literature for PV systems, lines, and cables.

Likewise, categories were defined in this review for diagnostic techniques, namely, expert knowledge, analytical models, and AI-based approaches. These diagnostic techniques aim at improving interpretability of test results. From Tables 4-9 and the extensive review, it is seen that most diagnostic techniques are based on expert knowledge (established in standards and widely accepted procedures and industry practice), whereas very few approaches use analytical models probably because of the complex derivation and implementation of physical models that hinders the applicability. However, analytical models are still valuable diagnostic techniques, e.g., in the analysis of PD and corona effect in lines and cables, and the analysis of FRA and acoustic signals in PT and DT.

Furthermore, AI-based diagnostic techniques have been introduced recently to overcome some limitations of conventional methods. For instance, AI models are useful to deal with large volumes of data and to represent complex phenomena with minimal requirement of knowledge and physical interpretation (black-box approaches). Examples of AI-based diagnostic techniques are ANN for multiple applications such as interpretation of FRA and DFR in PT and DT, model parameters identification in PV systems, and PD and acoustic analysis in switchgear, lines, and cables. Other AI-based diagnostic techniques have been applied including decision trees, SVM, fuzzy logic, optimization methods, expert systems, kNN, and, to a lesser extent, deep learning models such as CNN, DBF, and LSTM.

9. Prospects for future research

The results of the review demonstrate that research on condition monitoring of some critical MV assets is incipient, especially regarding indices for asset management. Therefore, the need for further research in this topic is highlighted. For instance, HI for PV systems, DT, switchgear, lines, cables, and IT, should be further investigated to obtain a proper health characterization of each asset and its subsystems according to the application, e.g., investment decision-making, network planning, and maintenance scheduling. In this context, straightforward computation of HI based on weighted sums should be explored given its applicability to real-world scenarios, and extensive study cases should be analyzed to demonstrate the benefits of condition monitoring of MV assets for cost-effective operation of distribution networks.

Root cause failure analysis, forecasting of health status, and incipient fault detection based on condition monitoring are also promising areas of research. Applications of these research lines include failure mitigation measures, and predictive maintenance to improve useful life of the assets. Several techniques have demonstrated satisfactory results and capabilities in this direction. For instance, root cause analysis using supervised AI tools such as ANN, SVM, and decision trees has been performed. However, these approaches need large volumes of labelled training data and suffer from overfitting. Moreover, Markov chains have been used for health status forecasting and informed decision-making in asset management. Nonetheless, forecasting capabilities of Markov chains are limited because of the lack of long-term dependencies or representation of trends on data. Other promising approaches for health status forecasting and incipient fault detection include supervised and unsupervised deep learning techniques such as LSTM and deep autoencoders. In this case, drawbacks are on the requirement of large datasets for training and the limited interpretability of models.

As mentioned above, the use of AI techniques for advanced applications in condition monitoring require large datasets. In this regard, new technologies are increasingly adopted by network operators for widespread acquisition of offline and real-time data, e.g., IoT, and advanced metering infrastructures. To properly exploit this huge amount of data, big data frameworks and cloud-based applications with high performance capabilities should be studied in detail, and potentialities should be envisaged.

Finally, lack of integrated tools for asset management and condition monitoring considering most critical assets of the MV network, e.g., PT, PV systems, DT, switchgear, lines, cables, and IT, is observed. Consequently, the analysis, implementation, and research on case studies with the application of comprehensive asset management and condition monitoring methodologies on distribution networks considering all the critical assets is highly encouraged to take advantage of current technology infrastructures and high availability of data to improve the efficient operation of modern power networks. In this context, the use of optimization methods for the integrated operation of distribution networks based on asset management and condition monitoring is seen as a prolific prospective area of research.

10. Conclusion

This paper extends the work presented in ARGENCON 2022 [1]. This extended review examines the literature on critical MV assets operating in distribution networks, namely, PT, PV systems, DT, switchgear, lines and cables, and IT. Moreover, this work is developed in the context of a government-funded research project in Colombia aimed at implementing a digital platform for asset management of distribution network components at the archipelago of San Andrés, Providencia, and Santa Catalina. In this context, the review offers a detailed study of the main technical aspects of each asset and its subsystems, as well as the main parameters for condition assessment. The contributions of this paper are summarized as follows:

- A systematic methodology is adopted for the review and bibliographic data retrieval, where advanced search rules are formulated, and the obtained academic publications are complemented with diverse literature including standards, reports, and relevant white papers.
- A quantitative analysis of the status of research on condition assessment of MV assets is provided, highlighting maturity on the study of PT, the increasing interest in PV systems, and the lack of research on DT, switchgear, lines and cables, and IT.
- A taxonomy for tests and diagnostic techniques used for condition monitoring of MV assets is defined and the extensive literature is classified accordingly (see Tables 4-9).
- Based on the classification above, the main technical aspects related to methods for diagnosis and condition assessment of each asset and indices related to condition assessment are developed.

Regarding the main findings and research gaps, the review identifies the types of tests and diagnostic techniques applied to MV assets, with electrical tests being the most common. Analytical models and AI-based approaches show promise in improving the capability to analyze test results. The prospects for future research lie in developing comprehensive HI for various MV assets, root cause failure analysis, forecasting health status, and incipient fault detection using advanced AI techniques such as deep learning. Leveraging big data and cloud-based applications are suggested to exploit the growing amount of data from emerging technologies, and the integration of condition monitoring and asset management methodologies is crucial for enhancing the efficient operation of modern power distribution networks.

List of symbols

B	Aging factor
F_{AA}	Aging acceleration factor
F_{EQA}	Equivalent acceleration factor
HI_{IT}	Health index of IT
$HI_{IT\text{-}0}$	Initial health index of IT
HI_{L}	Health index of line
$HI_{L,i}$	Health index of line component i
HI_{SG}	Health index of switchgear
$HI_{SG,n}$	Health index obtained in test n on switchgear
LOL	Percentage of loss of life
t	Duration of stress
T_1	Year corresponding to HI_0
$T_{1\text{-}0}$	Expected life
T_2	Year under study
$W_{L,i}$	Weighting factor of line component i
$W_{SG,n}$	Weight for test n on switchgear
θ_H	Hot Spot Temperature

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

Financial support by Fondo de Ciencia, Tecnología e Innovación (FCTeI) of Sistema General de Regalías (SGR) under project BPIN 2021000100223 entitled “Implementation of a digital platform to increase the use of the electrical network assets of the Archipelago of San Andrés and Providencia” is greatly acknowledged.

References

[1] C. Madrid-Chirinos, A.A. Romero-Quete, L. Ontiveros, A. Sarasua, “Condition Assessment of High Voltage Instrument Transformers: A Bibliometric Analysis and Literature Review,” in 2022 IEEE Biennial Congress of Argentina (ARGENCON), IEEE: 1–8, 2022, doi:10.1109/ARGENCON55245.2022.9939980.

[2] ISO-55000, Asset Management - Overview, principles and terminology, Switzerland, 2014.

[3] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, D. Altman, G. Antes, D. Atkins, V. Barbour, N. Barrowman, J.A. Berlin, J. Clark, M. Clarke, D. Cook, R. D’Amico, J.J. Deeks, P.J. Devereaux, K. Dickersin, M. Egger, E. Ernst, P.C. Gøtzsche, J. Grimshaw, G. Guyatt, J. Higgins, J.P.A. Ioannidis, J. Kleijnen, T. Lang, N. Magrini, D. McNamee, L. Moja, et al., Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement, PLoS Medicine, 2009, doi:10.1371/journal.pmed.1000097.

[4] R.D. Stebbins, D.S. Myers, A.B. Shkolnik, “Furanic compounds in dielectric liquid samples: Review and update of diagnostic interpretation and estimation of insulation ageing,” in Proceedings of the IEEE International Conference on Properties and Applications of Dielectric Materials, 921–926, 2003.

[5] H. de Faria, J.G.S. Costa, J.L.M. Olivas, “A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis,” Renewable and Sustainable Energy Reviews, **46**, 201–209, 2015, doi:10.1016/j.rser.2015.02.052.

[6] C. AJ, M.A. Salam, Q.M. Rahman, F. Wen, S.P. Ang, W. Voon, “Causes of transformer failures and diagnostic methods – A review,” Renewable and Sustainable Energy Reviews, **82**, 1442–1456, 2018, doi:10.1016/j.rser.2017.05.165.

[7] R. Soni, B. Mehta, “Review on asset management of power transformer by diagnosing incipient faults and faults identification using various testing methodologies,” Engineering Failure Analysis, **128**, 105634, 2021, doi:10.1016/j.engfailanal.2021.105634.

[8] A.R. Abbasi, “Fault detection and diagnosis in power transformers: a comprehensive review and classification of publications and methods,” Electric Power Systems Research, **209**, 107990, 2022, doi:10.1016/j.epsr.2022.107990.

[9] L. Jin, D. Kim, A. Abu-Siada, S. Kumar, “Oil-Immersed Power Transformer Condition Monitoring Methodologies: A Review,” Energies, **15**(9), 3379, 2022, doi:10.3390/en15093379.

[10] A. Azmi, J. Jasni, N. Azis, M.Z.A.A. Kadir, “Evolution of transformer health index in the form of mathematical equation,” Renewable and

Sustainable Energy Reviews, **76**, 687–700, 2017, doi:10.1016/j.rser.2017.03.094.
[11] IEEE Std 62-1995, IEEE Guide for Diagnostic Field Testing of Electric Power Apparatus - Part 1: Oil Filled Power Transformers, Regulators, and Reactors, 1–64, 1995.
[12] IEC-60450:2004, Measurement of the Average Viscometric Degree of Polymerization of New and Aged Cellulosic Electrically Insulating Materials, 1–45, 2004.
[13] IEEE PC57.149/D9.2, IEEE Draft Guide for the Application and Interpretation of Frequency Response Analysis for Oil Immersed Transformers, 1–68, 2012.
[14] IEEE Std C57.152-2013, IEEE Guide for Diagnostic Field Testing of Fluid-Filled Power Transformers, Regulators, and Reactors, 1–121, 2013.
[15] IEEE PC57.104/D6.2, IEEE Approved Draft Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers, 1–103, 2019.
[16] S. Birlasekaran, F. Fetherston, "Off/on-line FRA condition monitoring technique for power transformer," IEEE Power Engineering Review, **19**(8), 54–56, 1999, doi:10.1109/39.780991.
[17] B. Noirhomme, E. David, H. Garbi, M.C. Lessard, R. Boissonneault, "Application of dielectric response techniques for the condition assessment of power transformers," in CEIDP '05. 2005 Annual Report Conference on Electrical Insulation and Dielectric Phenomena, 2005., IEEE: 273–276, 2005, doi:10.1109/CEIDP.2005.1560674.
[18] F.A. Wegelin, R.S. Magalhães, L.A.L. De Almeida, M. Fontana, "Condition monitoring of power transformers using acoustic signal Prony's analysis," in Conference Record - IEEE Instrumentation and Measurement Technology Conference, 1384–1387, 2005, doi:10.1109/IMTC.2004.1351324.
[19] M. Duval, "Dissolved Gas Analysis and the Duval Triangle," 1–20, 2006.
[20] Z. Abdul-Malek, N. Bashir, H. Ismail, "Furan analysis on power transformers in Malaysia: A field investigation," International Review on Modelling and Simulations, **4**(5), 2227–2233, 2011.
[21] G. Csépes, B. Németh, Cs. Vörös, "Practicable expert system for the improved interpretation of dielectric response diagnostic methods of power transformers," in 44th International Conference on Large High Voltage Electric Systems 2012, 2012.
[22] M.J. Liu, J.Z. Zou, Q.K. Zhou, Y.L. Li, N.P. Yan, "On-Line Condition Monitoring for Power Transformers Based on UHF PD Measurements," Advanced Materials Research, **614–615**, 1158–1162, 2012, doi:10.4028/www.scientific.net/AMR.614-615.1158.
[23] S. Tenbohlen, M. Heindl, M. Jovalekic, "On-site PD Diagnostics, FRA and Moisture Measurement for Power Transformers," IEEJ Transactions on Power and Energy, **132**(2), 134–138, 2012, doi:10.1541/ieejpes.132.134.
[24] H. Gago Garcia, F. Garnacho Vecino, M.A. Sánchez-Uran, J. Ortego La Moneda, I. Uliarte Ranea, "Condition assessment of power transformers in service using pd monitoring," in CIGRE Session 46, 2016.
[25] A.A. Romero-Quete, E.E. Mombello, G. Rattá, "Assessing the loss-of-insulation life of power transformers by estimating their historical loads and ambient temperature profiles using ANNs and Monte Carlo simulations," DYNA (Colombia), **83**(197), 104–113, 2016, doi:10.15446/dyna.v83n197.48134.
[26] A. Hekmati, R. Hekmati, "Optimum acoustic sensor placement for partial discharge allocation in transformers," IET Science, Measurement & Technology, **11**(5), 581–589, 2017, doi:10.1049/iet-smt.2016.0417.
[27] J. Liu, H. Zheng, Y. Zhang, H. Wei, R. Liao, "Grey Relational Analysis for Insulation Condition Assessment of Power Transformers Based Upon Conventional Dielectric Response Measurement," Energies, **10**(10), 1526, 2017, doi:10.3390/en10101526.
[28] R.D. Medina, A.A. Romero, E.E. Mombello, G. Rattá, "Assessing degradation of power transformer solid insulation considering thermal stress and moisture variation," Electric Power Systems Research, **151**, 1–11, 2017, doi:10.1016/j.epsr.2017.04.006.
[29] N. Roopnarine, A. Singh, C.J. Ramlal, S. Rocke, "Network Synthesis of Transformer Winding for FRA Interpretation," in 2016 8th International Conference on Computational Intelligence and Communication Networks (CICN), IEEE: 493–497, 2016, doi:10.1109/CICN.2016.102.
[30] N. Yao, X. Wu, "Transformer fault detection based on infrared power image," Acta Technica CSAV (Ceskoslovensk Akademie Ved), **62**(2), 237–243, 2017.
[31] S. Seifi, A. Akbari, P. Werle, H. Mohseni, A.A.S. Akmal, "A sensitivity analysis of RLC transformer winding model parameters used for a PD localization," in 2018 IEEE 2nd International Conference on Dielectrics (ICD), IEEE: 1–4, 2018, doi:10.1109/ICD.2018.8468521.
[32] M. Koch, M. Anglhuber, "Experiences with Measurement and Analysis of the Dielectric Response of Instrument Transformers," in 2018 IEEE International Conference on High Voltage Engineering and Application (ICHVE), IEEE: 1–4, 2018, doi:10.1109/ICHVE.2018.8642206.
[33] E.T. Mharakurwa, R. Goboza, "Multiparameter-Based Fuzzy Logic Health Index Assessment for Oil-Immersed Power Transformers," Advances in Fuzzy Systems, **2019**, 1–12, 2019, doi:10.1155/2019/2647157.
[34] Y.O. Shaker, "Detection of partial discharge acoustic emission in power transformer," International Journal of Electrical and Computer Engineering (IJECE), **9**(6), 4573, 2019, doi:10.11591/ijece.v9i6.pp4573-4579.
[35] A.F. Cerón, R.A. Lozano, G. Aponte, A.A. Romero, "Hot spot temperature estimation in mineral oil immersed power transformers using support vectors regression," Informacion Tecnologica, **31**(4), 35–44, 2020, doi:10.4067/S0718-07642020000400035.
[36] R. Cselkó, M. Szirtes, G. Csépes, Speciality and Applicability of Advanced Response Methods in Contrast to the Traditional Dielectric Measurements for Condition Assessment of Power Transformers, 1003–1014, 2020, doi:10.1007/978-3-030-31676-1_94.
[37] S.J.T. Shahrabad, V. Ghods, M.T. Askari, "Power transformer fault diagnosis using DGA and artificial intelligence," Recent Advances in Computer Science and Communications, **13**(4), 579–587, 2020, doi:10.2174/2213275912666190212124133.
[38] H.C. Chen, Y. Zhang, M. Chen, "Online DGA Sensor Calibration Using GANN and Data Augmentation," in Proceedings of 2022 IEEE 5th International Electrical and Energy Conference, CIEEC 2022, 1426–1430, 2022, doi:10.1109/CIEEC54735.2022.9846119.
[39] R. Emadifar, N.T. Kalantari, V. Behjat, R. Najjar, "Monitoring and Condition Assessment of Insulation Paper of Distribution Transformers With Novel Oil Spectroscopy Method," IEEE Transactions on Dielectrics and Electrical Insulation, **29**(5), 1904–1912, 2022, doi:10.1109/TDEI.2022.3199687.
[40] X. Ma, Y. Luo, J. Shi, H. Xiong, "Acoustic Emission Based Fault Detection of Substation Power Transformer," Applied Sciences, **12**(5), 2759, 2022, doi:10.3390/app12052759.
[41] M. Zulkifli, M.Z. Muthaza, I. Safaat, S.S. Yi, N. Sahari, "Thermography Study on Distribution Transformer in Substation: Transformer Core Material," International Journal of Engineering Trends and Technology, **71**(3), 9–16, 2022, doi:10.14445/22315381/IJETT-V71I3P202.
[42] A. Naderian, S. Cress, R. Piercy, F. Wang, J. Service, "An Approach to Determine the Health Index of Power Transformers," in Conference Record of the 2008 IEEE International Symposium on Electrical Insulation, IEEE: 192–196, 2008, doi:10.1109/ELINSL.2008.4570308.
[43] M. Arshad, S. Islam, A. Khaliq, "Fuzzy logic approach in power transformers management and decision making," IEEE Transactions on Dielectrics and Electrical Insulation, **21**(5), 2343–2354, 2014, doi:10.1109/TDEI.2014.003859.
[44] A.F. Cerón, D.F. Echeverry, G. Aponte, A.A. Romero, "Health index for power transformers immersed in mineral oil with voltages between 69kV and 230kV using fuzzy logic," Informacion Tecnologica, **26**(2), 107–116, 2015, doi:10.4067/S0718-07642015000200013.
[45] Juan.P. Lata, Diego.P. Chacon-Troya, R.D. Medina, "Improved tool for power transformer health index analysis," in 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON), IEEE: 1–4, 2017, doi:10.1109/INTERCON.2017.8079656.
[46] C. Ranga, A.K. Chandel, "Expert System for Health Index Assessment of Power Transformers," International Journal on Electrical Engineering and Informatics, **9**(4), 850–865, 2017, doi:10.15676/ijeei.2017.9.4.16.
[47] A.M. Abdullah, R. Ali, S.B. Yaacob, K. Ananda-Rao, N.A. Uloom, "Transformer Health Index by Prediction Artificial Neural Networks Diagnostic Techniques," Journal of Physics: Conference Series, **2312**(1), 012002, 2022, doi:10.1088/1742-6596/2312/1/012002.
[48] D. Luo, R. Xi, L. Che, H. He, "Health condition assessment of transformers based on cross message passing graph neural networks," Frontiers in Energy Research, **10**, 2022, doi:10.3389/fenrg.2022.973736.
[49] K. Ramalingam, C. Indulkar, Solar Energy and Photovoltaic Technology, 2017, doi:10.1016/B978-0-12-804208-3.00003-0.
[50] International Energy Agency, Trends in photovoltaic applications 2022 - Report IEA PVPS T1-43:2022, 2022.
[51] L. Hernández-Callejo, S. Gallardo-Saavedra, V. Alonso-Gómez, "A review of photovoltaic systems: Design, operation and maintenance," Solar Energy, **188**, 426–440, 2019, doi:10.1016/j.solener.2019.06.017.
[52] S. Lu, B.T. Phung, D. Zhang, "A comprehensive review on DC arc faults and their diagnosis methods in photovoltaic systems," Renewable and Sustainable Energy Reviews, **89**, 88–98, 2018, doi:10.1016/j.rser.2018.03.010.

[53] A. Mellit, G.M. Tina, S.A. Kalogirou, "Fault detection and diagnosis methods for photovoltaic systems: A review," Renewable and Sustainable Energy Reviews, **91**, 1–17, 2018, doi:10.1016/j.rser.2018.03.062.

[54] D.S. Pillai, N. Rajasekar, "A comprehensive review on protection challenges and fault diagnosis in PV systems," Renewable and Sustainable Energy Reviews, **91**, 18–40, 2018, doi:10.1016/j.rser.2018.03.082.

[55] D.S. Pillai, N. Rajasekar, "Metaheuristic algorithms for PV parameter identification: A comprehensive review with an application to threshold setting for fault detection in PV systems," Renewable and Sustainable Energy Reviews, **82**, 3503–3525, 2018, doi:10.1016/j.rser.2017.10.107.

[56] Syafaruddin, D.S. Zinger, "Review on methods of fault diagnosis in photovoltaic system applications," Journal of Engineering Science and Technology Review, **12**(5), 53–66, 2019, doi:10.25103/jestr.125.07.

[57] L.B. Bosman, W.D. Leon-Salas, W. Hutzel, E.A. Soto, "PV system predictive maintenance: Challenges, current approaches, and opportunities," Energies, **16**(3), 2020, doi:10.3390/en13061398.

[58] A. Kumaradurai, Y. Teekaraman, T. Coosemans, M. Messagie, "Fault detection in photovoltaic systems using machine learning algorithms: A review," Institute of Electrical and Electronics Engineers Inc., Automotive Technology Research Group (MOBI), Vrije Universiteit Brussel (VUB), Brussels, Belgium, 2020, doi:10.1109/ICOT51877.2020.9468768.

[59] T. Berghout, M. Benbouzid, T. Bentrcia, X. Ma, S. Djurović, L.-H. Mouss, "Machine learning-based condition monitoring for pv systems: State of the art and future prospects," Energies, **14**(19), 2021, doi:10.3390/en14196316.

[60] T. Berghout, M. Benbouzid, X. Ma, S. Djurovic, L.-H. Mouss, "Machine Learning for Photovoltaic Systems Condition Monitoring: A Review," IEEE Computer Society, University of Batna2, Department of Industrial Engineering and Manufacturing, Batna, Algeria, 2021, doi:10.1109/IECON48115.2021.9589423.

[61] C. Feng, Y. Liu, J. Zhang, "A taxonomical review on recent artificial intelligence applications to PV integration into power grids," International Journal of Electrical Power and Energy Systems, **132**, 2021, doi:10.1016/j.ijepes.2021.107176.

[62] A.W. Kandeal, M.R. Elkadeem, A. Kumar Thakur, G.B. Abdelaziz, R. Sathyamurthy, A.E. Kabeel, N. Yang, S.W. Sharshir, "Infrared thermography-based condition monitoring of solar photovoltaic systems: A mini review of recent advances," Solar Energy, **223**, 33–43, 2021, doi:10.1016/j.solener.2021.05.032.

[63] M. Mansouri, M. Trabelsi, H. Nounou, M. Nounou, "Deep Learning-Based Fault Diagnosis of Photovoltaic Systems: A Comprehensive Review and Enhancement Prospects," IEEE Access, **9**, 126286–126306, 2021, doi:10.1109/ACCESS.2021.3110947.

[64] A.K. V de Oliveira, M. Aghaei, R. Rüther, "Automatic Inspection of Photovoltaic Power Plants Using Aerial Infrared Thermography: A Review," Energies, **15**(6), 2022, doi:10.3390/en15062055.

[65] M.I.S. Guerra, F.M.U. de Araújo, J.T. de Carvalho Neto, R.G. Vieira, "Survey on adaptative neural fuzzy inference system (ANFIS) architecture applied to photovoltaic systems," Energy Systems, 2022, doi:10.1007/s12667-022-00513-8.

[66] Z. Yuan, G. Xiong, X. Fu, "Artificial Neural Network for Fault Diagnosis of Solar Photovoltaic Systems: A Survey," Energies, **15**(22), 8693, 2022, doi:10.3390/en15228693.

[67] D.C. Jordan, S.R. Kurtz, "Photovoltaic Degradation Rates-an Analytical Review," Progress in Photovoltaics: Research and Applications, **21**(1), 12–29, 2013, doi:10.1002/pip.1182.

[68] O.A. Alimi, E.L. Meyer, O.I. Olayiwola, "Solar Photovoltaic Modules' Performance Reliability and Degradation Analysis—A Review," Energies, **15**(16), 5964, 2022, doi:10.3390/en15165964.

[69] IEC 62446-1:2016, Photovoltaic (PV) systems - Requirements for testing, documentation and maintenance - Part 1: Grid connected systems - Documentation, commissioning tests and inspection, 1–83, 2016.

[70] IEEE Std 1547-2018, IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces, 1–138, 2018.

[71] IEEE Std 1526-2020, IEEE Recommended Practice for Testing the Performance of Stand-Alone Photovoltaic Systems, 1–32, 2020.

[72] I. Bodnár, D. Matusz-Kalász, R.R. Boros, R. Lipták, "Condition Assessment of Solar Modules by Flash Test and Electroluminescence Test," Coatings, **11**(11), 1361, 2021, doi:10.3390/coatings11111361.

[73] V. Esen, S. Saglam, B. Oral, O. Ceylan Esen, "Toward Class AAA LED Large Scale Solar Simulator With Active Cooling System for PV Module Tests," IEEE Journal of Photovoltaics, **12**(1), 364–371, 2022, doi:10.1109/JPHOTOV.2021.3117912.

[74] S. Choi, S. Park, J. Hong, J. Won, "A Design and Validation of 400 W PV Emulator Using Simple Equivalent Circuit for PV Power System Test," Energies, **16**(4), 1561, 2023, doi:10.3390/en16041561.

[75] C. Han, H. Lee, "A field-applicable health monitoring method for photovoltaic system," Reliability Engineering & System Safety, **184**, 219–227, 2019, doi:10.1016/j.ress.2018.01.002.

[76] P. Ramanan, K. Kalidasa Murugavel, A. Karthick, "Performance analysis and energy metrics of grid-connected photovoltaic systems," Energy for Sustainable Development, **52**, 104–115, 2019, doi:10.1016/j.esd.2019.08.001.

[77] SolarPower Europe, Asset management best practice guidelines Version 2.0, 2020.

[78] IEC, IEC 60076-7. Power transformers - Loading guide for mineral-oil-immersed power transformers, Genova, Switzerland, 2018.

[79] IEC, IEC 60076-7. Power transformers - Loading guide for oil-immersed power transformers., Genova, Switzerland, 2005.

[80] F. Kabir, B. Foggo, N. Yu, "Data driven predictive maintenance of distribution transformers," IEEE Computer Society, Department of Electrical and Engineering, University of California, Riverside, CA 92507, United States: 312–316, 2018, doi:10.1109/CICED.2018.8592417.

[81] G.C. Jaiswal, M.S. Ballal, P.A. Venikar, D.R. Tutakne, H.M. Suryawanshi, "Genetic algorithm–based health index determination of distribution transformer," International Transactions on Electrical Energy Systems, **28**(5), 2018, doi:10.1002/etep.2529.

[82] IEEE, IEEE Std C57.12.00-2010 - Standard for General Requirements for Liquid-Immersed Distribution, Power, and Regulating Transformers, New York, United State, 2010.

[83] K. Azimi Hosseini, M. Hajiaghapour-Moghimi, E. Hajipour, M. Vakilian, "Advanced distribution transformer asset management embracing its relocation: A cost saving approach," IET Generation, Transmission and Distribution, **16**(12), 2370–2385, 2022, doi:10.1049/gtd2.12456.

[84] Meghalaya State Electricity Board, Manual for Maintenance of Distribution Transformers, 2010.

[85] M.M. Youssef, R.A. Ibrahim, H. Desouki, M.M.Z. Moustafa, "An Overview on Condition Monitoring & Health Assessment Techniques for Distribution Transformers," Institute of Electrical and Electronics Engineers Inc., Arab Academy for Science, Technology & Maritime Transport, Alexandria, Egypt: 187–192, 2022, doi:10.1109/ICGEA54406.2022.9791900.

[86] IEEE, IEEE Std 62-1995 - Guide for Diagnostic Field Testing of Electric Power Apparatus - Part 1: Oil Filled Power Transformers, Regulators, and Reactors., 1995.

[87] R. Behkam, H. Karami, M.S. Naderi, G.B. Gharehpetian, "Condition Monitoring of Distribution Transformers Using Frequency Response Traces and Artificial Neural Network to Detect the Extent of Windings Axial Displacements," Institute of Electrical and Electronics Engineers Inc., Amirkabir University of Technology (Tehran Polytechnic), Department of Electrical Engineering, Iran: 18–23, 2022, doi:10.1109/EPDC56235.2022.9817296.

[88] A. Sbravati, R. Breazeal, D. Robalino, "Practical Considerations for the Usage of Ester Fluids in Distribution Transformers," Institute of Electrical and Electronics Engineers Inc., Cargill Bioindustrial, Minneapolis, United States: 366–371, 2022, doi:10.1109/EIC51169.2022.9833213.

[89] R. Chumnuan, D. Rerkpreedapong, "A practicable framework for risk assessment of distribution transformers using PEA smart meter data," in: Y., K., ed., Institute of Electrical and Electronics Engineers Inc., Provincial Electricity Authority (PEA), Engineering Division Region 4, Bangkok, Thailand: 590–594, 2021, doi:10.1109/ECTI-CON51831.2021.9454682.

[90] R. Prananda, H.S. Dini, T.W.O. Putri, "Design of Load Balancing Method on Secondary Distribution Network Using Artificial Intelligence Based on Fuzzy Logic," Institute of Electrical and Electronics Engineers Inc., Institut Teknologi PLN, Faculty of Electricity and Renewable Energy, Jakarta, Indonesia: 372–377, 2021, doi:10.1109/ICHVEPS53178.2021.9600977.

[91] C.-S. Tu, C.-Y. Yang, M.-T. Tsai, "An optimal phase arrangement of distribution transformers under risk assessment," Energies, **13**(21), 2020, doi:10.3390/en13215852.

[92] Z. Yang, Y. Shen, R. Zhou, F. Yang, Z. Wan, Z. Zhou, "A transfer learning fault diagnosis model of distribution transformer considering multi-factor situation evolution," IEEJ Transactions on Electrical and Electronic Engineering, **15**(1), 30–39, 2020, doi:10.1002/tee.23024.

[93] S. Kulkarni, K. Ashok, F. Lambert, D. Divan, "Asset Monitoring using Smart Sensing and Advanced Analytics for the Distribution Network," Institute of Electrical and Electronics Engineers Inc., Center for Distributed Energy, Georgia Institute of Technology, Atlanta, GA, United States, 2019, doi:10.1109/NAPS46351.2019.9000254.

[94] G.C. Jaiswal, M.S. Ballal, D.R. Tutakne, P. Vishnu, "Intelligent condition monitoring system for distribution transformer and health status diagnosis," Institute of Electrical and Electronics Engineers Inc., Additional Executive Engineer, Maharashtra State Elec. Distr. Co. Ltd. and Res. Scholar Shri Ramdeobaba Coll. of Eng. and Mgmt., Nagpur, India: 1–6, 2018, doi:10.1109/PICC.2018.8384786.

[95] M. Mahoor, A. Majzoobi, A. Khodaei, "Distribution asset management through coordinated microgrid scheduling," IET Smart Grid, **1**(4), 159–168, 2018, doi:10.1049/iet-stg.2018.0076.

[96] A.A. Nelson, G.C. Jaiswal, M.S. Ballal, D.R. Tutakne, "Remote condition monitoring system for distribution transformer," Institute of Electrical and Electronics Engineers Inc., Dept of Elect Engg, VNIT, Nagpur, India, 2015, doi:10.1109/NPSC.2014.7103848.

[97] Y.B. Chen, Y. Sun, X.R. Sun, G.H. Sheng, X.C. Jiang, Real-time temperature on-line monitoring and analysis system for transformers, **521**, 409–413, 2014, doi:10.4028/www.scientific.net/AMM.521.409.

[98] B. Noirhomme, J. Jalbert, M.-C. Lessard, F. Brochu, S. Hachichi, "Pole mounted distribution transformers insulation condition assessment," Hydro-Québec - IREQ, Varennes, Canada: 239–243, 2013, doi:10.1109/EIC.2013.6554241.

[99] S. Carvalhosa, H. Leite, F. Branco, C.A. Sá, A.M. Moura, R.C. Lopes, M. Soares, "Survey on the Advancements of Dielectric Fluids and Experiment Studies for Distribution Power Transformers," Renewable Energy and Power Quality Journal, **19**, 97–102, 2021, doi:10.24084/repqj19.225.

[100] Y. Wang, L. Hu, J. Chen, Y. Ren, "Health status diagnosis of distribution transformers based on big data mining," Transactions on Emerging Telecommunications Technologies, **33**(2), 2022, doi:10.1002/ett.3759.

[101] N.S. Shariffuddin, N. Azis, A.M. Selva, M.S. Yahaya, J. Jasni, M.A. Talib, "Investigation on the Relationship between Failure Rates and Health Index of Distribution Transformer Population," Institute of Electrical and Electronics Engineers Inc., Universiti Putra Malaysia, Advanced Lightning, Power and Energy Research Centre (ALPER), Serdang, Malaysia: 119–122, 2021, doi:10.1109/ICPADM49635.2021.9493977.

[102] S.A. Ghani, Y.H.M. Thayoob, Y.Z.Y. Ghazali, M.S.A. Khiar, I.S. Chairul, "Condition monitoring of distribution transformer's mechanical parts using sweep frequency response analysis (SFRA)," Elsevier Ltd, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia: 469–476, 2013, doi:10.1016/j.proeng.2013.12.208.

[103] S. Zhu, C.J. Kikkert, N. Ertugrul, "A wide bandwidth, on-line impedance measurement method for power systems, based On PLC techniques," in 2014 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, School of Electrical and Electronic Engineering, University of Adelaide, Adelaide, SA, Australia: 1167–1170, 2014, doi:10.1109/ISCAS.2014.6865348.

[104] J. Javid, M.A. Mughal, M. Karim, "Using kNN Algorithm for classification of Distribution transformers Health index," Institute of Electrical and Electronics Engineers Inc., HITEC University Taxila, Department of Electrical Engineering, Taxila, Pakistan, 2021, doi:10.1109/ICIC53490.2021.9693013.

[105] R.R. Pawar, S.B. Deosarkar, "Health condition monitoring system for distribution transformer using Internet of Things (IoT)," Institute of Electrical and Electronics Engineers Inc., Department of Electronics and Telecommunication Engineering, Dr. Babasaheb Ambedkar Technological University, Lonere, Raigad, Maharashtra, India: 117–122, 2018, doi:10.1109/ICCMC.2017.8282650.

[106] P. Henault, "Detection of internal arcing faults in distribution transformers," IFD Corporation, United States, 2011, doi:10.1109/TDCLLM.2011.6042220.

[107] L. Bin, S. Sheng, L. Yuan, F. Qiang, X. Yunfeng, L. Guanke, "Spectro-Temporal Self-similarity based Identification of Corrupted Acoustic Signal of Distribution Transformer in Noisy Environment," IEEE Transactions on Power Delivery, 1–12, 2022, doi:10.1109/TPWRD.2022.3181978.

[108] IEC, IEC 60270 - High-voltage test techniques-Partial discharge measurements., 2001.

[109] Network Operators in Great Britain, DNO COMMON NETWORK ASSET INDICES METHODOLOGY Health & Criticality-Version 2.1, 2021.

[110] G.C. Jaiswal, M.S. Ballal, D.R. Tutakne, "Health index based condition monitoring of distribution transformer," Institute of Electrical and Electronics Engineers Inc., Department of Electrical Engineering, Shri Ramdeobaba College of Engineering & Management, Nagpur, India: 1–5, 2017, doi:10.1109/PEDES.2016.7914222.

[111] G.C. Montanari, R. Ghosh, L. Cirioni, G. Galvagno, S. Mastroeni, "Partial Discharge Monitoring of Medium Voltage Switchgears: Self-Condition Assessment Using an Embedded Bushing Sensor," IEEE Transactions on Power Delivery, **37**(1), 85–92, 2022, doi:10.1109/TPWRD.2021.3053658.

[112] I.A. Metwally, "Technology progress in high-voltage gas-insulated substations," IEEE Potentials, **29**(6), 25–32, 2010, doi:10.1109/MPOT.2010.939085.

[113] X. Zhang, E. Gockenbach, "Component reliability modeling of distribution systems based on the evaluation of failure statistics," IEEE Transactions on Dielectrics and Electrical Insulation, **14**(5), 1183–1191, 2007, doi:10.1109/TDEI.2007.4339478.

[114] X. Zhang, E. Gockenbach, V. Wasserberg, H. Borsi, "Estimation of the lifetime of the electrical components in distribution networks," IEEE Transactions on Power Delivery, **22**(1), 515–522, 2007, doi:10.1109/TPWRD.2006.876661.

[115] K. Niayesh, M. Runde, Power Switching Components, Springer International Publishing, Cham, 2017, doi:10.1007/978-3-319-51460-4.

[116] Power and Water Corporation, "Asset Management Plan – Distribution Switchgear," 1–54, 2019.

[117] C. Lowsley, N. Davies, D. Miller, "EFFECTIVE CONDITION ASSESSMENT OF MV SWITCHGEAR," in ea technology, 1–8.

[118] Cigré, Non-intrusive methods for condition assessment of distribution and transmission switchgear, Cigré, 2018.

[119] Y. Alsumaidaee, C. Yaw, S. Koh, S. Tiong, C. Chen, K. Ali, "Review of Medium-Voltage Switchgear Fault Detection in a Condition-Based Monitoring System by Using Deep Learning," Energies, **15**(18), 6762, 2022, doi:10.3390/en15186762.

[120] K. Perdon, L. Cavalli, M. Testa, "INNOVATIVE ANALYTICS TO ESTIMATE THE PROBABILITY OF FAILURE AND REMAINING USEFUL LIFE OF MEDIUM VOLTAGE BREAKERS," in Cired 2017, 12–15, 2017.

[121] A.A. Bhole, W.Z. Gandhare, "An Overview of Dynamic Contact Resistance Measurement of HV Circuit Breakers," Journal of The Institution of Engineers (India): Series B, **97**(2), 219–226, 2016, doi:10.1007/s40031-014-0164-2.

[122] IEC, IEC 62271-1:2017/AMD1:2021 Amendment 1 - High-voltage switchgear and controlgear - Part 1: Common specifications for alternating current switchgear and controlgear, 2021.

[123] X. Jiang, B. Stephen, T. Chandarasupsang, S.D.J. McArthur, B.G. Stewart, "A Gaussian Process Based Fleet Lifetime Predictor Model for Unmonitored Power Network Assets," IEEE Transactions on Power Delivery, **38**(2), 979–987, 2023, doi:10.1109/TPWRD.2022.3203161.

[124] J. Li, Y. Sun, N. Dong, Z. Zhao, "A Novel Contact Temperature Calculation Algorithm in Distribution Switchgears for Condition Assessment," IEEE Transactions on Components, Packaging and Manufacturing Technology, **9**(2), 279–287, 2019, doi:10.1109/TCPMT.2018.2886404.

[125] IEC, IEC 62271-100:2021 High-voltage switchgear and controlgear - Part 100: Alternating-current circuit-breakers, 600, 2021.

[126] EPRI, Plant Support Engineering : Life Cycle Management Planning Sourcebook : Medium-Voltage Switchgear, 2006.

[127] F.Á. Gómez, R. Albarracín-Sánchez, F.G. Vecino, R.G. Arrabé, "Diagnosis of insulation condition of mv switchgears by application of different partial discharge measuring methods and sensors," Sensors (Switzerland), **18**(3), 2018, doi:10.3390/s18030720.

[128] G.A. Hussain, M. Shafiq, L. Kumpulainen, F. Mahmood, M. Lehtonen, "Performance evaluation of noise reduction method during on-line monitoring of MV switchgear for PD measurements by non-intrusive sensors," International Journal of Electrical Power and Energy Systems, **64**, 596–607, 2015, doi:10.1016/j.ijepes.2014.07.057.

[129] W. Kang, Dohoon. Lee, S. Ham, J. Lee, J. Lee, D. Kim, T. Kwon, "Development of a UHF PD detection system to estimate the dielectric condition of a Medium Voltage Switchgear," in Conference Record of the 2008 IEEE International Symposium on Electrical Insulation, 377–380, 2008, doi:10.1109/ELINSL.2008.4570353.

[130] J.A. Kay, G.A. Hussain, M. Lehtonen, L. Kumpulainen, "New preemptive arc-fault detection techniques in medium-voltage switchgear and motor controls," IEEE Transactions on Industry Applications, **52**(1), 740–750, 2016, doi:10.1109/TIA.2015.2466620.

[131] S. Ishak, S.P. Koh, J.D. Tan, S.K. Tiong, C.P. Chen, C.T. Yaw, "Arcing Faults Detection in Switchgear with Extreme Learning Machine," Journal of Physics: Conference Series, **2319**(1), 012007, 2022, doi:10.1088/1742-6596/2319/1/012007.

[132] X. Zheng, Z. Wang, "Partial Discharge Fault Diagnosis of Switchgear Based on APSO-BP Algorithm," in 2021 IEEE 2nd China International Youth Conference on Electrical Engineering (CIYCEE), IEEE, Dalian

University of Technology, China: 1–7, 2021, doi:10.1109/CIYCEE53554.2021.9676811.

[133] Q. Che, H. Wen, X. Li, Z. Peng, K.P. Chen, "Partial Discharge Recognition Based on Optical Fiber Distributed Acoustic Sensing and a Convolutional Neural Network," IEEE Access, **7**, 101758–101764, 2019, doi:10.1109/ACCESS.2019.2931040.

[134] S. Barrios, D. Buldain, M.P. Comech, I. Gilbert, "Partial Discharge Identification in MV Switchgear Using Scalogram Representations and Convolutional AutoEncoder," IEEE Transactions on Power Delivery, **36**(6), 3448–3455, 2021, doi:10.1109/TPWRD.2020.3042934.

[135] H. Li, X. Tong, F. Yang, D. Huang, "Switchgear Insulation Condition Assessment with Multi-Dimensional Feature Database Based on an Optimal Grade Clustering Method," in: B., X. and K., M., eds., in 2021 IEEE 5th Information Technology,Networking,Electronic and Automation Control Conference (ITNEC), IEEE, Shanghai Marine Equipment Research Institute, Shanghai, China: 294–299, 2021, doi:10.1109/ITNEC52019.2021.9587116.

[136] Z. Zhang, H. Xia, S. Cao, C. Liu, "Research on feature extraction and fault diagnosis of mechanical vibration and sound signal of Disconnector," in 2022 IEEE 5th International Electrical and Energy Conference (CIEEC), IEEE, Shandong University, School of Electrical Engineering, Jinan, China: 1119–1124, 2022, doi:10.1109/CIEEC54735.2022.9846229.

[137] Z. Qiu, J. Ruan, D. Huang, "Mechanical fault diagnosis of outdoor high-voltage disconnector," IEEJ Transactions on Electrical and Electronic Engineering, **11**(5), 556–563, 2016, doi:10.1002/tee.22273.

[138] C. Nicolaou, A. Mansour, K. Van Laerhoven, "On-site Online Condition Monitoring of Medium-Voltage Switchgear Units," ACM International Conference Proceeding Series, 1–8, 2021, doi:10.1145/3494322.3494323.

[139] S. Yan, G. Zhang, B. Chen, C. Li, Y. Liu, J. Yang, Fault Diagnosis Technology of Low Voltage Circuit Breaker Based on Multilayer Perceptron, Springer Science and Business Media Deutschland GmbH, Power Supply Branch of Beijing Metro Operation Co., Ltd., Beijing, 100084, China: 332–344, 2022, doi:10.1007/978-981-19-1532-1_35.

[140] A.S.N. Huda, S. Taib, "Application of infrared thermography for predictive/preventive maintenance of thermal defect in electrical equipment," Applied Thermal Engineering, **61**(2), 220–227, 2013, doi:10.1016/J.APPLTHERMALENG.2013.07.028.

[141] Z. Chen, T. Yuan, M. Xu, Y. Yu, Z. Xie, Y. Liu, "Research on Fault Diagnosis of Switchgear Based on Temperature Cloud Image Technology," IOP Conference Series: Earth and Environmental Science, **1044**(1), 012007, 2022, doi:10.1088/1755-1315/1044/1/012007.

[142] Y. Yan, D. Qi, H. Gu, "A Real-Time IR-Fusion Switchgear Contact Monitoring System (SCMS)," IEEE Access, **5**, 12114–12124, 2017, doi:10.1109/ACCESS.2017.2698060.

[143] R. Salustiano, R.M. Capelini, S.R. De Abreu, M.L.B. Martinez, I.C. Tavares, G.M.F. Ferraz, M.A.A. Romano, "Development of new methodology for insulators inspections on aerial distribution lines based on partial discharge detection tools," ICHVE 2014 - 2014 International Conference on High Voltage Engineering and Application, 2014, doi:10.1109/ICHVE.2014.7035429.

[144] M. Shafiq, I. Kiitam, P. Taklaja, L. Kutt, K. Kauhaniemi, I. Palu, "Identification and Location of PD Defects in Medium voltage Underground Power Cables Using High Frequency Current Transformer," IEEE Access, **7**, 103608–103618, 2019, doi:10.1109/ACCESS.2019.2930704.

[145] N. Davari, G. Akbarizadeh, E. Mashhour, "Intelligent Diagnosis of Incipient Fault in Power Distribution Lines Based on Corona Detection in UV-Visible Videos," IEEE Transactions on Power Delivery, **36**(6), 3640–3648, 2021, doi:10.1109/TPWRD.2020.3046161.

[146] H. Manninen, J. Kilter, M. Landsberg, "Advanced condition monitoring method for high voltage overhead lines based on visual inspection," IEEE Power and Energy Society General Meeting, **2018-August**, 2018, doi:10.1109/PESGM.2018.8586498.

[147] D.R. Morais, J.G. Rolim, J. Coser, H.H. Zürn, "Reliability centered maintenance for capacitor voltage transformers," in 2006 9th International Conference on Probabilistic Methods Applied to Power Systems, PMAPS, 2006, doi:10.1109/PMAPS.2006.360281.

[148] Y. Wu, M. Liu, Z. Ji, Z. Diao, Y. Wu, F. Yang, "A Fault Diagnosis System of All-Fiber Optical Current Transformer with Correlation Detection," in Journal of Physics: Conference Series, 2020, doi:10.1088/1742-6596/1550/4/042041.

[149] P. Werelius, B. Holmgren, U. Gafvert, "Diagnosis of medium voltage XLPE cables by high voltage dielectric spectroscopy," in ICSD'98. Proceedings of the 1998 IEEE 6th International Conference on Conduction and Breakdown in Solid Dielectrics (Cat. No.98CH36132), 79–84, 1998, doi:10.1109/ICSD.1998.709231.

[150] H. van Jaarsveldt, R. Gouws, "Condition Monitoring of Medium Voltage Electrical Cables by Means of Partial Discharge Measurements," SAIEE Africa Research Journal, **105**(4), 136–146, 2014, doi:10.23919/SAIEE.2014.8538345.

[151] C.-K. Chang, C.-S. Lai, R.-N. Wu, "Decision tree rules for insulation condition assessment of pre-molded power cable joints with artificial defects," IEEE Transactions on Dielectrics and Electrical Insulation, **26**(5), 1636–1644, 2019, doi:10.1109/TDEI.2019.008208.

[152] R.M. Arias Velásquez, J.V. Mejía Lara, "Ruptures in overhead ground wire — Transmission line 220 kV," Engineering Failure Analysis, **87**, 1–14, 2018, doi:10.1016/J.ENGFAILANAL.2018.01.003.

[153] International Electrotechnical Commission - IEC, INTERNATIONAL STANDARD - IEC 61395:1998 Overhead electrical conductors - Creep test procedures for stranded conductors, 1998.

[154] American Society for Testing and Materials International - ASTM, ASTM B1008-18 Standard Test Method for Stress-Strain Testing for Overhead Electrical Conductors.

[155] American Society for Testing and Materials International - ASTM, ASTM D1036-99 Standard Test Methods of Static Tests of Wood Poles, 2017.

[156] American Society for Testing and Materials International - ASTM, ASTM C1089-19 Standard Specification for Spun Cast Prestressed Concrete Poles.

[157] International Electrotechnical Commission - IEC, NTERNATIONAL STANDARD - IEC 60383-1 Insulators for overhead lines with a nominal voltage above 1000 V - Part 1: Ceramic or glass insulator units for a.c. systems - Definitions, test methods and acceptance crite, 1993.

[158] ANSI/NEMA, ANSI/NEMA C29.1-2018 Test Methods for Electrical Power Insulators, 2018.

[159] K. Ibrahim, R.M. Sharkawy, H.K. Temraz, M.M.A. Salama, "Reliability calculations based on an enhanced transformer life expectancy model," Ain Shams Engineering Journal, **13**(4), 2022, doi:10.1016/j.asej.2021.101661.

[160] D.M.R. Vanegas, S.M. Mahajan, "Effects of thermal accelerated ageing on a medium voltage oil-immersed current transformer," Conference Record of IEEE International Symposium on Electrical Insulation, 470–473, 2008, doi:10.1109/ELINSL.2008.4570375.

[161] IEEE Standard Requirements for Instrument Transformers, IEEE Std C57.13-2016 (Revision of IEEE Std C57.13-2008), 1–96, 2016, doi:10.1109/IEEESTD.2016.7501435.

[162] International Electrotechnical Commission - IEC, INTERNATIONAL STANDARD IEC 60044-2 Instrument transformers -Part 2: Inductive voltage transformers, 2003.

[163] International Electrotechnical Commission - IEC, INTERNATIONAL STANDARD - IEC 60044-1 Instrument transformers - Part 1: Current transformers, 2003.

[164] Z. Tiancheng, Y. Ying, S. Jinhai, X. Zhihao, L. Jiashuai, "Infrared Image Segmentation Method of Current Transformer Based on DeepLabv3+ Network," in 2022 IEEE 5th International Electrical and Energy Conference (CIEEC), IEEE, State Grid Jilin Electric Power Co., Ltd., Electric Power Research Institute, Changchun, China: 3111–3115, 2022, doi:10.1109/CIEEC54735.2022.9846719.

[165] B.K. Gupta, J. Densley, A. Narang, "Review of diagnostic techniques for oil-paper insulated instrument transformers," in 2010 IEEE International Symposium on Electrical Insulation, 1–5, 2010, doi:10.1109/ELINSL.2010.5549786.

[166] A. Jeanneton, C. Perrier, A. Beroual, "Thermal Aging Study of Alternative Liquids for Oil-Paper High-Voltage Insulation in Current Transformers," Proceedings - IEEE International Conference on Dielectric Liquids, **2022-May**, 2022, doi:10.1109/ICDL49583.2022.9830941.

[167] International Electrotechnical Commission - IEC, INTERNATIONAL STANDARD IEC 60599:2022 Mineral oil-filled electrical equipment in service - Guidance on the interpretation of dissolved and free gases analysis, 2022.

[168] Ning Hao, Zhuo Dong, "Condition assessment of current transformer based on multi-classification support vector machine," in Proceedings 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE), IEEE, NORTH CHINA BAODING ELECTRIC POWER VOC. and TECH.COLLEGE, Baoding, Hebei, China: 2402–2405, 2011, doi:10.1109/TMEE.2011.6199705.

[169] O.D.L. Campos, H. Zhu, "Condition assessment of 653 instrument transformers with on-line PD testing," in 2016 International Conference on Condition Monitoring and Diagnosis (CMD), IEEE, Costa Rican Institute of Electricity, Costa Rica: 168–171, 2016, doi:10.1109/CMD.2016.7757797.

[170] G.A.T.N. Aravinda, D.C. Attaragama, I.T. Palawatte, M.A.R.M. Fernando, G.A. Jayantha, J.R.S.S. Kumara, "Condition monitoring of current transformers using frequency domain dielectric spectroscopy," in 2013 IEEE 8th International Conference on Industrial and Information Systems, IEEE, Faculty of Engineering, University of Peradeniya, Sri Lanka: 168–173, 2013, doi:10.1109/ICIInfS.2013.6731975.

[171] G. Iyer, R. Gorur, A. Krivda, P. Mahonen, "Prediction of electrical performance of medium voltage epoxy insulated equipment," IEEE Transactions on Dielectrics and Electrical Insulation, **17**(2), 334–342, 2010, doi:10.1109/TDEI.2010.5448086.

[172] T. Lei, M. Faifer, R. Ottoboni, S. Toscani, "On-line fault detection technique for voltage transformers," Measurement: Journal of the International Measurement Confederation, **108**, 193–200, 2017, doi:10.1016/j.measurement.2017.03.002.

[173] L. Zhang, H. Chen, Q. Wang, N. Nayak, Y. Gong, A. Bose, "A Novel On-Line Substation Instrument Transformer Health Monitoring System Using Synchrophasor Data," IEEE Transactions on Power Delivery, **34**(4), 1451–1459, 2019, doi:10.1109/TPWRD.2019.2905426.

[174] B. Cui, A.K. Srivastava, P. Banerjee, "Synchrophasor-Based Condition Monitoring of Instrument Transformers Using Clustering Approach," IEEE Transactions on Smart Grid, **11**(3), 2688–2698, 2020, doi:10.1109/TSG.2019.2960043.

[175] S. Tee, Q. Liu, Z. Wang, F. Hafid, P. Tournet, "Failure investigation and asset management of combined measuring instrument transformers," High Voltage, **6**(1), 61–70, 2021, doi:10.1049/hve2.12029.

[176] H. Guo, L. Guo, "Health index for power transformer condition assessment based on operation history and test data," Energy Reports, **8**, 9038–9045, 2022, doi:10.1016/j.egyr.2022.07.041.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 6, 37-45 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Infrastructure-as-a-Service Ontology for Consumer-Centric Assessment

Thepparit Banditwattanawong[1], Masawee Masdisornchote[*,2]

[1]*Department of Computer Science, Kasetsart University, Krung Thep Maha Nakhon, 10900, Thailand*

[2]*School of Information Technology, Sripatum University, Krung Thep Maha Nakhon, 10900, Thailand*

ARTICLE INFO

Article history:
Received:07 September, 2023
Accepted:09 November, 2023
Online: 30 November, 2023

Keywords:
Cloud computing
Infrastructure-as-a-Service (IaaS)
Knowledge engineering
Ontology
Taxonomy

ABSTRACT

In the context of adopting cloud Infrastructure-as-a-Service (IaaS), prospective consumers need to consider a wide array of both business and technical factors associated with the service. The development of an intelligent tool to aid in the assessment of IaaS offerings is highly desirable. However, the creation of such a tool requires a robust foundation of domain knowledge. Thus, the focus of this paper is to introduce an ontology specifically designed to characterize IaaSs from the consumer's perspective, enabling informed decision-making. The ontology additionally serves two purposes of other relevant parties besides the consumers. Firstly, it empowers IaaS providers to better tailor their services to align with consumer expectations, thereby enhancing their competitiveness. Additionally, IaaS partners can play a pivotal role in supporting both consumers and providers by understanding the protocol outlined in the ontology that governs interactions between the two parties. By applying principles of ontological engineering, this study meticulously examined the various topics related to IaaS as delineated in existing cloud taxonomies. These topics were subsequently transformed into a standardized representation and seamlessly integrated through a binary integration approach. This process resulted in the creation of a comprehensive and cohesive ontology that maintains semantic consistency. Leveraging Protégé, this study successfully constructed the resultant ontology, comprising a total of 340 distinct classes. The study evaluated the syntactic, semantic, and practical aspects of the ontology against a worldwide prominent IaaS. The results showed that the proposed ontology was syntactically and semantically consistent. Furthermore, the ontology successfully enabled not only the assessment of a real leading IaaS but also queries to support automation tool development.

1. Introduction

In the process of cloud adoption, there are three parties involved [1]. First, cloud service customers need to assess potential services to determine the best fit for their business and technical requirements. Presently there are numerous Infrastructure as a Service (IaaS) options available across multiple providers. This makes it challenging for potential IaaS customers to assess the different options and make optimal decisions about cloud adoption and/or migration. Second, cloud service providers strive to deliver cloud services that fully satisfy customers' expectations. However, these expectations remain partially unknown due to the lack of comprehensive checklists. This hinders the readiness improvement of offered IaaSs. Finally, cloud service partners facilitate the activities of both customers and providers based on mutually agreed-upon protocols between both IaaS customers and providers. Unfortunately, such protocols are currently lacking.

One promising solution to address these problems is to enhance the comprehension of customers' concerns and service judgment criteria via the introduction of an IaaS knowledge base that is understandable by all parties. As such, the contributions of this paper are twofold.

- Firstly, this paper proposes a semantic model, an ontology, that facilitates the decision making of IaaS adoption. The ontology is constructed from a consumer perspective and potentially serves as a foundational knowledge base for the development of various IaaS assessment tools, such as score-based comparing systems, recommendation systems, and expert systems.

[*]Corresponding Author: Masawee Masdisornchote, masawee.ma@spu.ac.th

- Secondly, this paper demonstrates an innovative approach to constructing this ontology by leveraging existing IaaS taxonomies. This approach not only extends benefits to researchers but also to industry practitioners. For instance, it can serve as a launchpad for further research and development concerning Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS) ontologies, thus fostering advancement in the field as a whole.

The paper is organized as follows. Next section reviews related ontologies and their limitations in addressing our research problem. Section 3 describes the systematic formulation of a proposed ontology using the ontology engineering principle. Section 4 evaluates the practicality of the ontology with real-world IaaSs. The paper concludes in Section 5, summarizing our key findings.

2. Related Work

Aiming to facilitate cloud customers as our work, the author in [2] recently proposed a KPI-based framework for evaluating and ranking IaaSs, PaaSs, and SaaSs to help cloud users identify cloud services that best suit their needs. However, the framework consists of only 41 KPIs, lacking several practical details for effective decision making. The framework has not been evaluated using real cloud service data. In [3], the authors proposed a recommender system using quality of cloud services for selecting cloud services that satisfied end user requirements. However, the quality of service (QoS) attributes were rather limited: response time, availability, throughput, dependability, reliability, price, and reputation. The authors in [4] proposed a signature-based QoS performance discovery algorithm to select IaaS. The algorithm leveraged the combination of service trial experiences and IaaS signatures. Each signature represented a provider's long-term performance behavior for a service over a fixed period. Similarities between users' service trial experiences and IaaS signatures were measured to select proper IaaSs. Nevertheless, the algorithm merely focused on a performance aspect. In [5], the authors proposed an IaaS selection algorithm based on utility functions and deployment knowledge base, which stored application execution histories. Unfortunately, the authors provided no details about the knowledge base's abstraction structure to be evaluated for its IaaS aspect coverage.

Cloud computing ontologies that encompass IaaS concepts, which particularly represented decision-making factors in cloud service adoption, are limited as follows. As many clouds vendors and standards employed inconsistent terminology to define their services, a common ontology in [6] was proposed to provide an approach to discover and use services in cloud federations. The ontology enables applications to negotiate cloud services as requested by users. Although the ontology included IaaS related terms such as some resources and specific services, it fell short of encompassing the complete spectrum of IaaS-related aspects essential for a comprehensive customer-oriented IaaS ontology. An ontology in [7] was proposed to implement intelligent service discovery and management systems for searching and retrieving appropriate cloud services in an accurate and quick manner. The ontology consisted of cloud computing concepts in general including some IaaS related ones such as compute, network, and storage services. Nevertheless, it remained incomplete as an IaaS ontology in its entirety. The authors in [8] proposed an ontology to define functional and non-functional concepts, attributes and relations of infrastructure services and used it to implement a cloud recommendation system. The ontology was composed of detailed services, compute, network, and storage, totally 26 classes. So, the ontology left several additional IaaS facets unaddressed. Several years later, the authors in [9] improved the ontology in [8] by adding 11 classes of price and quality-of-service concepts. The ontology was evaluated by being deployed in the development of semantic data sets sourced from Azure and Google Cloud services.

In summary, the IaaS concepts in aforementioned ontologies are only parts of our proposed ontology, which is more comprehensive by incorporating numerous customer-perspective IaaS taxonomies. Our preliminary work [10] is significantly extended in this paper to incorporate more recent taxonomies as follows. The authors in [11] extended [8] to derive the taxonomy of interoperability in IaaS cloud. The taxonomy's main topics were access mechanism, virtual appliance, network, and service level agreement (SLA). In [12], the author proposed the taxonomy of IaaS services where its main topics were Hardware as a Service (HaaS), which provisions hardware resources, and Infrastructure Services as a Service (ISaaS), which provided a set of auxiliary services to enable successful HaaS provision. The authors in [13] proposed a cloud computing services taxonomy including main topics: main service category, license type, intended user group, payment system, formal agreement, security measures, and standardization efforts. In [14], the authors proposed the taxonomy of fundamental IaaS components consisting of main topics: support layer, management layer, security layer, and control layer. In [15], the authors proposed a comprehensive taxonomy of cloud pricing consisting of three pricing strategies (i.e., value-based pricing, market-based pricing, and cost-based pricing) and nine pricing categories (e.g., retail-based pricing and utility-based pricing). The authors in [16] proposed a SLA taxonomy for PaaS and SaaS besides IaaS.

Furthermore, this paper resolves all conflicts and redundancy among the taxonomies and merges them by using a binary approach, which is more natural but less automated than an identified tabular list previously introduced in [17]. In addition, this paper transforms the integrated taxonomies into a unified ontology in order to allow wider applicability.

3. Ontology Formulation

This section presents the formulation of our proposed ontology based on the principle of ontology engineering [18]. The formulation process entails four steps: taxonomy reuse, refinement, formalization, and evaluation.

3.1. Taxonomy Reuse

A taxonomy serves as a structured arrangement of pertinent topics and subtopics designed for classification purposes. This study conducted a thorough examination of existing IaaS-related taxonomies. The period from 2009 to 2011 marked a significant juncture for cloud computing when Gartner Inc. positioned it at the zenith of the emerging technology hype cycle, garnering worldwide attention. Consequently, our focus was directed exclusively towards taxonomies created after 2009, as they began to gain widespread conceptual clarity and recognition during this

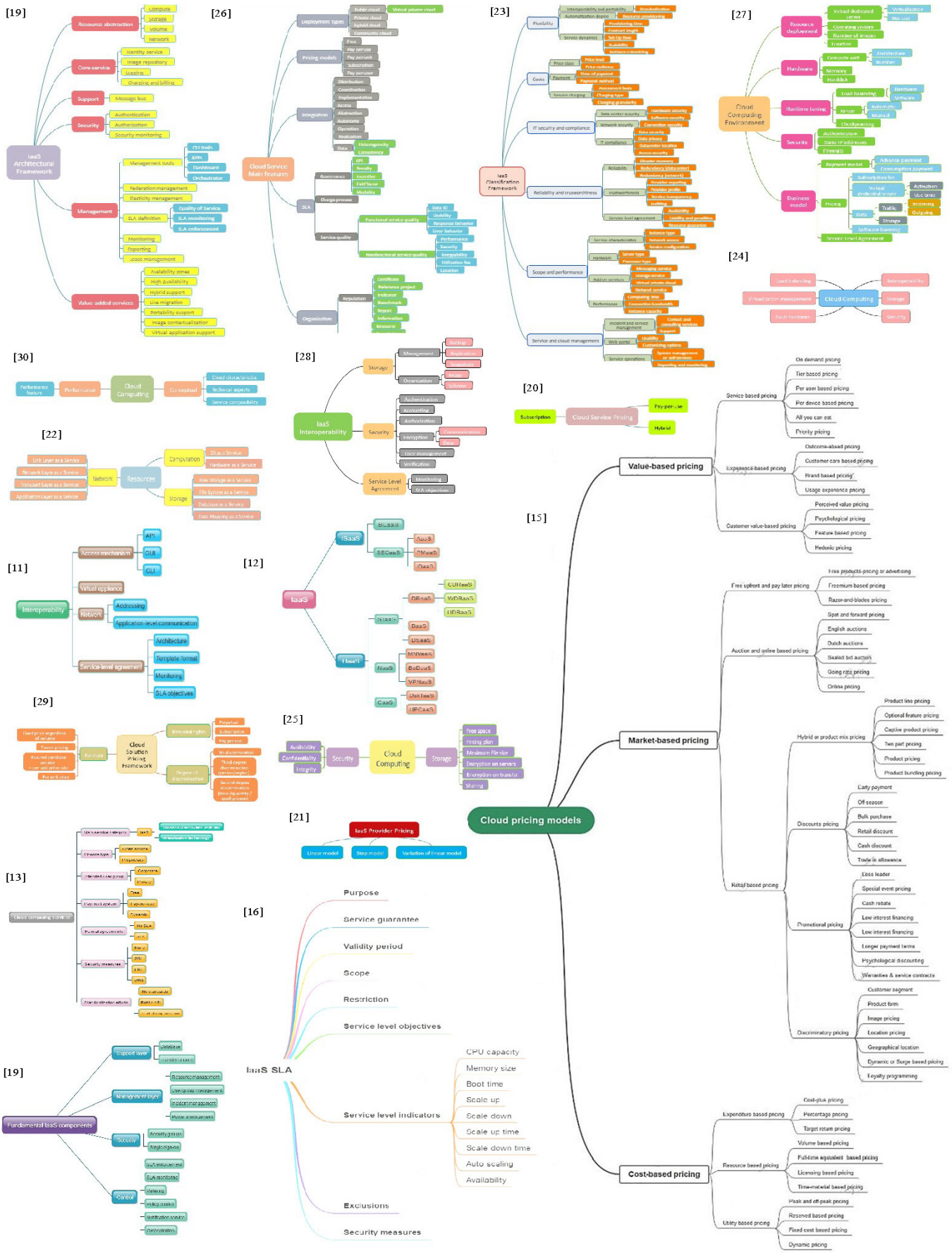


Figure 1: Tenant-centric IaaS excerpted and transformed taxonomies

period. While these taxonomies encompassed concepts found in all the ontologies examined in the preceding section, they were not primarily geared towards IaaS. Therefore, a careful analysis was required to identify IaaS-specific topics that are directly relevant to customers. Table 1 provides an overview of the taxonomies employed in this paper, along with the proportion of tenant-centric IaaS topics extracted from each taxonomy.

Table 1: Customer-centric topic excerption from IaaS taxonomies

Taxonomy	Excerption	Taxonomy	Excerption
[19]	100%	[24]	67%
[20]	100%	[25]	65%
[21]	100%	[26]	61%
[15]	100%	[27]	60%
[22]	98%	[11]	54%
[12]	96%	[28]	49%
[23]	95%	[29]	41%
[19]	75%	[16]	41%
[13]	73%	[30]	33%

Reusing these taxonomies in their original formats presents a challenge, given their diverse presentation styles, which encompass textual descriptions and various graphical models such as mind maps, feature models, decision trees, layered block diagrams, SBIFT models, and textual lists. To create a cohesive and consistent representation while eliminating conflicts and redundancies, this study chose to employ a standardized mind map approach. This allowed us to amalgamate all the extracted taxonomies into a unified model. Figure 1 visually depicts the consumer-centric IaaS excerpts that have been individually transformed into mind maps.

3.2. Refinement

This step analyzed semantic consistencies among the taxonomies from the previous step and subsequently merged them into a unified taxonomy by employing a binary integration approach. The algorithm of the binary integration is as follows.

- Step 1: A pair of taxonomies from Figure 1 that have some common topic(s) (which will be used as a merging point) is selected each time. For example, taxonomies in [20] and [26] have cloud service pricing and pricing models, respectively, as a common topic.
- Step 2: Any redundant and inconsistent topics between both taxonomies from step 1 were identified. For example, subscription in [20] is redundant with subscription in [26].
- Step 3: All redundant topics if there is any, except the one to be used as merging point(s), in both taxonomies are removed to retain the topic's uniqueness. For example, only subscription in [26] is removed.
- Step 4: Any synonymous topics are resolved by choosing the most appropriate topic and renamed the others to be the chosen one. For example, since cloud service pricing in [20] is synonymous with pricing models in [26], cloud service pricing in [20] is renamed to pricing models as that of [26].
- Step 5: Any homonymous topics are resolved by renaming each of them to a distinct term. For example, since availability in [25] and availability in [23] refer to the availability aspect of security and SLA, respectively, availability in [23] is renamed to availability aspect instead.
- Step 6: Merge both taxonomies into a single one by using the merging points. For example, taxonomy in [20] and taxonomy in [26] are merged by using the same topic pricing models.
- Step 7: Repeat step 1 to step 6 for the remaining pairs of taxonomies, including the merged one resulting from step 6, until a unified and consistent taxonomy is achieved.

As a result of refinement, the algorithm resolved 121 (sub)topics out of the total (sub)topics that were redundant and 67 inconsistent (sub)topics that held synonyms and homonyms to obtain a unified taxonomy at last.

3.3. Formalization

All of nonredundant topics and subtopics in the unified taxonomy were converted into classes and subclasses that were connected to other classes and subclasses based on semantic domains. The root class is denoted as "owl: Thing". Consequently, the proposed ontology is composed of 15 direct subclasses, as illustrated in Figure 2: Performance, Resource abstraction, Core service, Support, Costs, Security, Management, Value-added services, Resource deployment, Control, Standardization efforts, Flexibility, Deployment types, Integration, and Organization. The total number of subclasses is 340, such as Performance feature, Computing time, and Connection bandwidth, that are arranged hierarchically in a tree structure with a height of 7, as depicted in Figure 3.

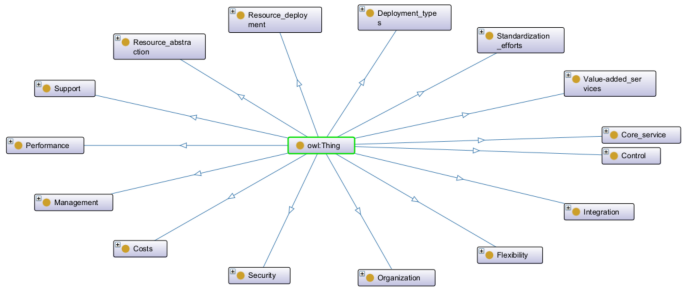


Figure 2: Proposed ontology and its direct subclasses

To provide a comprehensive understanding of the proposed ontology, each of direct subclasses is essentially described one by one, including its nested subclasses enclosed within curly brackets. Furthermore, supplementary explanations are provided within parentheses immediately following respective (sub)classes.

- Performance of IaaS has the following subclasses : Performance feature (identifying the atomic elements of cloud-service performance evaluation), Computing time, Connection bandwidth, Instance capacity, and Load balancing of Hardware or software type.
- Resource abstraction is various resources offered as services. Its subclass and nested subclasses are as follows. Compute = {OS as a Service = {Software licensing = {Open-source,Proprietary}}, Desktop as a Service (DskTaaS), High Performance Computing as a Service (HPCaaS)}, Storage = {Raw Storage as a Service (i.e., block level storage), File

System as a Service, Storage management = {Free space, Maximum file size, Encryption on servers, Encryption on transfer, Sharing, Backup, Replication, Snapshots}, Storage organization = {Image, Scheme (i.e., block, file, or object storage)}, Data Storage as a Service (DRaaS) = {Cold-site DRaaS (CDRaaS), Warm-site DRaaS (WDRaaS), Hot-site DRaaS (HDRaaS)}, Backup as a Service (BaaS) = { Data Storage as a Service (DSaaS)}}. Network = {Link Layer as a Service, Network Layer as a Service, Transport Layer as a Service, Application Layer as a Service, Traffic = {Incoming, Outgoing}, Addressing (providing accessibility to applications and underlying virtual machines once moved to new networks), Application-level communication (specifying API to be RESTful to decouple client from server components and advocate interoperable IaaSs via standard interfaces), Mobile Network Virtualization as a Service (MNVaaS), Bandwidth on Demand as a Service (BoDaaS), Virtual Private Network as a Service (VPNaaS)}, Infrastructure Services as a Service (ISaaS) (which provides a set of auxiliary services) = {Billing as a Service (BLaaS), Security as a Service (SECaaS) (encompassing Identity as a Service (IDaaS) for managing authentication and authorization), Auditing as a Service (AaaS) (for checking providers for standard compliance), Policy Management as a Service (PMaaS) (for handling all access policies across multiple providers)}}.

- Core service is fundamental to all other services. Its subclasses are {Image repository, Charging and billing, SECaaS = {Auditing as a Service (AaaS) (i.e., Logging), Policy Management as a Service (PMaaS), Identity as a Service (IDaaS)}}.
- Support is services used to operate some other services. It has three subclasses: Message bus (providing a means for passing messages between different cloud services), Database, and Transfer service (for other layers to communicate and interact).
- Costs = {Price class = {Price level (all factors affecting resulting cost directly), Price resilience (price options for flexibility purpose)}, Payment = {Time of payment, Payment method, Payment model = {Advance payment, Consumption payment}}, Service charging = {Assessment basis (how regular billing occurs such as hourly or monthly), Charging granularity}, Pricing models = {Value-based pricing (estimating customers' satisfaction) = {Service-based pricing (focusing on service content) = {On-demand pricing, Tier-based pricing, Per-user-based pricing, Per-device-based pricing, All-you-can eat (buffet pricing), Priority pricing}, Experience-based pricing (based on performance) = {Outcome-based pricing, Customer-care-based pricing, Brand-based pricing, Usage-experience pricing}, Customer-value-based pricing (a price from a subjective view of a customer) = {Perceived-value pricing, Psychological pricing, Feature-based pricing, Hedonic pricing}}, Market-based pricing (equilibrium of customers and providers) = {Free-upfront-and-pay-later pricing = {Free products-pricing-on-advertising, Freemium-based pricing, Razor-and-blades pricing (giving away nonconsumable element and charging consumable replacement element)}, Auction-and-online-based pricing = {Spot-and-forward pricing (i.e., current and future prices), English auctions (Open Ascending), Dutch auctions (Open Descending), Sealed-bid auction, Going-rate pricing, Online pricing}, Retail-based pricing (for small quantity purchase) = {Hybrid-or-product-mix pricing (combining different pricing models) = {Product-line pricing, Optional-feature pricing, Captive-product pricing (i.e., cheap core part with costly accessory), Two-part pricing, By-product pricing, Product-bundling pricing}, Discounts pricing = {Early payment, Off-season, Bulk purchase, Retail discount, Cash discount, Trade in allowance (discount in exchange of buyer's asset)}, Promotional pricing = {Loss leader (selling below market price), Special event pricing, Cash rebate, Low interest financing, Longer payment terms, Psychological discounting, Warranties & service contracts}, Discriminatory pricing (charging different prices to different customers) = {Customer segment, Product form (different prices for different versions of a product), Image pricing, Location pricing, Geographical location, Dynamic or surge-based pricing (based on current market demands), Loyalty programming (rewarding customers to continue buying from the brand}}}, Cost-based pricing (covering Capex and Opex) = {Expenditure-based pricing = {Cost-plus pricing (cost plus margin), Percentage pricing, Target-return pricing}, Resource-based pricing = {Volume-based pricing, FTE (full-time equivalent)-based pricing, Licensing-based pricing, Time-material-based pricin}, Utility-based pricing = {Peak-and-off-peak pricing, Reserved-based pricing (e.g., Subscription or no up-front, partial up-front, and all up-front), Fixed-cost-based pricing, Dynamic pricing}}}}, Formula = {Linear model, Step model, Variation of linear model}, Degree of discrimination (how a service is offered for different buyers for different prices) = {No discrimination, Second degree discrimination (when providers sell different units for different prices where customers must do self-selection to choose from the offers), Third degree discrimination (vendor identifies different customer groups based on their willingness-to-pay and can be personal (e.g., student discounts) or regional (e.g., different prices for developing countries))}}}.
- Security = {Availability = {Confidentiality, Integrity, Fault tolerance, Authentication, Authorization, Accounting, Security monitoring, Static IP address, Firewalls, Data center security = {Hardware security, Software security, Data center redundancy}, Network security = {Connection security, Network redundancy}, IT compliance = {Data security, Data privacy, Datacenter location, Access security, Disaster recovery}, User management, Verification, Security groups, Single sign-on, Security measures = {None, SSL, PKI, VPN}}.
- Management = {Management tools = {CLI tools, APIs, Dashboard, Orchestrator, Federation management, Elasticity management}, Formal agreement = {No SLA, SLA = {SLA objectives (defining quality-of-service measurement in SLAs), Scope, Quality of Service, Functional service quality = {Data IO, Usability, Response behavior, Error behavior}, Nonfunctional service quality = {Availability aspect, Performance aspect, Security aspect, Integrability, Utilization fee, Location}, SLA monitoring, SLA enforcement, Governance = {KPI, Penalty, Incentive, Exit clause, Modality}, Charge process, Liability and penalties, Validity period (of negotiated SLA), Resource guarantee, Architecture

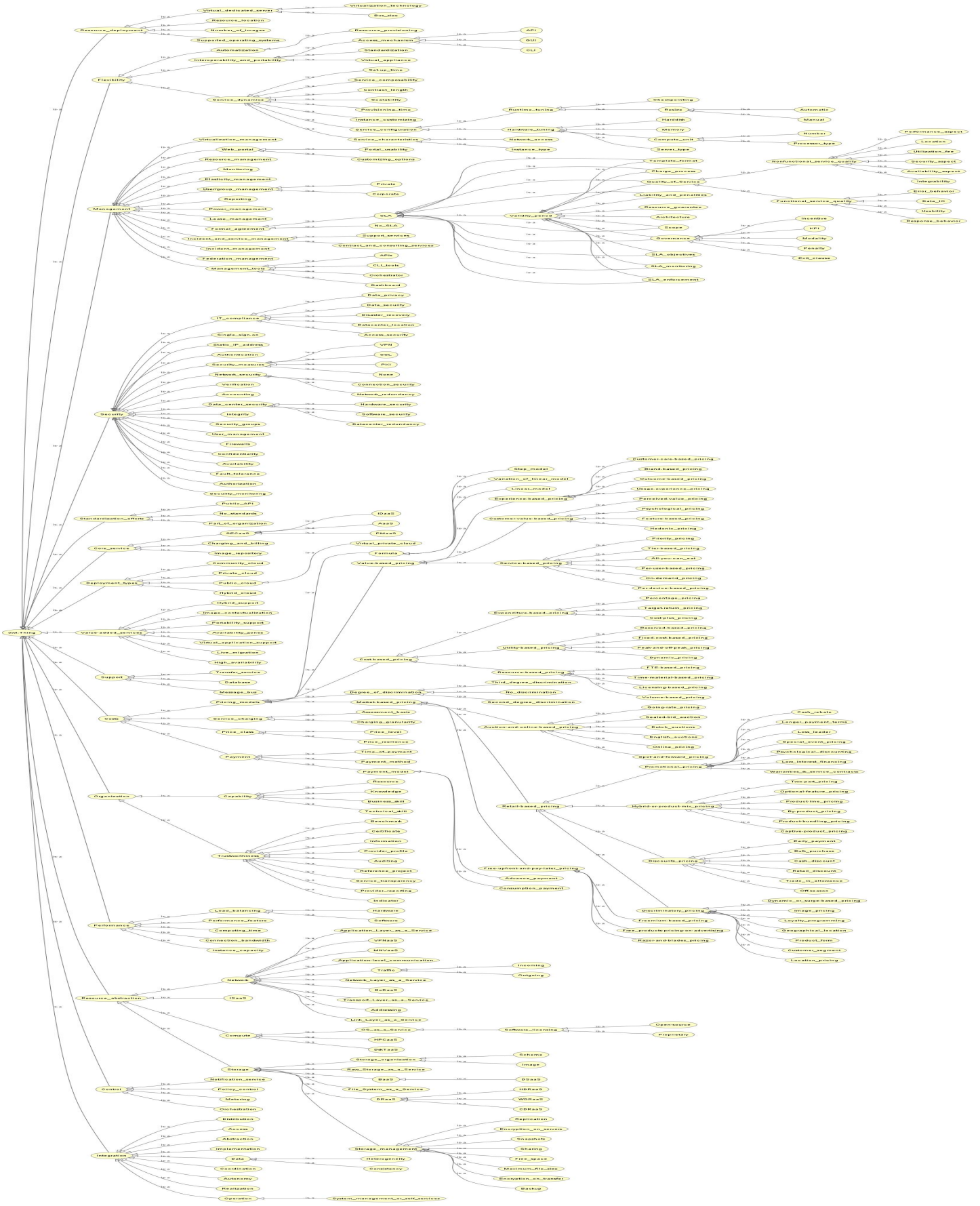


Figure 3: Proposed ontology's complete class hierarchy

(of SLA measures and SLA requirement management for different IaaSs; For example, Web Service Agreement Specification (WSA) is the standard for the SLA management architecture in Web service environments), Template format (used to electronically represent SLA for automated management)}, Monitoring, Reporting, Lease management, Incident and service management = {Contract and consulting services, Support services}, Web portal = {Portal_usability, Customizing options}, Virtualization management, Resource management, User/group management = {Corporate, Private}, Incident management, Power management}.

- Value-added services = {Availability zones, High availability, Hybrid support (facilitating the implementation of hybrid cloud by resource extension to external), Live migration, Portability support, Image contextualization (enabling virtual machine (VM) instance to be deployed in the form of a shared customized image for specific context such as VM with a turnkey database), Virtual application support (i.e., containers consisting of several VMs and allowing design and configuration of multi-tier applications)}.
- Resource deployment = {Virtual dedicated server = {Virtualization technology, Bus size (or processor register size e.g. 64 bits)}, Number of images, Resource location, Supported operating systems (OS supported by providers)}.
- Control provides cloud systems with basic control features. Its subclasses are Metering, Policy control, Notification service, and Orchestration.
- Standardization efforts = {No standards, Public API (i.e., common API enabling interoperability and customization), Part of organization (i.e. organization involves in public standardization)}.
- Flexibility = {Interoperability and portability = {Standardization, Access mechanism = {API, GUI, CLI}, Virtual appliance (delivering a service as a complete software stack installed on a VM)}, Automatization = {Resource provisioning}, Service dynamics = {Provisioning time, Contract length, Set-up time, Scalability, Instance customizing, Service composability, Service characteristics = {Instance type, Network access}, Service configuration = {Hardware tuning = {Compute unit = {Number, Processor type}, Server type, Memory, Harddisk}, Runtime tuning = {Resize = {Automatic, Manual}, Checkpointing}}.
- Deployment types = {Public cloud = {Virtual private cloud}, Private cloud, Hybrid cloud, Community cloud}.
- Integration = {Distribution, Coordination, Implementation, Access, Abstraction, Autonomy, Operation = {System management or self-services}, Realization, Data = {Heterogeneity, Consistency}}.
- Organization = {Trustworthiness = {Certificate, Reference project, Indicator, Benchmark, Provider reporting, Information, Provider profile, Service transparency, Auditing}, Capability = {Resource, Knowledge, Technical skill, Business skill}}.

The properties of the classes are hasDirectCost, hasMgmtAspect, hasSecurityAspect, and useTechnology as depicted in Figure 4.

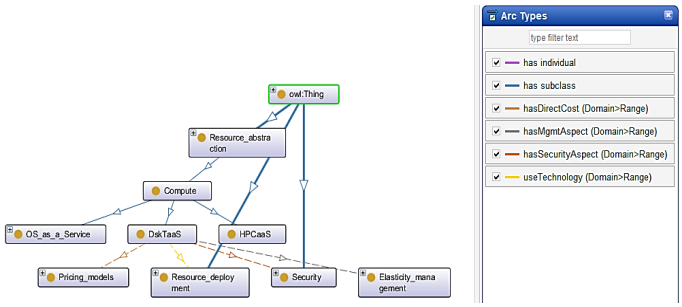


Figure 4: Ontology's class properties

4. Evaluation

The proposed ontology is evaluated into two crucial parts, a technology-focused evaluation and a user-focused evaluation, as follows.

For technology-focused evaluation, this paper employed a HermiT reasoner to rigorously determine whether the proposed ontology is syntactically and semantically consistent and identify subsumption relationships between classes. The result of this evaluation is an inferred ontology that is prominently displayed in yellow background in Figure 5 without any error (which will be indicated by red text if there is any).

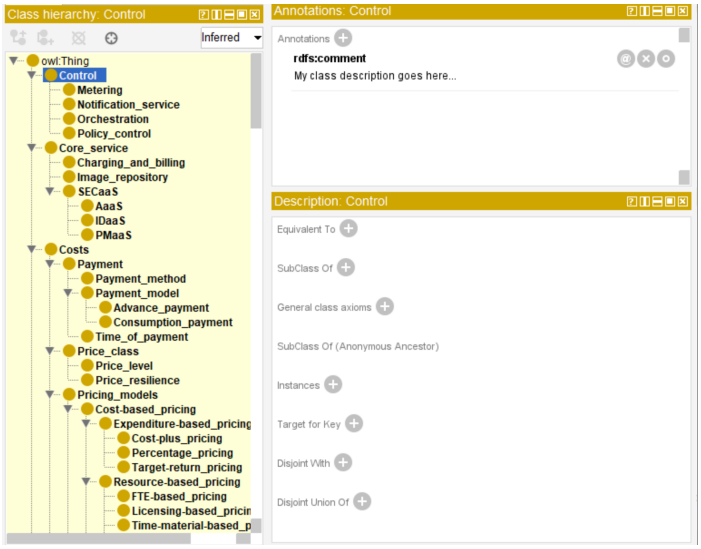


Figure 5: Inferred ontology by using HermiT

Complementing the technological assessment, the study conducted a comprehensive user-focused evaluation. This evaluation entailed applying our ontology to assess a worldwide recognized Infrastructure as a Service (IaaS) platform, namely, AWS EC2. To accomplish this, the study carefully generated ontology individuals by drawing from publicly available information pertaining to AWS EC2, as exemplified in Figure 6, where these individuals are denoted by prefix violet diamond symbols. All pieces of information regarding AWS EC2 offerings for consumers can be seamlessly mapped into the ontology's

existing classes (denoted by prefix orange oval symbols) to facilitate comprehensive IaaS selection.

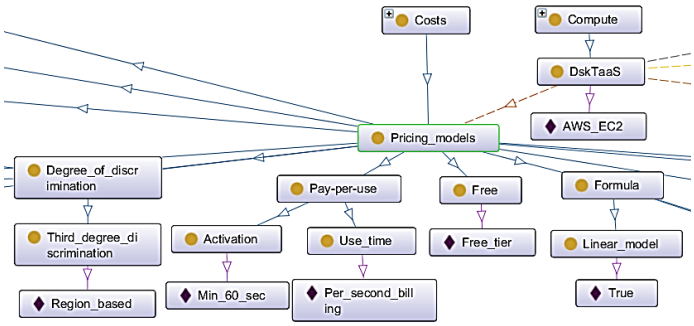


Figure 6: Ontology's instance portion for AWS EC2's offerings

Furthermore, in Figure 7, the study performs two distinct SPARQL queries against the ontology. The first query below

```
SELECT ?subject WHERE {
    ?subject rdf:type owl:Class.
}
```

aims to list all (sub)classes. Part of the resulting classes of the query are listed in the bottom pane in the figure. The second query below

```
SELECT ?subject WHERE {
    ?subject rdf:type owl:NamedIndividual.
}
```

identifies and enumerates individual instances as partially displayed in the resulting pane in the figure. The successful execution of both queries substantiates the potential for automating IaaS assessment processes using our ontology.

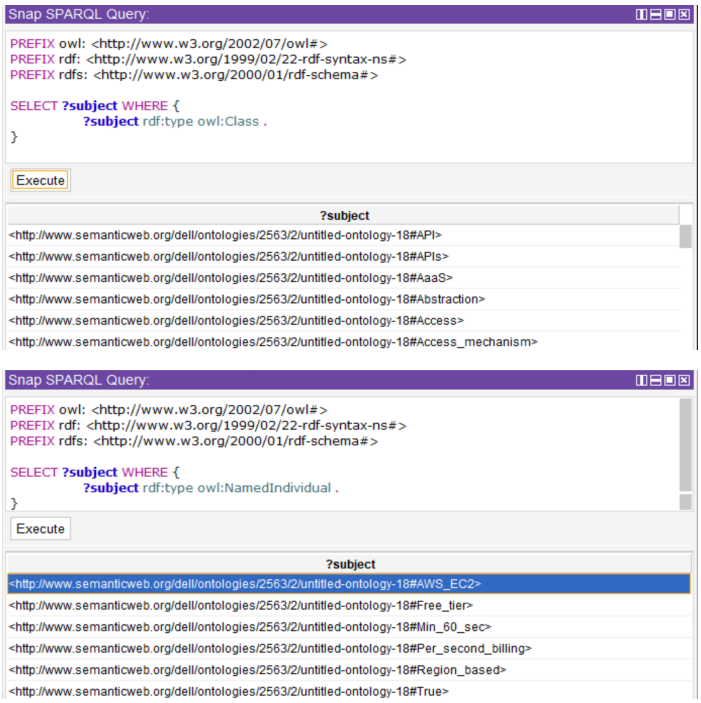


Figure 7: SPARQL queries for classes and individuals

5. Conclusion

This paper presents a novel customer-perspective IaaS ontology developed from various 18 IaaS taxonomies in present existence. This ontology stands out for its exceptional comprehensiveness, encompassing a total of 15 primary subclasses (e.g., performance, costs, and security) and 340 individual classes (e.g., instance capacity, availability, and price class). The evaluation shows that the proposed ontology is syntactically and semantically consistent. Furthermore, the ontology successfully enables not only the assessment of AWS EC2 IaaS but also SPARQL queries. This has affirmed that the ontology holds significant semantic value, offering utility not only to researchers but also to practitioners by leveraging it as a foundational component to develop a sophisticated assessment tools for facilitating effective IaaS adoption. The tool will be definitely helpful for IaaS customers, IaaS providers, and IaaS partners. The future work of this study is to develop an expert system in the form of SaaS to facilitate IaaS selection based on the proposed ontology.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research was financially supported by Department of Computer Science, Kasetsart University, Krung Thep Maha Nakhon, Thailand.

References

[1] ISO/IEC, ISO/IEC 17789: Information technology — Cloud computing — Reference Architecture, ISO/IEC, 2014.

[2] F. Nadeem, "Evaluating and Ranking Cloud IaaS, PaaS and SaaS Models Based on Functional and Non-Functional Key Performance Indicators," IEEE Access, **10**, 63245-63257, 2022, doi: 10.1109/ACCESS.2022.3182688.

[3] E. Al-Masri, L. Meng, "A Quality-Driven Recommender System for IaaS Cloud Services," in 2018 IEEE International Conference on Big Data (Big Data), 5288-5290, 2018, doi: 10.1109/BigData.2018.8622017.

[4] S. M. M. Fattah, A. Bouguettaya, S. Mistry, "Signature-based Selection of IaaS Cloud Services," in 2020 IEEE International Conference on Web Services (ICWS), 50-57, 2020, doi: 10.1109/ICWS49710.2020.00014.

[5] K. Kritikos, K. Magoutis, D. Plexousakis, "Towards Knowledge-Based Assisted IaaS Selection," in 2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), 431-439, 2016, doi: 10.1109/CloudCom.2016.0073.

[6] F. Moscato, R. Aversa, B. Di Martino, T. -F. Fortiş, V. Munteanu, "An analysis of mOSAIC ontology for Cloud resources annotation," in 2011 Federated Conference on Computer Science and Information Systems (FedCSIS), 973-980, 2011.

[7] A. Abdullah, S.M. Shamsuddin, F.E. Eassa, Ontology-based Cloud Services Representation. Research Journal of Applied Sciences, Engineering and Technology. **8**(1), 83-94, 2014, doi:10.19026/rjaset.8.944.

[8] M. Zhang, R. Ranjan, A. Haller, D. Georgakopoulos, M. Menzel, S. Nepal, "An ontology-based system for Cloud infrastructure services' discovery," in 8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 524-530, 2012.

[9] Q. Zhang, A. Haller, Q. Wang, CoCoOn: Cloud Computing Ontology for IaaS Price and Performance Comparison. in International Semantic Web Conference. 325-341, 2019, doi.org/10.1007/978-3-030-30796-7_21.

[10] T. Banditwattanawong, "The survey of infrastructure-as-a-service taxonomies from consumer perspective," in the 10th International Conference on e-Business, 2015.

[11] Z. Zhang, C. Wu, D.W.L. Cheung, "A survey on cloud interoperability: taxonomies, standards, and practice," ACM SIGMETRICS Performance Evaluation Review, **40**(4), 13-22, 2013, doi.org/10.1145/2479942.2479945.

[12] M. Firdhous, "A Comprehensive Taxonomy for the Infrastructure as a Service in Cloud Computing," in 2014 Fourth International Conference on Advances in Computing and Communications, 2014, 158-161, doi: 10.1109/ICACC.2014.45.

[13] C.N. Höfer, G. Karagiannis, "Cloud computing services: taxonomy and comparison," Journal of Internet Services and Applications, **2**, 81-94, 2011, doi.org/10.1007/s13174-011-0027-x.

[14] R. Dukarić, J. Matjaz, "A Taxonomy and Survey of Infrastructure-as-a-Service Systems," Lecture Notes on Information Theory, **1**, 29-33, 2013, doi: 10.12720/lnit.1.1.29-33.

[15] C. Wu, R. Buyya, K. Ramamohanarao, "Cloud Pricing Models: Taxonomy, Survey, and Interdisciplinary Challenges," ACM Computing Survey, 52(6), 1-36, 2019, doi.org/10.1145/3342103.

[16] Shivani and A. Singh, "Taxonomy of SLA violation minimization techniques in cloud computing," in Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 1845-1850, 2018, doi:10.1109/ICICCT.2018.8473230.

[17] T. Banditwattanawong, M. Masdisornchote, "Federated Taxonomies Toward Infrastructure-as-a-Service Evaluation," International Journal of Management and Applied Science, **2**(8), 33-38, 2016.

[18] S. Staab, R. Studer, Handbook on Ontologies (2nd. ed.). Springer Publishing Company, Inc., 2009.

[19] R. Dukaric, M.B. Juric, "Towards a unified taxonomy and architecture of cloud frameworks," Future Generation Computer System, **29**(5), 1196–1210, 2013, doi.org/10.1016/j.future.2012.09.006..

[20] S. Kansal, G. Singh, H. Kumar, S. Kaushal, "Pricing models in cloud computing," in International Conference on Information and Communication Technology for Competitive Strategies, 33:1–33:5, 2014, doi.org/10.1145/2677855.2677888.

[21] M.K.M. Murthy, H.A. Sanjay, J.P. Ashwini, "Pricing models and pricing schemes of iaas providers: A comparison study," in Proceedings of the International Conference on Advances in Computing, Communications and Informatics, 143–147, 2012, doi:10.1145/2345396.2345421.

[22] S. Kächele, C. Spann, F. J. Hauck, J. Domaschka, "Beyond IaaS and PaaS: An Extended Cloud Taxonomy for Computation, Storage and Networking," in IEEE/ACM 6th International Conference on Utility and Cloud Computing, 75-82, 2013, doi: 10.1109/UCC.2013.28.

[23] J. Repschlaeger, S. Wind, R. Zarnekow, K. Turowski, "A Reference Guide to Cloud Computing Dimensions: Infrastructure as a Service Classification Framework," in The 45th Hawaii International Conference on System Sciences, 2178-2188, 2012, doi: 10.1109/HICSS.2012.76.

[24] B.P. Rimal, E. Choi, I. Lumb, "A Taxonomy and Survey of Cloud Computing Systems," in The Fifth International Joint Conference on INC, IMS and IDC, 2009, 44-51, doi: 10.1109/NCM.2009.218.

[25] H. Kamal Idrissi, A. Kartit, M. El Marraki, "A taxonomy and survey of Cloud computing," in National Security Days (JNS3), 1-5, 2013, doi:10.1109/JNS3.2013.6595470.

[26] S. Gudenkauf, M. Josefiok, A. Göring, O. Norkus, "A Reference Architecture for Cloud Service Offers," in The 17th IEEE International Enterprise Distributed Object Computing Conference, 227-236, 2013, doi: 10.1109/EDOC.2013.33.

[27] R. Prodan, S. Ostermann, "A survey and taxonomy of infrastructure as a service and web hosting cloud providers," in The 10th IEEE/ACM International Conference on Grid Computing, 17-25, 2009, doi: 10.1109/GRID.2009.5353074.

[28] R. Teckelmann, C. Reich, A. Sulistio, "Mapping of Cloud Standards to the Taxonomy of Interoperability in IaaS," in IEEE Third International Conference on Cloud Computing Technology and Science, 522-526, 2011, doi:10.1109/CloudCom.2011.78.

[29] G. Laatikainen, A. Ojala, O. Mazhelis, "Cloud services pricing models," in Software Business. From Physical Products to Software Services and Solutions - 4th International Conference, 117–129, 2013, doi: 10.1007/978-3-642-39336-5_12.

[30] F. Polash, A. Abuhussein, S. Shiva, "A survey of cloud computing taxonomies: Rationale and overview," in The 9th International Conference for Internet Technology and Secured Transactions (ICITST-2014), 459-465, 2014, doi:10.1109/ICITST.2014.7038856.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 9, No. 1, 42-48 (2024)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Analysis of Emotions and Movements of Asian and European Facial Expressions

Ajla Kulaglic*,[1], Zeynep Örpek[2], Berk Kayı[3], Samet Ozmen[2]

[1]*College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait*

[2]*Research and Development Department, Vakif Katilim Bank, Istanbul 34000, Türkiye*

[3]*Data Engineering Department, Trendyol Group, Istanbul 34000, Türkiye*

ARTICLE INFO

Article history:
Received: 02 October, 2023
Revised: 01 January, 2024
Accepted: 01 January, 2024
Online: 20 January, 2024

Keywords:
Emotion recognition
Movement detection
Asian Facial Expression
European Facial Expression

ABSTRACT

The aim of this study is to develop an advanced framework that not only recognize the dominant facial emotion, but also contains modules for gesture recognition and text-to-speech recognition. Each module is meticulously designed and integrated into unified system. The implemented models have been revised, with the results presented through graphical representations, providing prevalent emotions in facial expressions, body language, and dominant speech/voice analysis. Current research, to identify the dominant facial emotion, involves two distinct approaches that autonomously determine the primary emotional label among seven fundamental emotions found in the input data: anger, disgust, happiness, fear, neutral expressions, sadness and surprise. The dataset utilized comprises over 292680 images sourced from the benchmark datasets FER-2013 and CK, enriched by images sourced from the Google search engine, along with 80 videos obtained during dedicated sessions, used for training and testing purposes. The Residual Masking Network (resmasknet) and CNN architectures are used as pre-trained models in this analysis. Resmasknet and CNN were chosen considering their superior performance compared to other algorithms found in the literature. The CNN architecture comprises 11 blocks, with each block containing a linear operator followed by ReLU or max-pooling layer. Starting with a convolutional layer that uses 32 filters and an 11x11x3 input, followed by a 3x3 max-pooling layer with a step of 2, the next layer includes a convolutional layer that uses 16 filters of size 9x9x16. The Residual Masking Network, contains four residual masking blocks operating on different feature sizes, each consisting a residual layer and masking block. The network initiates with a 3x3 convolution layer, followed by 2x2 max-pooling, effectively downsizing the image to 56x56. Successive transformations within four residual masking blocks generate different maps of sizes 56x56, 28x28, 14x14, and 7x7, culminating in an average pooling layer and a fully connected SoftMax output layer. The significance of this project lies in its focus on a comprehensive analysis of emotions and movements characteristic of Asian and European facial expressions. Showing promising accuracy rates, the proposed solutions achieve 75.2% accuracy for Asian and 86.6% for European individuals. This performance demonstrates the potential of this multidisciplinary framework in understanding and interpreting different facial expressions in different cultural settings.

1. Introduction

Emotional recognition (ER) enabled by artificial intelligence (AI) and machine learning (ML) models is pivotal for detecting and understanding emotions. This study introduces an integrated framework for facial emotion recognition for facial emotion recognition within the EUREKA project [1], exploring facial emotions, gestures, and social communication using AI platform. This paper is an extension of work originally presented in 2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET) [2]. Recognizing

*Corresponding Author: Ajla Kulaglic, ajla.kulaglic@aum.edu.kw

emotions plays a key role for effective human-AI interaction, as more than 65% of communication is non-verbal, while the remaining 35% is influenced by verbal and physical gestures [1]. Emotional states are represented as a discrete base (the main emotions), complex categories of expression or activation of facial action units (FAC), where each action unit (AU) corresponds to a certain facial muscle movement increased with emotions [3-5]. Video-based emotion recognition is a multidisciplinary area and covers various fields of psychology, affective computing and human-robot interaction. Audio features predict emotion status from audio within video. However, as text recognition can be unexpected and confusion because it refers to the context and the language, we focused only on the implementation and analysis of the dominant emotion in the facial expression, based on detected ethnic group of the person found in the video.

Advancements in real-time applications related to face detection, recognition, face enhancement, reconstruction and facial expression have evolved [6]. The lack of all these tasks can be seen in recognizing unknown faces within test datasets. Early approaches extract geometric features using FAC, but recent focus shifted to ML and deep learning (DL) models to better detect facial movements and expression indicating emotions with extraordinary accuracy [7-10]. For non-temporal features, InceptionNet [11] and DenseNet [12] form an essential part. However, for the tasks where the video is used as input, it is not enough to consider the features of the appearance of the images in the spatial dimension. For that purpose, the information about the movement in the video is also important. The spatial feature generated by pure Deep Neural Network (DNN) is not directly suitable for videos due to the lack of motion modeling. On the other hand, Recurrent Neural Network (RNN) provides an attractive framework for propagating information over a sequence using a continuous value hidden layer representation to capture temporal features [13]. It is efficient for sequence tasks, such as action recognition, speech recognition, natural language processing (NLP), etc. [14]. Long short-term memory (LSTM) [15] developed from RNN has shown its potential in time sequence analysis and is widely used in emotion recognition analysis [16]. In [17], the author combined Convolutional Neural Network (CNN) with RNN to model the spatio-temporal evolution of visual information. In [18], the authors used CNN and bidirectional RNN architecture to learn texture appearance models for video sequences. Unlike the composite CNN-RNN structure, the 3-dimensional CNN presented in [19] uses 3-dimension convolution operations instead of 2-dimensional convolution in the network structure generating multiple channels of information from adjacent video frames. Currently, the 3D-CNN is widely used in solving various tasks of video analysis. In [16], the researchers are also using 3D-CNN proposed hybrid network for encoding information about appearance and movement in face expressions and improved the results.

The wide application of machine learning and deep learning techniques in commercial, health and security environments has created engagement and higher morale among users with an increasing representation of trained models. In this study, the implemented software is intended for the Human Resource (HR) department. The proposed solution offers several benefits in the recruitment process analyzing facial expressions and body language during video interviews, providing insights into candidates' emotional intelligence, confidence and engagement level; offering a more comprehensive understanding of a candidate's suitability for a role. Using the AI approach the bias can be reduced evaluating candidates based on their expressions and responses rather than subjective impressions. It can be used as predictive performance indicator where certain emotions and expressions found in the video might correlate with job performance. For example, a high level of positivity might indicate strong motivation. On the other hand, automated emotion recognition tools can help in the efficient selection of a large number of candidates, prioritizing those who exhibit desired emotional traits. Understanding candidates' emotions during the requirement process enables custom interactions, addressing concern or tailor communication styles based on detected emotional cues, enhancing the overall candidate experience.

The implemented software is divided into five steps. In the first step, we set the performance (personal information) and competence goals. In step 2, the person is evaluated by completing a survey. A huge number of tests are available based on areas of interest and role-based competencies. In Step 3, you have to incorporate your social behavioral cues by taking a video to capture information about your emotions, gestures, and body language using a laptop or smartphone camera. This data is recorded, saved and then processed. In the fourth step, we explore and get real-time insights through the platform. It uses machine intelligence to identify the causes of human behavior. In the final step, the final results are presented. The user can get personalized real-time insight about the gestural articulation patterns and behaviors of your competency displayed on the dashboard.

The main contributions of the proposed work can be summarized as:

1. Ethnicity extraction from facial data based on the input data;

2.Machine learning-based determination of dominant emotions in video;

3. Validation of the proposed module within the Eureka360 project.

The rest of the paper is organized as follows: section 2 discusses emotional face recognition datasets used in this study and data preprocessing performed. Section 3 details implementation specifics - data augmentation, network architectures, parameter settings, and training methods. Section 4 presents experimental validations, with conclusions in Section 5.

2. Dataset and data preprocessing

2.1. Data

The study integrates publicly available datasets, namely the Facial Expression Recognition 2013 (FER-2013) [20] and the Extended Cohn-Kanade Dataset (CK+) [21], together with images and videos gathered during the project implementation. Both, FER-2013 and CK+ datasets contain labeled images categorized into seven emotions (0=Anger, 1=Disgust, 2=Fear, 3=Happiness, 4=Sadness, 5=Surprise, 6=Neutral).

CK+ consists of a total of 981 marked pictures of 123 people. The images are placed with a similar background containing only the front face (Figure 1). FER-2013 (Figure 2) dataset consisting

of 28709 labeled images in the training set, 7178 labeled images in the validation and test set together. The validation test contains 3589 images, and the test set consists of another 3589 images. The CK+ dataset does not contain a default test split. In this study, a 70-30 split was made, where 70% of CK+ images were used for model training, and 30% of images were used for testing. Happiness and surprise are the most represented feelings in the FER-2013 and the CK+ data set.

However, reference datasets mainly contain posed facial expressions captured in controlled environments, with consistent lighting and head positioning. Algorithms trained on such a dataset might underperform when faced with different preconditions, especially in natural settings. To address this limitation, additional data were collected from Google searches along with videos of project participants. A total of 291,770 Google images (Figure 3) and 80 videos were collected. The Google images were obtained by searching for expressions related to each emotion and their synonyms, expanding the dataset's diversity beyond the controlled environment. The videos were recorded during meetings and dedicated sessions to capture facial expressions in diverse environments. A total of 80 videos were collected, pre-processed and used in this project. This additional dataset encompasses various conditions, contributing to a more diverse dataset, especially in terms of ethnic representation, with a focus on Asian and European ethnicities.

Figure 1: Emotional Expression Samples: CK+ Dataset

Figure 2: Instances of Facial Expressions: FER-2013 Dataset

In this study, we utilized established benchmark image datasets. However, our input data for analysis encompassed videos rather than individual static images. Video data provides richer source of data, capturing changes of facial expressions that static images might not fully capture. The videos are converted into individual frames, where each frame represents image captured at a specific point in time thought the video. Analyzing these frames sequentially enables us to reconstruct the temporal facial expressions, capturing the subtle changes in emotions displayed through the video. Transforming videos into frames for analysis, we bridge the gap between the continuous nature of video data and the discrete analysis required to interpret facial expressions, allowing us to dive deeply into the temporal dynamics of emotional displays.

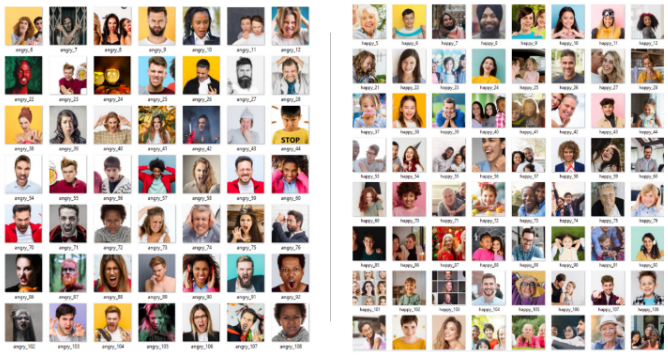


Figure 3: A collection of images obtained through a Google search

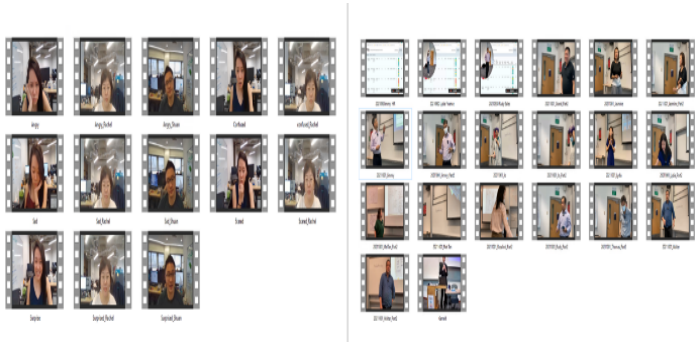


Figure 4: Instances of Collected Videos: Diverse Meetings and Special Sessions

2.2. Data preprocessing

The collected video data consists of many irregularities like changes in illumination, oclussions, perspective shifts, and scale variations. With these anomalies, it is difficult to accurately extract emotional information features from face images in videos. To address this challenge, our study aligns features using the detector to match the expected size of the front faces to the Region of Interest (ROI). We use a detector available in OpenCV based on ResNet-10, which has been shown to be a very efficient CNN model for face detection [22]. Utilizing the ResNet-10 detector, we preprocess video data by analysing it frame by frame. This process involves identifying faces withing bounding boxes, avoiding a separate face detection step. This preprocessing step enables speeding up the training process by isolating faces beforehand, eliminating the need for various data agumentation techniques such as histogram equalization, cropping, or image rotation. By aligning and padding faces withing the frames, we achieve standardized facial presentations. This alignment ensures consistency in facial sizes across varied video conditions, reducing the impact of anomalies like mentioned above, and improving the efficiency of the overall training phase.

3. Methodology

The project is designed to recognize the predominant emotion in a video, emphasizing ethnicity as a crucial factor. The input video is divided into 40 equal parts, and each of those parts have been used for analysis of dominant emotion. Additionally, it analyses signs such as eye movements, blinks, hand gestures and

facial touches. Following the ethnicity detection phase, the project employs two distinct neural network models for emotion recognition- one for identifying emotions within Asian ethnicity and the other trained for European ethnicity. These networks are fine-tuned using the training set to enhance their accuracy and performance. Figure 5 illustrates the proposed framework. Bellow, we provide an explanation for each implemented module.

3.1. Asian and European face recognition

The difference in facial expressions for various emotions between Asian and European ethnicities are result of a complex interaction between cultural, societal, and individual factors. Facial recognition algorithms detect the ethnicity by checking facial landmarks, shapes and features unique to different ethnicities. Ethnicity identification poses significant challenges, as evidenced by studies found in the literature, both for humans and computer vision algorithms, especially when restricted to facial attributes [23]. According to the studies and implemented libraries, in this project we use the DeepFace Python library [24] as a solution to estimate the ethnicity in the input data. DeepFace relies on a pre-trained Convolutional Neural Network (CNN) architecture that stands out in determining ethnicity from image or video frames. The mentioned library is trained across a spectrum of data including Asian, White, Middle Eastern, Native American, Latino, and Black races, DeepFace serves as a robust tool. Once the ethnicity is detected based on the input data, the system then selectively activates an emotion recognition model trained for that specific ethnicity. This approach facilitates the accurate assessment of emotions by ethnicity, as shown in Figure 6.

3.2. Dominant emotion recognition

Similar to the ethnicity classification, emotion classification recognizes emotions that are involved in understanding facial expressions, body language, and context. The models analyze facial expressions using feature extraction techniques to capture emotions, which include identifying movements in specific facial muscles, eye expressions associated with different emotions. Upon successfully identifying the ethnicity of the subjects in the video, the project engages different models for recognizing the primary emotion. Two distinctive pre-trained network architectures are used, DeepFace and Python Toolbox for Facial Expression Analysis (Py-Feat) [24, 25]. Within the DeepFace library, an advanced convolutional neural network (CNN) model for emotion recognition is employed (Figure 7). This model has been intentionally selected based on its superior performance compared to other algorithms found in literature. Emotion identification within images and collected videos is done by a CNN architecture consisting of 11 blocks. Each block contains a linear operator followed by at least one nonlinear layer like ReLU or max pooling. The initial phase takes an image input traversing through a convolutional layer with 32 filters, followed by an 11x11x3 and 3x3 max-pooling layer with a stride of 2. Subsequently, another convolutional layer with 16 filters sized at 9x9x16 is employed to extract low-level features from edges and textures within the input image. The subsequent three layers are locally linked layers with different types of filters based on their distinctive feature map. Finally, the last two layers consist of fully connected layers aimed to establish correlation between two distant parts of the face. The network's output is a SoftMax layer, with total of 120 million parameters in this architecture. Each convolutional layer is integrated with a ReLU activation function. Additionally, dropout is used in the initial fully connected layer, and L2-regularization is employed in the final stages [25,26,27].

In contrast, the Py-Feat library introduces the Residual Masking Network (resmasknet) [28]. This network consists of four primary resmasking blocks, each representing residual masking functionality. These blocks operate on diverse feature sizes, containing both residual layer and masking block. The input image will traverse through 3x3 convolutional layer with a stride of 2, followed by a 2x2 max pooling layer, effectively reducing the image dimensions to 56x56. The subsequent feature maps, derived after a 2x2 maximum shrinkage layer, undergo transformation via four residual masking blocks, generating four distinct maps sized at 56x56, 28x28, 14x14, and 7x7, respectively. The network ends with an average pooling layer and a fully connected layer integrated with SoftMax to generate the final output, as depicted in Figure 8.

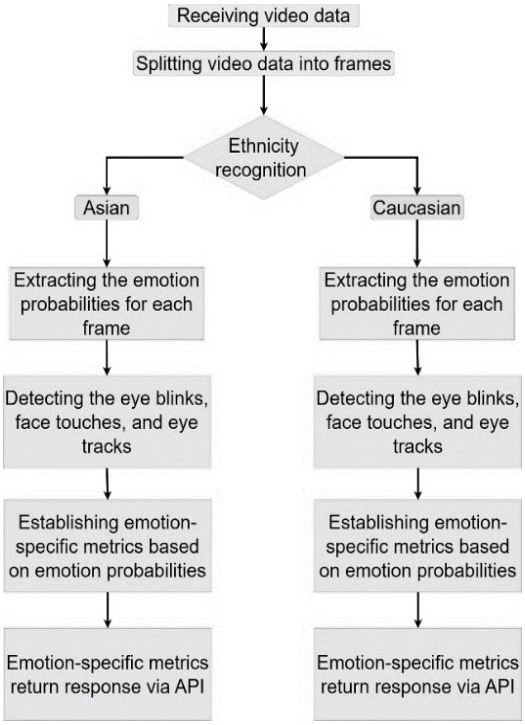


Figure 5: Emotion recognition framework for Asian and European facial expressions

Figure 6: Example of dominant race recognition

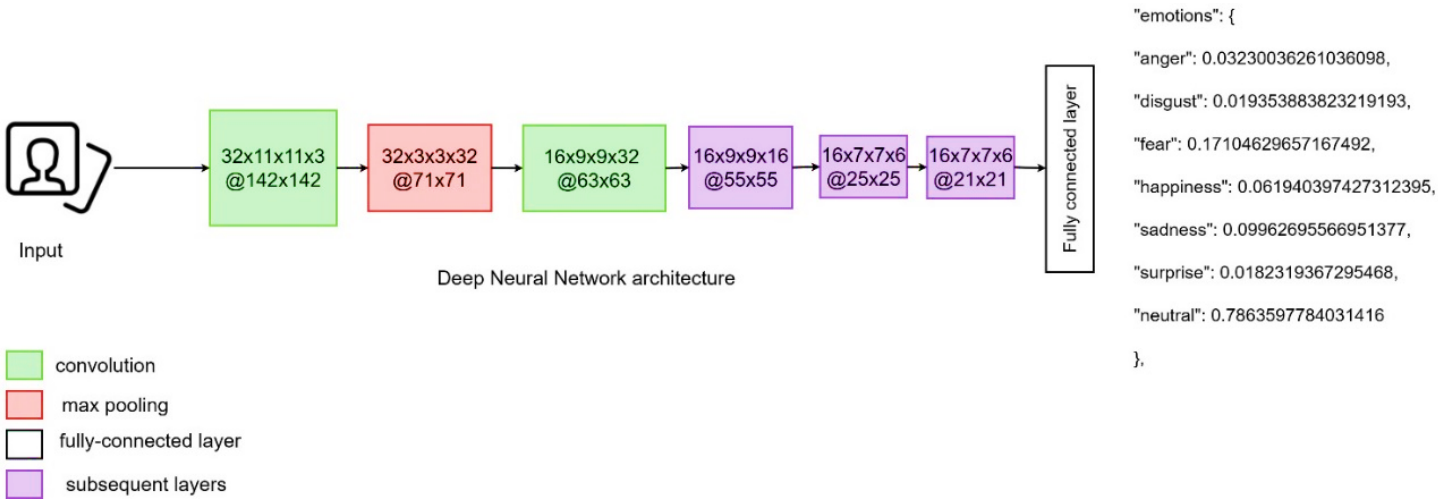


Figure 7: Schematic representation of DeepFace – Deep Neural Network architecture

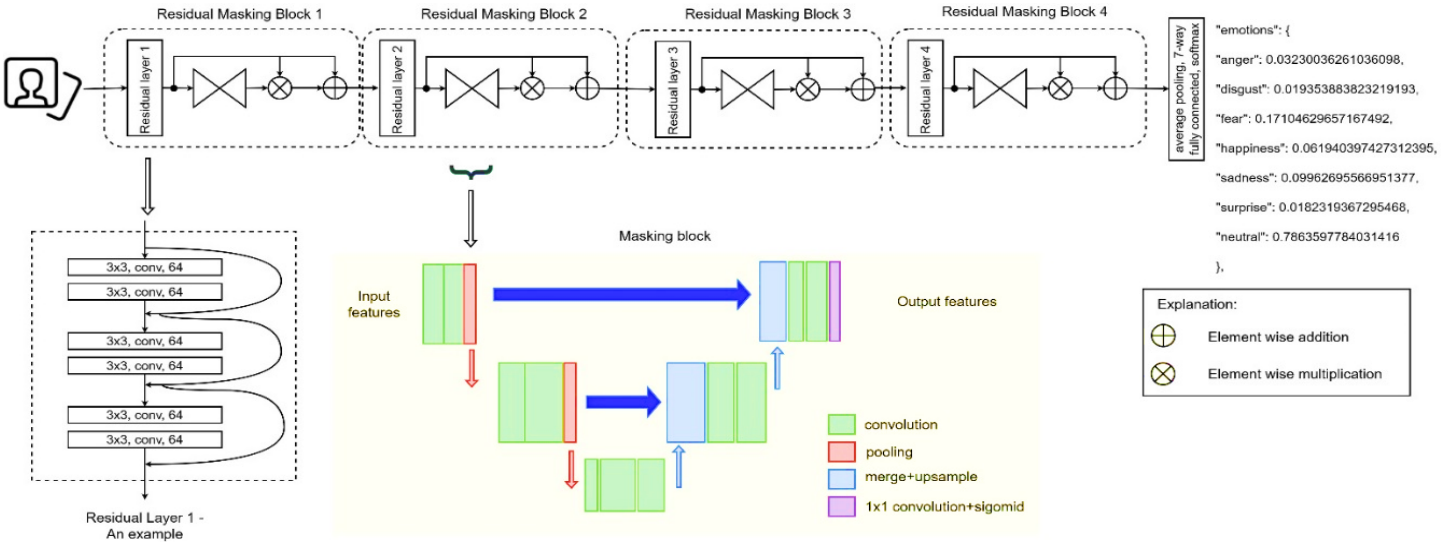


Figure 8: Schematic representation of Residual Masking Network (resmasknet) architecture

3.3. Eye tracking, blinking, and face touching detectors

Beyond emotion analysis, this project includes examination of eye movements, blinking frequencies, and facial touches. These non-verbal signs are considered expressive elements that contribute to a better understanding of emotional states. For instance, increased eye blink rates might signify nervousness or stress. Conversely, a reduction in blink frequency might indicate focus or intense concentration. Tracking the movement of eyes can offer valuable information about attention, interest, or cognitive processing related to emotions. People use their hands during conversations to further highlight some of the emotions. Touching face can indicate anxiety, discomfort, or even deception in some context. These non-verbal cues complement facial expressions and body language in emotion recognition. Integrating eye movements, blinking and facial touches into analysis enhance the accuracy and depth of understanding emotional states.


```
{'ErrorCode': 0,
 'datetime': '2022-06-27 12:58:21.638802',
 'emotions': {'average_eye_blink_time': 7,
              'average_of_emotion_results': {'anger': 0.02,
                                             'disgust': 0.02,
                                             'fear': 0.19,
                                             'happiness': 0.04,
                                             'neutral': 0.83,
                                             'sadness': 0.11,
                                             'surprise': 0.0},
              'average_of_face_touching': 2,
              'dominant_emotions_in_difference_time_intervals': {'First': 'neutral',
                                                                 'First-Mid': 'neutral',
                                                                 'Last': 'neutral',
                                                                 'Mid': 'neutral',
                                                                 'Mid-Last': 'neutral'},
              'longest_observed_emotions': 'neutral',
              'number_of_hard_emotions_transitions_by_minutes': 0.0,
              'number_of_observed_emotions_from_video': 7.0,
              'number_of_soft_emotions_transitions_by_minutes': 14.0,
              'number_of_transition_by_minutes': 14.0,
              'shortest_observed_emotions': 'surprise'},
 'ethnicity': 'white',
 'filename': '720p_Medium.mp4',
 'response_time': '50.89487981796265 second'}
```


Figure 9: Probability results of dominant emotions in the analyzed video

The project employs a fixed time window to count occurrences of blinks and facial touches. Moreover, the project extracts diverse features including blink duration, eyelid closure time, blink speed, and hand movement speed. It is important to note that the nearness of individuals to the camera significantly influences

the frequency of these observed events, resulting in higher event frequency for closer subjects.

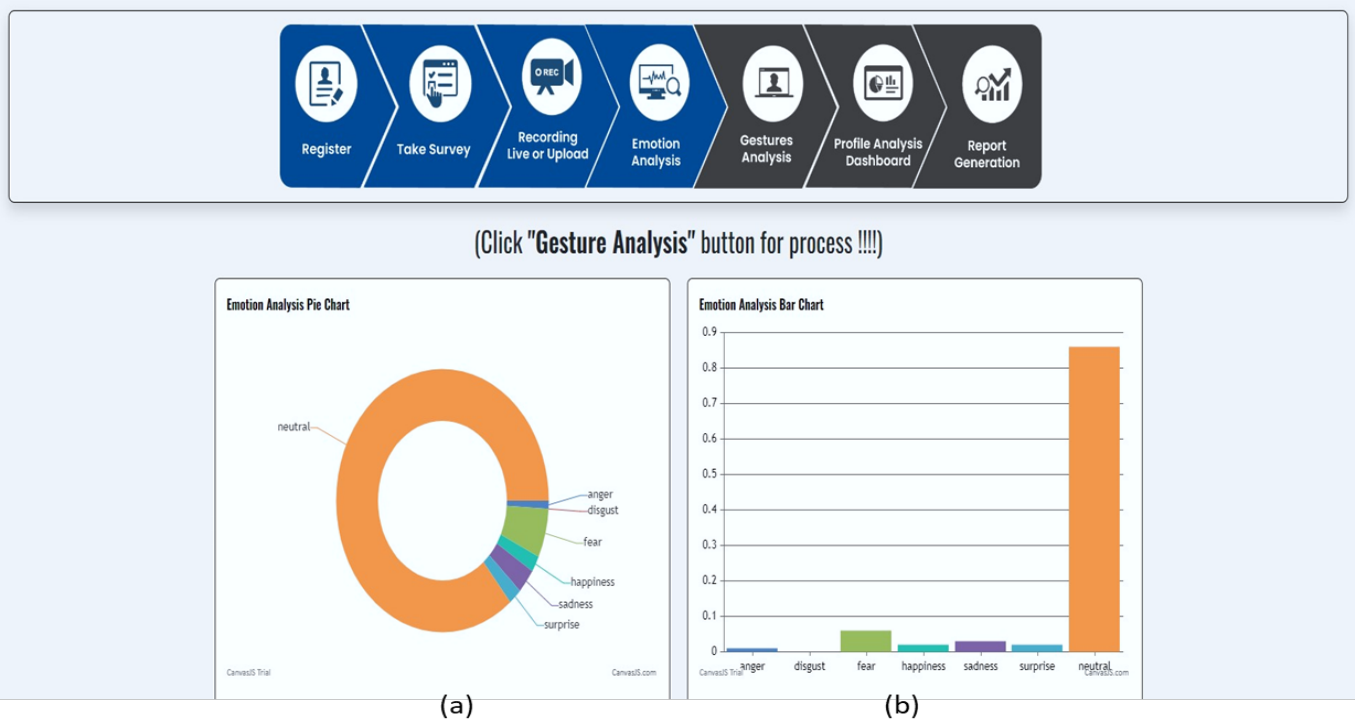


Figure 10: Visualized Result Representation. Emotion Analysis Pie Chart (a) represents the results in form of pie chart. Emotion Analysis Bar Chart (b) shows seven emotions (0=Anger, 1=Disgust, 2=Fear, 3=Happiness, 4=Sadness, 5=Surprise, 6=Neutral) and the numerical value (probability) of each emotion detected in the video (y - axis).

4. Results

The Eureka Framework comprises several components, including registration, surveys/tests, video recording or uploading, emotional analysis, gesture analysis, a profile analysis dashboard, and a comprehensive report derived from the collected and analyzed data. This paper focuses exclusively on outcomes related to emotional analysis. The initial phase involves preprocessing the collected data and utilizing pre-trained models. Each data set used was analyzed and prepared separately, and at the end all data were used together for training and testing purposes. The reference data sets are recorded in controlled environments, while the collected Google images and videos contain different environments and conditions, which contributes to the diversity of the data set used in addition to ethnic diversity. The 80 videos collected and preprocessed, converting them into data frames. The first step is ethnicity detection within the input frames. A model is deployed to ascertain whether the individual in the input data is Asian or European person. Subsequently, depending on the ethnicity detected, specific models intended for different ethnic groups - Asian or European – are invoked. The accuracy of each model is separately calculated. Based on the input data in this work, the accuracy for Asian individuals stands at 75.2%, while for European individuals it reaches 86.6%. One reason for this may be due to the dataset's imbalance, where significantly more data for European individuals was available for training the models. Notably, the previously trained models from Py-Feat and DeepFace libraries, undergo tuning, involving adjustments in periods and parameters (learning rates, activation functions, etc.) to optimize their accuracy. The results for video recorded on June 27, 2022 are shown in Figures 9 and 10. The person appearing in the video is identified as European, or white. Moreover, the analysis detected 7 instances of eye blinks, 2 facial touches and a dominant neutral emotion. The overall emotion analysis, based on seven emotions (0=Anger, 1=Disgust, 2=Fear, 3=Happiness, 4=Sadness, 5=Surprise, 6=Neutral), reveals on 83% of neutrality in the video. The video segments are analyzed separately, where the dominant emotions are calculated within each segment (first, first-middle, middle, middle-last, and last). Within segments the neutral emotion was detected. The results are graphically presented in the Figure 10, offering a view of dominant emotion through the analyzed video.

5. Conclusion

The presented study demonstrates the implementation of a dominant emotion recognition model applicable across diverse sectors, notably crucial in HR for candidate assessment. Through questionnaires, competency-related tests and recorded or uploaded video introductions, an analysis based on collected data offers insights into candidate suitability. The proposed project empowers individuals to refine their skills and advance in their careers, simultaneously enabling corporations to maximize their investment in employees while exploring new opportunities. The main drawback of the above methods lies in their reliance on detecting overt facial movements, however, individuals might intentionally try to mask their emotions.

Employing AI in candidate interviews ensures unbiased selection. Leveraging unbiased datasets and interview videos,

well-trained models exhibit remarkable precision in detecting dominant emotions, eye blinks, facial touches, and eye tracking. Training models with varied datasets and interview videos resulted in high accuracy in recognizing emotions. This study highlights the model's efficacy not only in research but also in practical applications across various fields.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This study was implemented thanks to the support provided by Vakıf Katılım Bank R&D Center. The opportunities offered by Vakıf Katılım R&D Center have added a huge impact on our scientific work. On this occasion, we would like to express our sincere gratitude to Vakıf Katılım Bank R&D Center for their valuable support.

References

[1] https://vimeo.com/775825642, accessed:28.12.2022
[2] B. Kayı, Z. Erbaşı, S. Özmen and A. Kulaglic, "Emotion and Movement Analysis Study from Asian and European Facial Expressions," 2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Mohammedia, Morocco, 2023, 1-5, doi: 10.1109/IRASET57153.2023.10152995.
[3] DeepFace - Most Popular Deep Face Recognition in 2022 (Guide) - viso.ai, accessed:19.12.2022
[4] Y. Taigman, M., Yang, M.A, Ranzato, & L. Wolf, L.," Deepface: Closing the gap to human-level performance in face verification", In Proceedings of the IEEE conference on computer vision and pattern recognition, 1701-1708, 2014, doi:10.1109/cvpr.2014.220
[5] G. E. Dahl, T. N. Sainath, and G. E. Hinton. "Improving deep neural networks for LVCSR using rectified linear units and dropout". In ICASSP, 2013., doi:10.1109/ICASSP.2013.6639346
[6] P. M. Ashok Kumar, Jeevan Babu Maddala & K. Martin Sagayam, "Enhanced Facial Emotion Recognition by Optimal Descriptor Selection with Neural Network", IETE Journal of Research, 2021, DOI: 10.1080/03772063.2021.1902868
[7] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," in IEEE Transactions on Affective Computing, **13**(3), 1195-1215, 1 July-Sept. 2022, doi: 10.1109/TAFFC.2020.2981446.
[8] D. Kollias, M. A., Nicolaou, I., Kotsia, G., Zhao, & S. Zafeiriou, "Recognition of affect in the wild using deep neural networks", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (26-33), 2017, DOI: 10.1109/CVPRW.2017.247
[9] H., Yang, U., Ciftci, & L., Yin, "Facial expression recognition by de-expression residue learning". In Proceedings of the IEEE conference on computer vision and pattern recognition (2168-2177)., 2018, DOI: 10.1109/CVPR.2018.00231
[10] S., Xie, & H. Hu," Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks". IEEE Transactions on Multimedia, **21**(1), 211-220., 2018, DOI: 10.1109/TMM.2018.2844085
[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, 2818-2826, doi: 10.1109/CVPR.2016.308.
[12] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, 2261-2269, doi: 10.1109/CVPR.2017.243.
[13] P. Scovanner, S. Ali, and M. Shah. "A 3-dimensional sift descriptor and its application to action recognition", In Proceedings of the 15th ACM international conference on Multimedia (MM '07). Association for Computing Machinery, New York, NY, USA, 357–360., 2007 https://doi.org/10.1145/1291233.1291311
[14] T. Mikolov, M. Karafiát, L. Burget, J. H. Černocký and S. Khudanpur. "Recurrent neural network-based language model." Interspeech (2010), doi: 10.21437/Interspeech.2010-343
[15] S. Hochreiter and J. Schmidhuber. "Long Short-Term Memory". Neural Comput. **9**, 8 (November 15, 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
[16] Y. Fan, X. Lu, D. Li, and Y. Liu. "Video-based emotion recognition using CNN-RNN and C3D hybrid networks". In Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16). Association for Computing Machinery, New York, NY, USA, 445–450. 2016, https://doi.org/10.1145/2993148.2997632
[17] S. E. Kahou, V. Michalski, K. Konda, R. Memisevic, and C. Pal. "Recurrent Neural Networks for Emotion Recognition in Video". In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15). Association for Computing Machinery, New York, NY, USA, 467–474. 2015. https://doi.org/10.1145/2818346.2830596
[18] J. Yan, W. Zheng, Z. Cui, C. Tang, T. Zhang, Y. Zong, and N. Sun. "Multi-clue fusion for emotion recognition in the wild". In Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16). Association for Computing Machinery, New York, NY, USA, 458–463. 2016, https://doi.org/10.1145/2993148.2997630
[19] D. Tran, L. Bourdev, R. Fergus, L. Torresani and M. Paluri, "Learning Spatiotemporal Features with 3D Convolutional Networks," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, 4489-4497, doi: 10.1109/ICCV.2015.510.
[20] FER2013 Dataset | Papers With Code, accessed:24.09.2023
[21] CK+ Dataset | Papers With Code, accessed:24.09.2023
[22] H., Kaushik, T., Kumar, & K. Bhalla, K. "iSecureHome: A deep fusion framework for surveillance of smart homes using real-time emotion recognition". Applied Soft Computing, 122, 108788. 2022. https://doi.org/10.1016/j.asoc.2022.108788.
[23] S. Fu, H. He and Z. -G. Hou, "Learning Race from Face: A Survey," in IEEE Transactions on Pattern Analysis and Machine Intelligence, **36**(12), 2483-2509, 1 Dec. 2014, doi: 10.1109/TPAMI.2014.2321570.
[24] https://viso.ai/computer-vision/deepface/, accessed:25.09.2023
[25] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, 1701-1708, doi: 10.1109/CVPR.2014.220.
[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet classification with deep convolutional neural networks". In Proceedings of the 25th International Conference on Neural Information Processing Systems - **1** (NIPS'12). Curran Associates Inc., Red Hook, NY, USA, 1097–1105. 2012. DOI: 10.1109/ACPR.2015.7486599
[27] G. E. Dahl, T. N. Sainath and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 2013, 8609-8613, doi: 10.1109/ICASSP.2013.6639346.
[28] L. Pham, T. H. Vu and T. A. Tran, "Facial Expression Recognition Using Residual Masking Network," 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021, 4513-4519, doi: 10.1109/ICPR48806.2021.9411919.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 6, 46-56 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

EEG Feature Extraction based on Fast Fourier Transform and Wavelet Analysis for Classification of Mental Stress Levels using Machine Learning

Ng Kah Kit[1,*], Hafeez Ullah Amin[1], Kher Hui Ng[1], Jessica Price[1], Ahmad Rauf Subhani[2]

[1]*School of Computer Science, University of Nottingham Malaysia, 43500 Semenyih, Selangor*

[2]*College of Electrical and Mechanical Engineering, National University of Science and Technology, Islamabad, Pakistan*

ARTICLE INFO

Article history:
Received:24 July, 2023
Accepted:05 November, 2023
Online: 30 November, 2023

Keywords:
Electroencephalography (EEG)
Feature Extraction
Mental Stress
Discrete Wavelet Transform
Fourier Transform
Machine Learning

ABSTRACT

Mental stress assessment remains riddled with biases caused by subjective reports and individual differences across societal backgrounds. To objectively determine the presence or absence of mental stress, there is a need to move away from the traditional subjective methods of self-report questionnaires and interviews. Previously, it has been evidence that EEG Oscillations can discriminate mental states, for instance, stressed and non-stressed. However, it is still not clear in which range of EEG oscillations the neural activities are associated with the mental states. This paper presents a wavelet-based EEG feature extraction method for the classification of mental stress using machine learning classifiers. An EEG dataset of 22 participants was used to test the performance of the proposed wavelet-based feature extraction method. The dataset includes both stress and control conditions, and the stress condition has multiple levels of stress, starting from low, mild, and high stress. The Daubechies mother wavelet of the fourth order was used to separate the EEG oscillations into 7 levels for the extraction of the absolute powers. Whereas Fast Fourier Transform were implemented to obtain the average power of the oscillations. The features were then used in support vector machine, decision tree, linear discriminant analysis and artificial neural network classifiers. A comparison between the classifiers using average power, absolute power, and a combination of both is provided. The EEG alpha, theta, and beta frequency bands showed promising results for the classification of mental stress vs. control conditions by achieving an average accuracy of 95% using the decision tree. The results of the proposed method suggest the potential use of wavelet analysis for mental stress detection despite FFT performing better. The proposed method has the potential to be used in Computer-Aided Diagnosis (CAD) systems for mental stress assessment in the future alongside the discovery of significant wave bands in relation to mental stress detection.

1. Introduction

The Latin verb ‘strictus’, which merely means to draw tight, is whence the word "stress" gets its original meaning. The term "stress" didn't have psychological connotations until the late 19th century, thanks to the groundbreaking work of Hans Selye, who is regarded as the father of stress study. Before then, stress was always thought of as the act of applying physical pressure or force to an object. However, we also experience internal emotional pressures and invisible forces, which has led to a biological investigation into the causes of stress. This was illustrated in the work of W.B. Cannon, who detailed how biological systems have developed an internal system to preserve homeostasis, a stable internal state. In order to test his views concerning acute stress responses as opposed to chronic stress, Hans Selye carried out experiments. He then came up with the phrase "stress responses," defining stress as pressures or mutual actions that occur across any part of the body, whether it be psychological or physical.

The broad term for stress can be further broken down into two categories: positive stress (Eustress) and negative stress (Distress). Eustress is perceived as a type of pressure that encourages a person to overcome obstacles by learning to see outside pressures as challenges rather than obstacles. On the other hand, distress results from a failure to use these needs as a motivator, which eventually stifles any advancement or success. Distress can also be divided

*Corresponding Author: Ng Kah Kit, kahkitng@gmail.com

into two categories: acute distress, which is short-term, and chronic distress, which is long-term. It should be highlighted that current definitions and understanding suggest that stress results in a physical reaction as a stressor rather than a physical reaction to perceived threats or challenges.

Both adults and children are afflicted by mental stress. Every human being suffers stress at some point in their lives, whether it is from work-related homework or just plain peer pressure from their employer. Short-term stress may be good for encouraging the improvement of work performed, while long-term stress can be destructive to one's physical and mental health [1]. If untreated or without an appropriate management strategy, it can seriously damage cognitive abilities and, as a result, the person's quality of life [2]. Studies have also revealed that prolonged work stress diminishes the grey matter volumes of the dorsolateral prefrontal cortex and the anterior cingulate cortex, which are accountable for memory, attention, and mood [3].

Despite advances in medicine, particularly in psychology, have made it possible for regular people to get the care or assistance they want. Patients only seek treatment when they can no longer tolerate to live in a situation of extended stress, therefore there is still much to learn and develop in this area. As a result, the general public continues to be untreated and lacks access to expert assistance to lessen future stress. Even worse, because conventional techniques for detecting mental stress primarily depend on self-report questionnaires and interviews, the results are still largely ambiguous. Therefore, subjective interpretations may be used to assess the degree to which a patient's stress level can be deemed harmful as well as whether or not the patient is experiencing stress. Therefore, the ability to anticipate a patient's likelihood of experiencing stress in the future or even just identify stress without consulting a medical practitioner may make it possible for patients to receive treatment and possibly even seek it out more voluntarily.

Many attempts have been made up to this point to use a machine learning technique to swiftly evaluate and forecast a patient's state of health. Results in certain instances point to machine learning's potential to diagnose conditions more accurately than qualified medical professionals. But in many of these instances, there are numerous flaws and biases, which contribute to the widespread belief that machine learning can supplement, if not completely replace, medical professionals [4]. Nevertheless, we think that machine learning applications in the healthcare industry will only grow. Medical practitioners should then use artificial intelligence and machine learning as a tool to give patients better and more advanced medical care.

By placing tiny electrodes on the scalp, an electroencephalogram (EEG) is a non-invasive, affordable, easily accessible, and painless test that looks for irregularities in brain waves. The potential difference between the cortical neuronal activity and the electrodes' detection of it is amplified and shown as a waveform. The cortical excitatory and inhibitory postsynaptic potential summations serve as the primary sources of electrographic activity [5]. Additionally, EEG scans reflect changes in brain activity almost instantaneously due to their high temporal resolution, while other scan types require several minutes following the occurrence of an event. Unfortunately, the limited spatial resolution of EEG makes it impossible to pinpoint the precise location of the cerebral waveforms. Furthermore, there will be significant contamination from other electrical noise as a result of the potential difference's amplification. Even though EEG signals are fascinating, they cannot be used by an interpreter to make future predictions. It's possible that any waveform anomalies that have happened before won't happen again.

Based on their frequency range, the waveforms found in an EEG can be grouped. The frequencies at which delta activity, theta activity, alpha activity, beta activity, and gamma activity occur are as follows in increasing order: delta activity occurs between 0.5 and 4 Hz, theta activity between 5 and 7 Hz, alpha activity between 8 and 13 Hz, beta activity between 14 and 30 Hz, and gamma activity between 30 and 80 Hz. Each of these waveforms, in turn, represents distinct brain states, including relaxation, sleep, anxiety, passive attention, and concentration [6]. EEG signal information is a well-known neuroimaging modality that records brain electrical activity for the diagnosis of various brain abnormalities, such as the identification of epileptic seizure activity, depression, stroke, and Alzheimer's disease [5]. Despite the vast amount of data that can be gleaned from an EEG reading, little research has been done on using it to identify mental stress. Furthermore, the research that has already been done does not consistently point to a general strategy for methodically combining machine learning with EEG for stress assessment [7]. This was demonstrated when various machine learning classifiers and feature extraction techniques were used in comparable circumstances.

Since it has been shown in numerous studies that the alpha, theta, and beta bands of an EEG reading correlate to cognitive workload processing and, in turn, mental stress, here we propose to extract information from these bands in order to classify mental stress [8,9]. Therefore, the purpose of this study is to propose a machine learning (ML) framework for the extraction of EEG features based on fast Fourier transform and wavelet transform for the classification of mental stress, including high- and low-level stress. EEG information from an earlier investigation [10] were decomposed with discrete wavelet transform and fast Fourier transform into different frequency bands, including alpha, theta, and beta, and computed EEG frequency bands power. Widely used ML classifiers, including Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Decision Trees (DT), and Artificial Neural Networks (ANN), were used to model the EEG dataset in order to illustrate the performance of the suggested ML framework. Next, the Fourier Transform method—a conventional transformation technique for determining power spectral density—is compared to the wavelet transform.

The remaining sections of the paper are organized as section II explains the dataset and feature extraction and classification methodology, section III reports the findings, followed by the discussion in section IV, and finally section V concludes the paper.

2. Literature Review

2.1. Stress Detection and Appraisal

It's interesting to note that stress is not a medical diagnosis or condition that calls for medical professionals to thoroughly and methodically evaluate their patients. In actuality, subjective questionnaires like the Perceived Stress Scale (PSS), which has 14 questions (later lowered to 10) that the patient must answer, are the closest measurements of stress that are utilized in a medical setting [11], Stress Response Inventory [11], Holmes and Rahe Stress Scale (Social Readjustment Rating Scale) [12], Depression Anxiety Stress Scale [13], The Hospital Anxiety and

Depression Scale [14], The State Trait Anxiety Inventory [15] and Life Events and Coping Inventory [16]. The PSS was created in 1983 and is still a widely used tool to help us comprehend how various circumstances impact our emotions and our perceived stress levels, which range from low to high. The current problem, however, is that despite all the subjective questionnaires currently in use, the only stress that patients perceive is the stress that they observe and rate for themselves without any concrete data to support the presence or absence of stress. Given that mental stress is the psychological and physiological condition that has afflicted humans for the longest duration, it is unexpected that there isn't a methodical way to conclusively determine whether stress is present.

Further, doctors and psychologists attempted to use physiological changes mentioned above for an objective measure such as increased heart rate through heart rate variability (HRV) [17]. HRV is where the amount of time between the heartbeats fluctuates slightly usually measured using an electrocardiogram (ECG) that detects the electrical activity of the heart using sensors attached to the skin of the chest. Galvanic Skin Responses (GSR) also known as electrodermal activity (EDA) measures the changes in sweat gland activity that are reflective of the intensity of our emotional state [18]. Moreover, research has tapped into the area of measuring stress through pupil dilation and blood pressure [19]. Consequently, salivary alpha-amylase and cortisol levels were used as a biomarker for stress indication due to their association with the activation of the sympathetic nervous system [20].

2.2. Electroencephalography

The use of electroencephalography (EEG) to measure stress has been the subject of recent research due to its relative affordability when compared to other methods that involve the collection of blood samples or the ingestion of radioactive chemicals, such as PET scans. But since this method is still relatively new, researchers have had differing degrees of success in identifying stress [5]. Many methods for using EEG signals to detect mental stress have been reported in the literature. These methods include using SVM, Multilayer Perceptron, and Convolutional Neural Networks in a virtual environment to detect mental stress [6]. The detection of stress is still very new and in its early stages, where researchers are still trying to determine electrical signals and patterns related to stress. One drawback of EEG is its limited spatial resolution, which makes it difficult to locate the precise region involved or responsible for stress because the electrical signals measured are only on the surface of the brain. Although there are studies that use direct experimentation and specific regions to identify stress using PET or MRI scans, these approaches are deemed impractical due to the high cost and limited accessibility of the necessary equipment. Furthermore, unlike the other resolution methods mentioned, using an EEG machine does not require extensive training.

2.3. Machine Learning Methods for Stress Detection

There has been much research that uses artificial intelligence for mental stress assessment through machine learning classifiers and algorithms. A machine learning classifier uses an independent set of information from a dataset known as features to predict the corresponding class it belongs to by having several parameters. Such features need to have unique characteristics that separate one class from another. Further, machine learning classifier will undergo either supervised or unsupervised training to predict the classes of new instances in an unseen testing dataset. Several machine learning techniques and algorithms often used in mental stress assessment are described below.

Machine learning classifiers like SVM, ANN and k-NN have shown immense potential in stress detection with each classifier achieving accuracies of more than 85% [6,21,22]. However, it seems that each classifier is feature specific when comparing each studies above. For example, SVM seems to favor frontal alpha asymmetry as a feature whereas convolutional neural network, a branch of ANN performed much better when simply considering all the brain waves. K-NN on the other hand used only a few selective electrodes and achieved accuracies of 94 and 93.7% [22]. Meanwhile, LDA relied on multiple bio-signals besides EEG such as electrocardiography (ECG), electromyography (EMG) and galvanic skin response (GSR) to perform well [23].

Further, given the complexities of the EEG signals alongside the variability in terms of the experimental conditions, a machine learning end-to-end approach may not be feasible. Among the many limitations of such an approach includes the huge amounts of data that is required to train the classifier alongside the difficulty to validate the output. Therefore, a traditional framework to train each classifier before forming an ensemble is required, especially in domain generalization, nullifying any opportunities to allow for an end-to-end deep learning model to be framed without considering the usual pre-processing and feature extraction process when training a classifier.

It is interesting to note that SVM tops the list of classifiers used when analyzing EEG for stress detection in several review papers [5,24–26]. In fact, SVM is used in almost all the experiments when the Montreal Imaging Stress Task is used as a stressor [24]. This may be because SVM has stronger discriminatory powers than LDA, and less overfitting issues compared to neural networks. However, the field of deep learning for signal processing has been growing lately, especially the usage of pretrained CNN models for robust BCI framework [27].

Recent studies performed also involved the use of SVM classifier alongside Naïve Bayes, and K-Nearest Neighbours (KNN) achieved accuracies of up to 99.98% [28] whereby the participants were induced with stress through the performance of the Stroop Colour Word Test (SCWT). In this study, four different bands were explored, namely the alpha, beta, theta, and delta band before concluding that alpha and beta bands showed a higher accuracy than the other two. Other studies involving the use of K-NN achieved maximum classification accuracy of 91.26% [29]. However, this may be caused by the low number of participants who participated in the study. Another paper involving the use of SVM and Naïve Bayes classifier successfully classified stress and

control subjects with up to 98.21% accuracies by focusing on all the power density of the frontal lobes [30]. While both achieved high accuracies in the experiment, it should be noted that these were conducted on subject-wise classification instead of a mixed classification such as control versus mental stress. When dealing with mixed classification, the accuracies dropped to around 80%, a problem that we are trying to address in our study.

2.4. EEG Wavebands

Alpha waves range from 8 to 12 Hz where it occurs in the occipital head area in the awake state. Regular meditation and relaxation have been shown to enhance alpha waves and the reason for why it is most recommended for lowering stress. Beta waves are most frequently observed around the frontal head zones and most closely associated with stress when there is an increased in beta waves. Delta waves are found in the frontocentral brain area, and these waves are associated with tiredness and early stages of sleep. Theta band is detected in anxiety activation and strongly observed in hypervigilance states such as meditation, prayer, and awareness. Finally, Gamma band is more related to ADHD and knowledge disabilities when there is an inadequate of the activity. Although the Gamma band has been associated with depression when there's inadequate gamma signals which in turn suggests a relation between mental stress and depression, it remains a relatively unexplored area. There have been multiple findings with regards to the relation of stress and the associated wavebands. Namely the alpha, theta, and beta waves [31,32]. Gamma and Delta bands has been omitted from our experimental design as it has been found that Gamma band is more closely associated to wakefulness than in relation to stress [33]. Whereas Delta band has always been associated with slow brain activity and its occurrence is most prominent during sleep. Nonetheless, some studies have suggested the delta-beta relationship with regards to anxiety [34], a trait that is closely related to mental stress.

3. Materials and Methodology

3.1. Dataset

Twenty-two healthy individuals (ages 19 to 25) without a history of illness or head trauma make up the EEG dataset. They don't take any kind of medication that could cause their heart rate to increase. Every participant participated in both the stress and control experimental sessions. They completed the Montreal Imaging Stress Task (MIST)-based Mental Arithmetic Task (MAT) [35] as it has shown the capability of producing stress-related responses involving the hypothalamic pituitary-adrenal (HPA) axis. Accordingly, eight distinct conditions—four stress levels and four control levels—were applied to each participant based on the task's level of difficulty. Only levels 1 and 4 of stress, along with the controls that go with them, were used in this experiment; level 1 is referred to as low stress, and level 4 as high stress. 128 channels are used to record the EEG data, with a 500 Hz sampling rate. In addition, the other channels are referenced to the 129th channel. Each subject has a total of two sets of 129 channels of EEG data—one for control and one for stress. Prior to being sent to the machine learning classifier for classification, the dataset is normalized. According to the experiment's original report [10], The EEG trials lasted 2200ms for high stress and control and 1100ms for low stress and control. Therefore, averaged EEG power rather than absolute power is used to compare stress vs. control trials. EEG data from two experimental conditions—low stress versus control condition and high stress versus control condition—were analyzed.

3.2. EEG Feature Extraction

EEG features are usually divided into three broad categories, the statistical time domain features, frequency domain features, and synchronicity domain features. Statistical time domain features are obtained directly from the raw EEG signals such as calculating the average amplitudes, standard deviation, and variance. Other time domain features frequently used are Hjorth parameters, entropies estimation and Higuchi's fractal dimension. It should be noted that the raw EEG signals are usually filtered for noise and artefact removal before obtaining any features for machine learning classification. Whereas frequency domain features are usually computed by first converting the raw EEG signals that are in the time domain to the frequency domain alone by applying Fourier transforms or a time-frequency domain via wavelet transforms. Common wavelet transforms usually use Daubechies set of wavelets in the fourth order. In frequency domain features, we can obtain the power spectral density, the distribution of power in its frequency components where power is defined as the amount of energy transferred per unit time, once the raw EEG has been converted. Moreover, absolute and relative powers are commonly used to check the rhythm of EEG signals.

Next, synchronicity domain features use an effective and functional measure of brain connectivity to examine significant coincidences that appear to have no apparent reason. The energy in each frequency is obtained by applying discrete wavelet transforms and Fourier transforms, which are then added to determine the absolute power or averaged to determine the power. With 129 channels and 22 participants from the experimental and control groups, the feature matrix for each EEG frequency band was 44x129x1, yielding a minimum of 5676 (22 x 2 x 129 x 1) features.

3.2.1 Wavelet Transforms

A collection of wave-like oscillations known as wavelets is produced by wavelet transforms, which break down EEG data. Wavelets are a class of functions with a variety of characteristics, including their ability to be stretched to capture high- or low-frequency data; a stretched wavelet will typically capture data at a lower frequency. Furthermore, the function's integral must have zero mean and finite energy, and its integral squared function must yield a finite number, for a function to be considered a wavelet function. This is how wavelet differs from Fourier analysis, which assumes an infinite integral squared of a sin wave. Subsequently, the wavelet's location can be adjusted to precisely pinpoint the oscillation point. The wavelet's location and "stretch-ness" can therefore be changed as we slide it across any given signal. By doing this, the signal input can be represented by the wavelet transforms in both the time and space domains.

Traditionally, the mathematical formula is represented in (1) using the first derivative of a Gaussian function. Increasing the value "a" will stretch the wavelet to capture low frequency information. In contrast, by lowering the value, the parameter "b"

establishes the wavelet's location; a left shift is necessary, and vice versa.

$$f(x) = -(x-b)e^{\frac{-(x-b)^2/(2a^2)}{\sqrt{2\pi}a^3}} \quad (1)$$

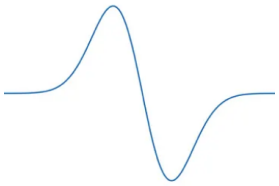


Figure 1: First Gaussian Derivative Function

Figure 2: Squishing (left) or Stretching (right) by decreasing or increasing the value of 'a'

Figure 3: Sliding the gaussian derivative left (left) or right (right) by decreasing or increasing the value of 'b'

Since a wavelet is made up of a chosen function, such as Gaussian, Harr, Daubechies, and so on, we have chosen to break down the EEG signal into sub-band frequencies using the Daubechies wavelet transform of fourth order (figure 4). This is mostly because, when compared to Haar, Morlet, and other wavelets, an EEG signal's Daubechies wavelet exhibits striking similarities to it. Moving on, we decompose the signal into 7 levels such that, in each level, half of the signal range is obtained, as shown in figure 1, where A1 to A7 is the approximate coefficient and D1 to D7 is the detailed coefficient. The sampling rate of the EEG data is set at 512 Hz.

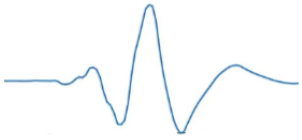


Figure 4:Daubechies Wavelet Function of the fourth order 'db4'

As a result, we can derive the EEG data's absolute power within its frequency range. To get the absolute power from all 129 channels, each of the estimated coefficients is squared before being added up. This gives us 129 channels in the alpha, beta, and theta bands, giving us three different sets of features.

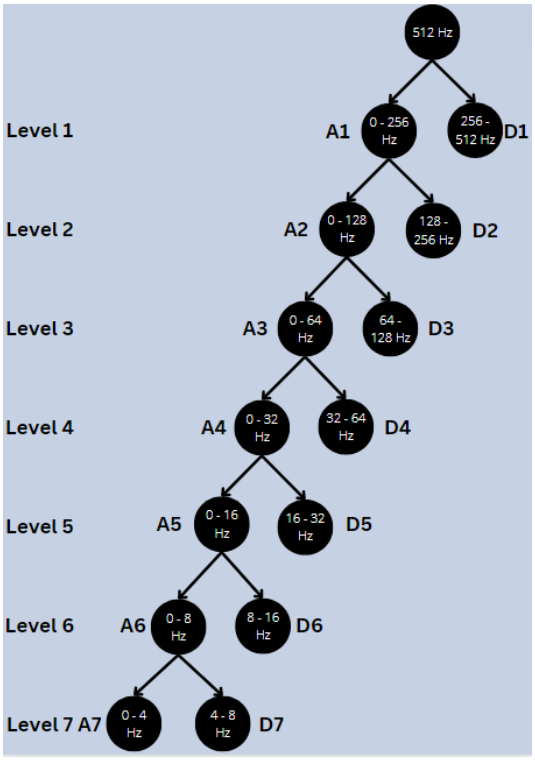


Figure 5: Wavelet Decomposition into 7 Levels

3.2.2 Fourier Transforms

Any given signal can be broken down using Fourier transforms by expanding a periodic function f(x) with an infinite sum of sines and cosines, which is a generalization of the complex Fourier series. It changes the time domain representation of the EEG data to the frequency domain. This is accomplished by using the formula in (2), where N is the number of time samples we have, n is the sample we are currently examining, x_n is the signal's amplitude at time n, k is the frequency, and X_k is the signal's amount of frequency k.

$$X_k = \sum_{n=0}^{N-1} x_n * e^{-i2\pi kn/N} \quad (2)$$

$$x_n = \frac{1}{N}\sum_{k=0}^{N-1} X_k * e^{-i2\pi kn/N} \quad (3)$$

We can determine the average power for each frequency band to be used as a feature by averaging the power obtained across the EEG dataset's sampling rate. Given that Fourier transforms data from the time domain to the frequency domain, it is limited in its applicability to time domain data. As a result, throughout the entire experimental process, we will be unable to determine when exactly the brain region responds to the stressful stimuli.

3.3. Machine Learning Classifiers

3.3.1 Support Vector Machines

A machine learning classifier called the Support Vector Machine (SVM) looks for a hyperplane in an N-dimensional space, where N is the number of features needed to clearly separate the data points. The SVM will identify the ideal hyperplane by maximizing the margins between two or more data points because there are numerous hyperplanes available to divide data points. In

order to construct the position and orientation of the hyperplane, data points that are closest to it are referred to as support vectors.

Since we are only classifying between stress and control classes, the kernel trick method suggested for non-linear dataset was not used. Consequently, MATLAB's default settings have our SVM's kernel function set to be linear, and its scale set to "automatic." The SVM's box constraint level is set to 1, and its multiclass method is configured as "One vs. One.".

3.3.2 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is an extension of Fisher's linear discriminant, which looks for a linear feature set that clearly divides two or more classes. The multivariate Gaussian function to be used for prediction is subjected to a calculation of mean and variance when there are multiple feature variables. Plotting the data, however, is assumed by (LDA) to follow the Gaussian function in a bell-curve fashion. It also presumes that the variance around the mean of each feature variable is the same. As a result, the mean, value of each feature variable, x for each class, k, can be calculated by dividing the total number of instances by the sum of values, as shown in the following formula.:

$$\mu_k = \frac{1}{N_k * \sum x} \tag{4}$$

The variance is then calculated across each class, k as follows:

$$\sigma^2 = \frac{1}{(N-K) * \sum (x-\mu)^2} \tag{5}$$

Using the input feature x from equation (6), where pi_k is the prior probability and $f_k(x)$ is the density function, LDA will apply the Bayes Theorem to estimate the probability of the predicted output class. LDA will use Bayes Theorem to estimate the probability of the predicted output class using the input feature, x that has been given in (6) where is the prior probability and is the density function.

$$P(Y = k \mid X = x) = \frac{pi_k f_k(x)}{\sum_{l=1}^{K} pi_k f_k(x)} \tag{6}$$

The default implementation of LDA from MATLAB's classification learner app was used whereby the covariance structure is set as "full".

3.3.3 Decision Tree

A decision tree is conceptualized as a tree root with numerous branches that eventually grow into leaves at the tip. But decision trees are illustrated in reverse, with the root at the top. This means that after applying a starting condition and a given value, the tree may split into branches based on a splitting criterion, ultimately leading to a final output at the leaf. Decision tree splitting criteria are typically based on the ecological diversity index, which provides a quantitative representation of the various species or classes within a dataset.

The Simpson index, on which the Gini's diversity index is based, gauges the concentration levels when people are categorized into types such that the following probabilities apply when two people are randomly selected from the dataset to represent the same type:

$$\lambda = \sum_{i=1}^{R} p_i^2 \tag{7}$$

Where R is the total number of classes in the dataset. Gini's diversity index transforms equation (5) to capture the probability that the two individual data represent different types from the following equation:

$$1 - \lambda = 1 - \sum_{i=1}^{R} p_i^2 = 1 - \frac{1}{^2D} \tag{8}$$

MATLAB's preset for a fine tree to model our decision tree such that the maximum number of splits is set at 100 while the split criterion is based on Gini's diversity index.

3.3.4 Artificial Neural Network

Artificial neural networks (ANN) are based on the idea of biological neural networks of the brain. Just as a biological neural network consists of the firing of neurons interconnected with synapses, ANN would generally consist of a few layers with connected neurons to simulate the human brain. Mainly, the first layer is known as the input layer as it receives external data as an input. The following layer is the hidden layer that obtains the raw information from the input layer and processes it by applying weights to the inputs and subsequently directs them through an activation function as the output. Finally, the output from the hidden layer is passed to the final layer known as the output layer where the ultimate result such as classifying between stress and non-stress is determined. Most ANNs allow for weight adjustments of the hidden layers by computing the gradient of the loss function with respect to its individual weight. To simplify our experimental set up for neural network models, we decided to use narrow neural network as preset by MATLAB's built in function whereby there is only one fully connected layer with the first layer size of 10 and the activation function of ReLU.

3.4. Machine Learning Framework

We extract the absolute and average power features from the collected data using the Fourier and Daubechies wavelet transforms. Next, we'll feed it into the four machine learning classifiers (SVM, LDA, DT, and NN) that were previously mentioned. Next, the dataset is divided using the ten-fold cross-validation method, which divides the training data into ten distinct subsets, or folds, with 90% of the data in each fold being used for training and the remaining 10% for validation. To get the expected validation accuracy, this is done ten times over. Lastly, ten independent trials are used to train and validate each model to determine the average accuracy, which is displayed in the results section.

4. Experimental Setup, Results and Discussion

MATLAB R2022a was used to develop our machine learning classifiers through the help of the classification learner application using a Windows 10 Operating System running with 8GB RAM and 11th Generation Intel® Core @ 2.40GHz.

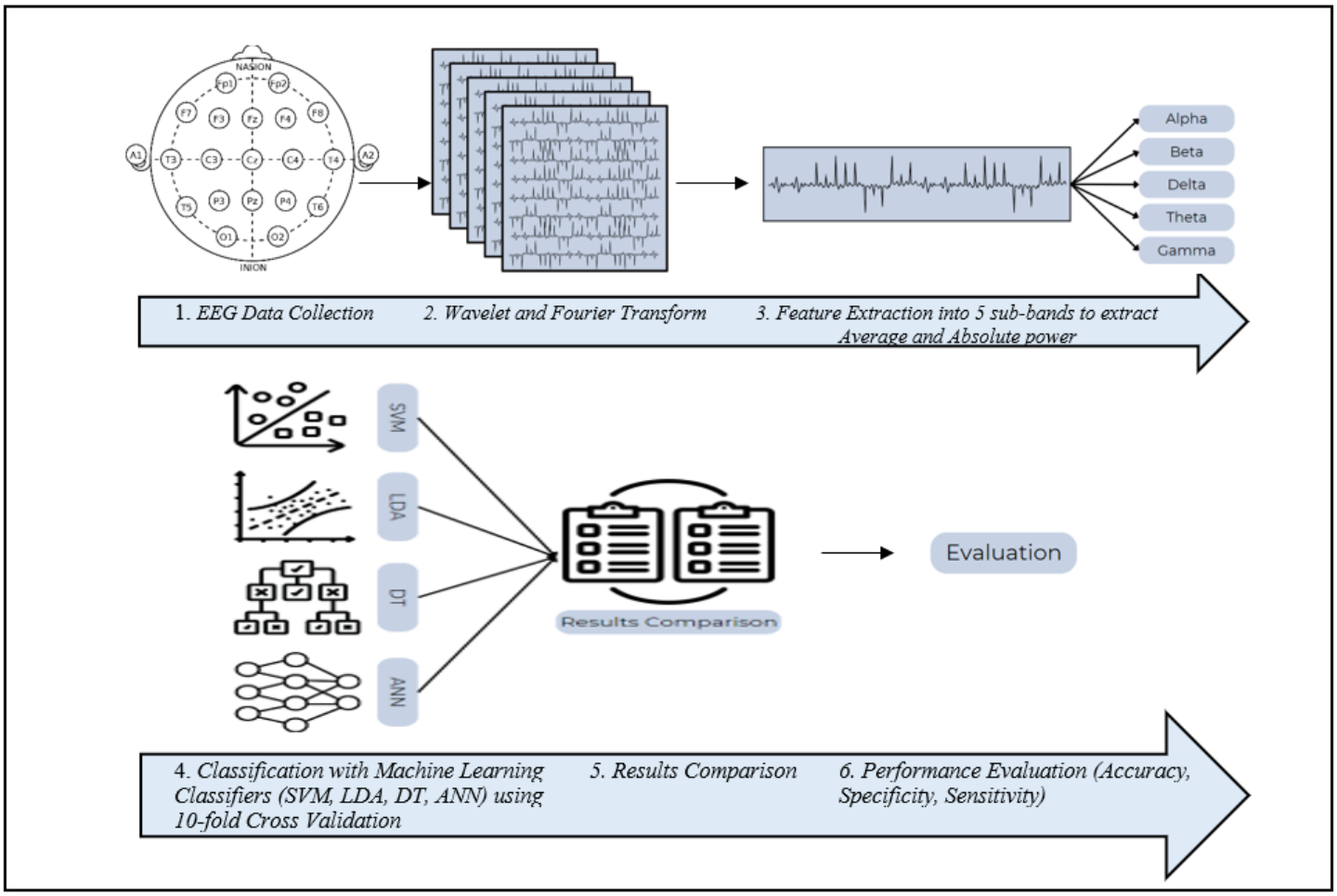


Figure 6: Proposed Machine Learning Framework

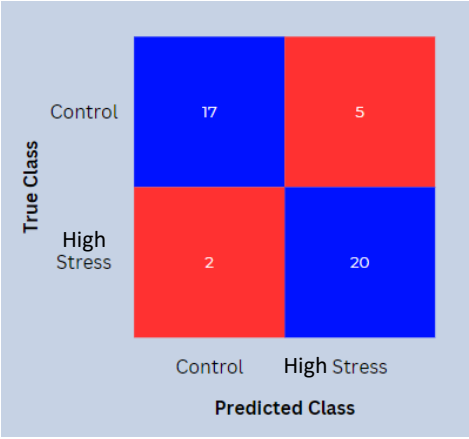


Figure 7 a: Sample of Confusion Matrix in Number of Participants

A fair assessment and evaluation are required to determine the usefulness and accuracy of the model. To validate the performance of the model, we have computed the accuracy, sensitivity and specificity of each model using the following equations in (5, 6, 7) based on the confusion matrix from figure 7. True Positive (TP) is used to denote correctly predicted cases while True Negative (TN) is used to denote correctly predicted non-cases. Likewise, False Positive (FP) and False Negative (FN) denotes incorrectly predicted cases and non-cases respectively. A sample of the confusion matrix is shown in figures 7a and 7b respectively where the control group is compared to stress such as low stress or high stress.

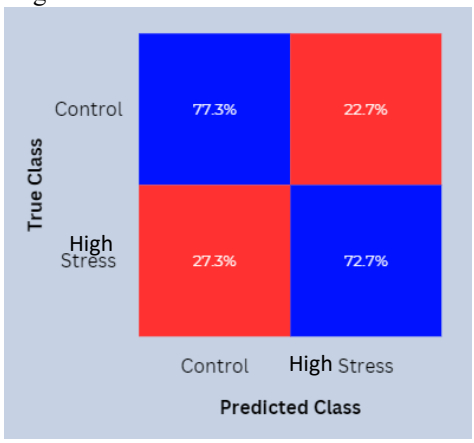


Figure 7 b: Sample of Confusion Matrix in Percentages

4.1. Accuracy

The accuracy of the classifier is denoted by the percentage of true positive (first quadrant) and true negative (fourth quadrant) over the total sum of true positive, true negative, false positive and false negative.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

4.2. Sensitivity

The sensitivity of a classifier is the percentage of correctly predicted cases (True Positives – first quadrant) over the sum of the true positives and false negatives (first and third quadrant).

$$Sensitivity = \frac{TP}{TP+FN} \quad (10)$$

4.3. Specificity

The specificity of a classifier is the percentage of correctly predicted non-cases (True Negatives – fourth quadrant) over the sum of the true negatives and false positives (fourth and second quadrant).

$$Specificity = \frac{TN}{TN+FP} \quad (11)$$

4.4. Standard Deviation

The standard deviation of each result mentioned is calculated based on formula (12) to indicate the average range of values obtained in 10 trials whereby a small standard deviation value tells us that the accuracy/sensitivity/specificity achieved does not deviate too far from the average value obtained. This is also an indication in our evaluation of whether the classifier performs in a stable manner instead of an erratic manner.

$$Standard\ Deviation = \sqrt{\frac{\sum(x_i-\mu)^2}{N}} \quad (12)$$

The rest of this section is arranged in the order of the type of machine learning classification used, followed by the experimental conditions of either control vs low stress or control vs high stress and a comparison between low stress vs high stress. Each table shows the result of the averaged validation accuracy, sensitivity, specificity, band spectrum and features used for each classification. Associated alongside is the standard deviation of the model after performing 10 trials. Finally, the best-performing features combination is bolded in each table. Given that absolute power was only used for comparison between stress and control of the same levels, it is not used in different levels as shown in control vs high stress and low stress vs high stress conditions.

Table 1: Support Vector Machine Classifier Results

Support Vector Machine Classification				
Control VS Low Stress				
Validation Accuracy (%)	***Sensitivity (%)***	***Specificity (%)***	***Band Spectrum***	***Features***
64.99±2.32	65.79±2.08	60.69±1.21	Alpha	Absolute Power
65.00±3.39	65.62±6.20	58.96±3.29	Theta	Absolute Power
67.05±3.26	72.38±10.42	63.45±5.12	Alpha, Theta	Absolute Power
65.92±3.53	72.79±6.40	63.39±3.16	Alpha, Theta, Beta	Absolute Power
Support Vector Machine Classification				
Control VS Low Stress				
Validation Accuracy (%)	***Sensitivity (%)***	***Specificity (%)***	***Band Spectrum***	***Features***
61.14±2.57	55.84±4.77	52.64±2.29	Alpha	Average Power
69.33±2.33	**75.68**±3.43	**65.24**±1.79	**Theta**	**Average Power**
57.28±3.77	63.92±4.59	58.42±2.77	Alpha, Theta, Beta	Average Power
63.86±2.14	68.54±4.70	62.03±2.76	Alpha, Theta, Beta	Absolute Power, Average Power
Control VS High Stress				
82.03±1.24	92.69±2.28	75.13±1.97	Alpha	Average Power
86.83±2.64	**94.61±0.23**	**83.10±2.60**	**Theta**	**Average Power**
82.88±1.24	91.80±2.68	76.36±2.33	Beta	Average Power
84.10±2.69	94.07±0.16	77.54±1.55	Beta, Theta	Average Power
Low Stress VS High Stress				
59.09±3.52	39.11±6.49	79.10±2.20	Beta	Average Power
72.95±4.01	66.82±8.12	79.10±2.20	Alpha	Average Power
69.54±5.38	70.00±7.94	69.07±3.96	Theta	Average Power
73.64±3.81	**74.54±5.45**	**72.72±2.88**	**Beta, Theta**	**Average Power**

Table 2: Linear Discriminant Analysis Classifier Results

Linear Discriminant Analysis Classification				
Control VS Low Stress				
Validation Accuracy (%)	***Sensitivity (%)***	***Specificity (%)***	***Band Spectrum***	***Features***
71.15±2.88	71.84±4.18	67.39±3.54	Alpha	Absolute Power
51.14±7.47	59.02±5.15	56.14±3.83	Theta	Absolute Power
73.19±5.16	80.82±4.60	71.19±2.33	Alpha, Theta	Absolute Power
75.91±1.82	80.77±5.08	67.34±2.93	Alpha, Theta, Beta	Absolute Power
47.74±4.55	51.62±4.42	51.2±3.18	Alpha	Average Power
72.95±5.50	72.48±2.39	69.04±1.97	Theta	Average Power
64.60±3.50	61.82±2.48	62.88±4.00	Alpha, Theta, Beta	Average Power
85.01±2.74	**93.25**±3.26	**79.53**±3.66	**Alpha, Theta, Beta**	**Absolute Power, Average Power**
Control VS High Stress				
71.82±3.81	73.70±5.89	70.45±3.46	Alpha	Average Power
68.41±4.14	72.32±6.12	68.04±3.34	Theta	Average Power
78.17±3.08	82.19±1.61	70.04±2.97	Beta	Average Power

Linear Discriminant Analysis Classification				
Control VS Low Stress				
Validation Accuracy (%)	*Sensitivity (%)*	*Specificity (%)*	*Band Spectrum*	*Features*
82.53±3.22	**86.08±4.90**	**77.83±2.65**	**Beta, Theta**	**Average Power**
Low Stress VS High Stress				
81.14±4.70	**72.27±8.75**	**90.02±4.46**	**Beta**	**Average Power**
63.43±7.49	67.26±8.33	59.53±10.1	Alpha	Average Power
63.18±3.49	61.35±5.48	64.98±6.13	Theta	Average Power
79.54±3.21	75.01±5.83	84.10±2.30	Beta, Theta	Average Power

Table 3: Decision Tree Classifier Results

Decision Tree Classification				
Control VS Low Stress				
Validation Accuracy (%)	*Sensitivity (%)*	*Specificity (%)*	*Band Spectrum*	*Features*
51.82±4.52	48.09±4.34	47.21±5.87	Alpha	Absolute Power
52.73±7.31	52.48±4.29	53.08±5.27	Theta	Absolute Power
54.57±3.81	52.91±5.35	53.46±5.33	Alpha, Theta	Absolute Power
58.88±3.11	56.32±5.12	58.14±7.26	Alpha, Theta, Beta	Absolute Power
47.73±4.55	49.00±7.71	49.60±12.78	Alpha	Average Power
62.30±6.67	**63.32**±2.91	**59.38**±2.73	**Theta**	**Average Power**
61.83±5.35	51.23±5.59	50.61±4.27	Alpha, Theta, Beta	Average Power
59.78±3.37	54.48±7.17	54.31±6.75	Alpha, Theta, Beta	Absolute Power, Average Power
Control VS High Stress				
70.92±3.92	68.76±4.19	73.85±3.34	Alpha	Average Power
89.99±2.32	90.1±2.28	90.46±2.19	Theta	Average Power
92.95±3.13	94.81±6.61	93.28±3.07	Beta	Average Power
94.55±2.52	**95.64±6.39**	**92.14±3.70**	**Beta, Theta**	**Average Power**
Low Stress VS High Stress				
54.54±5.65	52.73±6.48	57.27±7.09	Beta	Average Power
67.98±2.37	**61.80±2.20**	**74.09±4.11**	**Alpha**	**Average Power**
61.57±8.53	67.72±10.24	55.44±13.3	Theta	Average Power
61.83±4.96	67.26±6.69	56.35±6.15	Beta, Theta	Average Power

Table 4: Artificial Neural Network Classifier Results

Artificial Neural Network Classification				
Control VS Low Stress				
Validation Accuracy (%)	*Sensitivity (%)*	*Specificity (%)*	*Band Spectrum*	*Features*
60.22±3.09	58.62±5.15	61.79±4.17	Alpha	Absolute Power
62.73±4.58	56.34±5.46	69.09±6.37	Theta	Absolute Power
71.37±3.55	65.45±5.07	77.27±5.38	Alpha, Theta	Absolute Power
70.69±5.60	67.27±6.03	74.10±6.75	Alpha, Theta, Beta	Absolute Power
42.73±2.66	41.82±6.98	43.65±5.44	Alpha	Average Power
68.65±3.91	68.18±7.04	69.09±5.69	Theta	Average Power
58.40±4.99	54.09±6.56	62.71±6.37	Alpha, Theta, Beta	Average Power
76.59±4.76	**79.99±9.59**	**73.18±5.17**	**Alpha, Theta, Beta**	**Absolute Power, Average Power**
Control VS High Stress				
74.32±2.29	75.69±2.07	72.99±4.45	Alpha	Average Power
81.35±2.46	84.56±3.64	78.18±4.46	Theta	Average Power
85.46±2.54	85.93±2.46	85.02±3.55	Beta	Average Power
86.83±1.34	**83.65±3.03**	**90.00±1.80**	**Beta, Theta**	**Average Power**
Low Stress VS High Stress				
64.38±1.43	63.63±4.54	65.13±3.52	Beta	Average Power
67.68±2.85	72.73±3.52	62.62±5.31	Alpha	Average Power
76.36±3.10	**76.81±5.95**	**75.92±2.01**	**Theta**	**Average Power**
73.64±2.58	71.91±7.16	76.37±2.61	Beta, Theta	Average Power

5. Discussion

The results show that we can effectively classify between low stress and control and high stress and control by using the average and absolute power in the alpha, theta, and beta bands. We expanded the experimental setup even further to compare high and low stress levels while utilizing an artificial neural network as a fourth classifier. Notably, as table I, II, III, and IV demonstrate, the absolute power feature by itself does not perform well across all classifiers. For the most part, however, average power works better as a feature than other combinations of features and bands; this is not the case for LDA, which needs a combination of features from the theta band's average power and the alpha and beta bands' absolute power in order to improve the classifier's performance.

According to a previous report, in the low stress condition, the average power obtained from the theta band using Fourier transforms analysis consistently performed better across all classifiers. This is in line with previous studies that propose using theta range oscillations as a tool to identify mental illnesses and markers of mental stress [31,36]. Furthermore, all three of the initial classifiers demonstrated a significant increase in classification accuracy between high stress and control, averaging

87% ((86.83 + 82.53 + 94.55)/3) for SVM, 82.53% for LDA, and 94.55% for DT. The accuracy falls between 50% and 75% when there is less stress and control. This is expected given the strong similarities between the low levels of stress and control in terms of brain region activation. Surprisingly, the LDA classifier outperforms the other classifiers in this condition, achieving an accuracy of 85% in the classification between low stress and control by aggregating the Alpha, Theta, and Beta bands in table II. In contrast, DT performs better than all other classifiers—including LDA—in the second condition of control vs. high stress. But LDA performs better than the other classifier once more when it comes to the final condition of low stress vs. high stress. Except in the second case, LDA outperforms the other classifier in general. However, because LDA has a simpler discriminatory power to classify between two classes linearly, it did not perform significantly worse than DT, staying above the 80% range.

Although it has been suggested that Alpha and Theta bands correlates to mental stress and cognitive workload [31,36], studies have suggest the capabilities of Beta waves associated to stress [37] as well. This explains the surprising effects of aggregating the average power from the beta band to the Theta band in the control VS high-stress condition as it has been shown that beta band is used as a dynamic marker for stress assessment [32,38]. Further, our results point to the fact that alpha band waves from average or absolute power features do not perform as well generally compared to beta and theta features. This is likely because alpha waves are often correlated to awake states where further and more distinctive pre-processing methods to further segment alpha wave bands is needed to use it as a feature for stress detection.

It is therefore very likely that future research will call for the reduction of such channels in order to prevent noisy data from being used as a feature by classifiers. Furthermore, we discovered that the validation accuracy for individual Alpha and Theta waves absolute power classification using SVM is 68% and 71% when utilizing only 13 channels, specifically those from the parietal, frontal, and temporal lobes. This yields an accuracy that is comparable to, and in the case of Alpha waves, better than, using 129 channels. We further hypothesize that, since a decision tree has low discriminatory power, reducing its features should be able to achieve higher accuracy because it won't overfit the classification model like other classifiers like SVM and LDA tend to do. Nevertheless, this will be examined in later research when examining the classification of stress and control using decision trees. Furthermore, since we only have 22 subjects in our dataset, our findings are not indicative of the whole population. Larger dataset samples will be needed for future research in order to properly extrapolate our findings to other populations.

Moving on, the results produced from the neural network classifier are comparatively well given that we only used the minimum settings provided by MATLAB's toolbox extension. To be able to perform with accuracies of 76% to 86% suggests that fine-tuning the hyperparameters of the neural network and the implementation of a search strategy will aid the improvement of the results. Moreover, genetic algorithm such as swarm optimization can be implemented in addition to usual search strategies such as was done in [39] to detect emotional stress.

Additionally, there are studies involving the use of alpha asymmetry and the frontal region of the brain to determine the presence of stress. These are then used in a neurofeedback system to train people how to better manage their stress [40]. The authors also suggested the potential use of gaming simulation to better alleviate stress among university students, providing a good platform for future works regarding this.

The overall sensitivities and specificities of the best performing models tend to classify the participants as stressed compared to control. Finally, our result is comparable to the original author's dataset and its classification with our proposed features and classifiers where linear regression and naïve bayes classifiers were used to obtain accuracies from 75% to 83% [10] whilst ours ranged from 68% to 83% with SVM, LDA, DT and ANN classifiers. One aspect to consider in our slightly underperforming accuracies could be caused by the lack of features used as the original author had use absolute powers, relative powers, coherence, amplitude asymmetry and phase lag. Furthermore, our study only used three sets of features, alpha, beta, and theta to achieve the reported accuracies. This contrasts with another study using four sets of features whereby the delta waveband is included to achieve 99.98% accuracy [28]. However, it should be emphasized that the accuracy reported in this study drops to around 80% when looking at mixed classification (between stress and control) as opposed to subject-wise classification, providing a useful platform for further studies. Our future work intends to focus on the usage of singular value decomposition (SVD) to obtain the features of all the channels with minimal loss of information and possibly the use of PCA-enabled classifiers for dimensionality reduction of features.

6. Conclusion

We have successfully shown the usefulness and effectiveness of classifying mental stress states from control conditions using alpha, theta, and beta waves with the LDA, SVM, DT and ANN classifiers of up to 95% accuracy. This is in line with existing work to suggest the usage of the potential application of discrete wavelet transform with Machine Learning classifiers for EEG extraction and classification of mental states efficiently. The proposed framework based on wavelet transform shows significant potential for mental stress assessment, which could be further improved for developing a Computer-Aided Diagnosis (CAD) technique for automatic mental assessment in the future.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research work was supported from the University of Nottingham, Malaysia. The authors, therefore, acknowledge that there is no potential conflict of interest in the experiment conducted.

References

[1] M. Esler, Mental stress and human cardiovascular disease, Neuroscience and Biobehavioral Reviews, **74**, 269–276, 2017, doi:10.1016/j.neubiorev.2016.10.011.

[2] M.F. Marin, C. Lord, J. Andrews, R.P. Juster, S. Sindi, G. Arsenault-Lapierre, A.J. Fiocco, S.J. Lupien, Chronic stress, cognitive functioning and mental health, Neurobiology of Learning and Memory, **96**(4), 583–595, 2011, doi:10.1016/j.nlm.2011.02.016.

[3] E. Blix, A. Perski, H. Berglund, I. Savic, "Long-Term Occupational Stress Is Associated with Regional Reductions in Brain Tissue Volumes," PLoS ONE, **8**(6), 2013, doi:10.1371/journal.pone.0064065.

[4] M. Nagendran, Y. Chen, C.A. Lovejoy, A.C. Gordon, M. Komorowski, H. Harvey, E.J. Topol, J.P.A. Ioannidis, G.S. Collins, M. Maruthappu, "Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies in medical imaging," The BMJ, **368**, 2020, doi:10.1136/bmj.m689.

[5] S. Gedam, S. Paul, A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques, IEEE Access, **9**, 84045–84066, 2021, doi:10.1109/ACCESS.2021.3085502.

[6] D. Kamińska, K. Smółka, G. Zwoliński, "Detection of mental stress through EEG signal in virtual reality environment," Electronics (Switzerland), **10**(22), 2021, doi:10.3390/electronics10222840.

[7] V. Sulimova, D. Windridge, S. Bukhonov, V. Mottl, Quick breast cancer detection via classification of evoked EEG potentials in the mammologist's brain.

[8] H.U. Amin, W. Mumtaz, A.R. Subhani, M.N.M. Saad, A.S. Malik, "Classification of EEG signals based on pattern recognition approach," Frontiers in Computational Neuroscience, **11**, 2017, doi:10.3389/fncom.2017.00103.

[9] H.U. Amin, A.S. Malik, N. Badruddin, W.T. Chooi, "Brain behavior in learning and memory recall process: A high-resolution EEG analysis," in IFMBE Proceedings, Springer Verlag: 683–686, 2014, doi:10.1007/978-3-319-02913-9_174.

[10] A.R. Subhani, W. Mumtaz, M.N.B.M. Saad, N. Kamel, A.S. Malik, "Machine learning framework for the detection of mental stress at multiple levels," IEEE Access, **5**, 13545–13556, 2017, doi:10.1109/ACCESS.2017.2723622.

[11] S. Cohen, T. Kamarck, R. Mermelstein, A Global Measure of Perceived Stress, 1983.

[12] T.H. Holmes, R.H. Rahe$, THE SOCIAL READJUSTMENT RATING SCALE"?, Pergamon Press, 1967.

[13] S.H. Lovibond, P.F. Lovibond, "Manual for the Depression Anxiety Stress Scales," in Psychology Foundation, 1995.

[14] A.S. Zigmond, R.P. Snaith, "The Hospital Anxiety and Depression Scale," Acta Psychiatrica Scandinavica, **67**(6), 361–370, 1983, doi:10.1111/j.1600-0447.1983.tb09716.x.

[15] C. Spielberger, R. Gorsuch, R. Lushene, P.R. Vagg, G. Jacobs, Manual for the State-Trait Anxiety Inventory (Form Y1 – Y2), 1983.

[16] J.E. Dise-Lewis, The Life Events and Coping Inventory: An Assessment of Stress in Children, 1988.

[17] T. Pereira, P.R. Almeida, J.P.S. Cunha, A. Aguiar, "Heart rate variability metrics for fine-grained stress level assessment," Computer Methods and Programs in Biomedicine, **148**, 71–80, 2017, doi:10.1016/j.cmpb.2017.06.018.

[18] S. Betti, R.M. Lova, E. Rovini, G. Acerbi, L. Santarelli, M. Cabiati, S. Del Ry, F. Cavallo, "Evaluation of an integrated system of wearable physiological sensors for stress monitoring in working environments by using biological markers," in IEEE Transactions on Biomedical Engineering, IEEE Computer Society: 1748–1758, 2018, doi:10.1109/TBME.2017.2764507.

[19] A. Barreto, J. Zhai, M. Adjouadi, Non-intrusive Physiological Monitoring for Automated Stress Detection in Human-Computer Interaction, 2007.

[20] S. Cozma, L.C. Dima-Cozma, C.M. Ghiciuc, V. Pasquali, A. Saponaro, F.R. Patacchioli, "Salivary cortisol and α-amylase: Subclinical indicators of stress as cardiometabolic risk," Brazilian Journal of Medical and Biological Research, **50**(2), 2017, doi:10.1590/1414-431X20165577.

[21] S.M.U. Saeed, S.M. Anwar, H. Khalid, M. Majid, U. Bagci, "EEG based classification of long-term stress using psychological labeling," Sensors (Switzerland), **20**(7), 2020, doi:10.3390/s20071886.

[22] L.D. Sharma, R.K. Saraswat, R.K. Sunkaria, "Cognitive performance detection using entropy-based features and lead-specific approach," Signal, Image and Video Processing, **15**(8), 1821–1828, 2021, doi:10.1007/s11760-021-01927-0.

[23] J. Minguillon, E. Perez, M.A. Lopez-Gordo, F. Pelayo, M.J. Sanchez-Carrion, "Portable system for real-time detection of stress level," Sensors (Switzerland), **18**(8), 2018, doi:10.3390/s18082504.

[24] R. Katmah, F. Al-Shargie, U. Tariq, F. Babiloni, F. Al-Mughairbi, H. Al-Nashash, A review on mental stress assessment methods using eeg signals, Sensors, **21**(15), 2021, doi:10.3390/s21155043.

[25] S.S. Panicker, P. Gayathri, A survey of machine learning techniques in physiology based mental stress detection systems, Biocybernetics and Biomedical Engineering, **39**(2), 444–469, 2019, doi:10.1016/j.bbe.2019.01.004.

[26] S. Lotfan, S. Shahyad, R. Khosrowabadi, A. Mohammadi, B. Hatef, "Support vector machine classification of brain states exposed to social stress test using EEG-based brain network measures," Biocybernetics and Biomedical Engineering, **39**(1), 199–213, 2019, doi:10.1016/j.bbe.2018.10.008.

[27] M.T. Sadiq, M.Z. Aziz, A. Almogren, A. Yousaf, S. Siuly, A.U. Rehman, "Exploiting pretrained CNN models for the development of an EEG-based robust BCI framework," Computers in Biology and Medicine, **143**, 2022, doi:10.1016/j.compbiomed.2022.105242.

[28] Y. Badr, F. Al-Shargie, U. Tariq, F. Babiloni, F. Al Mughairbi, H. Al-Nashash, "Classification of Mental Stress using Dry EEG Electrodes and Machine Learning," in 2023 Advances in Science and Engineering Technology International Conferences, ASET 2023, Institute of Electrical and Electronics Engineers Inc., 2023, doi:10.1109/ASET56582.2023.10180884.

[29] M. Maruf Hossain Shuvo, T. Rahman, A. Kumer Ghosh, M. Mostafizur Rahman, Mental Stress Recognition using K-Nearest Neighbor (KNN) Classifier on EEG Signals, 2015.

[30] O. AlShorman, M. Masadeh, M.B. Bin Heyat, F. Akhtar, H. Almahasneh, G.M. Ashraf, A. Alexiou, "Frontal lobe real-time EEG analysis using machine learning techniques for mental stress detection," Journal of Integrative Neuroscience, **21**(1), 2022, doi:10.31083/j.jin2101020.

[31] S.A. Awang, P.M. Pandiyan, S. Yaacob, Y.M. Ali, F. Ramidi, F. Mat, "Spectral density analysis: theta wave as mental stress indicator," in Communications in Computer and Information Science, 103–112, 2011, doi:10.1007/978-3-642-27183-0_12.

[32] H.M. Díaz, F.M. Cid, J. Otárola, R. Rojas, O. Alarcón, L. Cañete, "EEG Beta band frequency domain evaluation for assessing stress and anxiety in resting, eyes closed, basal conditions," in Procedia Computer Science, Elsevier B.V.: 974–981, 2019, doi:10.1016/j.procs.2019.12.075.

[33] E.T. Attar, Review of electroencephalography signals approaches for mental stress assessment, Neurosciences, **27**(4), 209–215, 2022, doi:10.17712/nsj.2022.4.20220025.

[34] K.L. Poole, B. Anaya, K.E. Pérez-Edgar, "Behavioral inhibition and EEG delta-beta correlation in early childhood: Comparing a between-subjects and within-subjects approach," Biological Psychology, **149**, 2020, doi:10.1016/j.biopsycho.2019.107785.

[35] K. Dedovic, R. Renwick, N. Khalili Mahani, V. Engert, S.J. Lupien, J.C. Pruessner, K. Mahani, P. -Douglas, The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain, 2005.

[36] T. Okonogi, T. Sasaki, "Theta-Range Oscillations in Stress-Induced Mental Disorders as an Oscillotherapeutic Target," Frontiers in Behavioral Neuroscience, **15**, 2021, doi:10.3389/fnbeh.2021.698753.

[37] Universiti Teknologi MARA. Faculty of Electrical Engineering, IEEE Control Systems Society. Chapter Malaysia, Institute of Electrical and Electronics Engineers, Proceedings : 2015 6th IEEE Control and System Graduate Research Colloquium (ICSGRC 2015) : 10 - 11 August 2015, Shah Alam, Malaysia : Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia.

[38] W. Junaidee bin Wan Hamat, H. Majdi bin Ishak, K. Hashikura, T. Suzuki, K. Yamada, Detection of Anxiety Expression From EEG Analysis Using Support Vector Machine.

[39] D. Shon, K. Im, J.H. Park, D.S. Lim, B. Jang, J.M. Kim, "Emotional stress state detection using genetic algorithm-based feature selection on EEG signals," International Journal of Environmental Research and Public Health, **15**(11), 2018, doi:10.3390/ijerph15112461.

[40] Y. Hafeez, S.S.A. Ali, R.A. Hasan, S.H. Adil, M. Moinuddin, M. Ebrahim, M.S.B. Yusoff, H. Amin, U. Al-Saggaf, "Development of Enhanced Stimulus Content to Improve the Treatment Efficacy of EEG-Based Frontal Alpha Asymmetry Neurofeedback for Stress Mitigation," IEEE Access, **9**, 130638–130648, 2021, doi:10.1109/ACCESS.2021.3114312.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 8, No. 5, 55-61 (2023)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

The Graded Multidisciplinary Model: Fostering Instructional Design for Activity Development in STEM/STEAM Education*

Mauricio Flores-Nicolás*, Magally Martínez-Reyes, Felipe de Jesús Matías-Torres

Universidad Autónoma del Estado de México – Centro Universitario Valle de Chalco, Estado de México, 56615, México

ARTICLE INFO

Article history:
Received: 01 August, 2023
Accepted: 30 September, 2023
Online: 30 October, 2023

Keywords:
21st Century Skills
Gamification
Process Model
Middle School
Educational Technology

ABSTRACT

In a challenging and increasingly technological world, it is important to promote critical thinking, multidisciplinary problem solving, and collaboration through STEAM education; however, there are important economic, administrative, and especially pedagogical management limitations for its implementation at the secondary level. Therefore, this paper presents systematic recommendations for an effective and sustainable implementation of STEAM education in educational institutions through the Gradual Multidisciplinary Model (GMM), which allows the identification and specific adaptation of STEAM knowledge through the topic of logic gates related to the representations of disjunction and conjunction in Boolean algebra (university content) to its physical representation in Minecraft (high school content). The quasi-experimental method allows to evaluate the results through the application of a pre-test designed to measure logical-mathematical thinking and a post-test designed to measure the level of understanding of practical skills and the students' perception of the learning experience. The results obtained by t-student show that there is a significantly high difference between the means and suggest that the educational intervention orchestrated by the GMM had a significant impact on the performance and skills of the participants, since different levels of understanding (gradualness) and perception of the concepts related to logic gates could be identified; while the qualitative assessment shows the group's willingness and enthusiasm to work with a practical and meaningful activity using Minecraft, which allows them to specifically apply their skills in science, technology, engineering, art and mathematics.

1 Introduction

STEAM education is an educational approach that integrates the disciplines of science, technology, engineering, arts, and mathematics into the teaching and learning process. This strategy aims to foster an interdisciplinary view of knowledge and encourages students to approach problems and challenges holistically and creatively [1]. STEAM education has now gained greater recognition and acceptance in various education systems around the world. It has been documented that schools and teachers are trying to adopt this approach to prepare students for the current and future challenges of the labor market. This is due to the promotion of 21st century skills such as critical thinking, problem solving, and collaboration, as well as preparing students for future careers that require STEAM [2, 3].

There are a number of benefits that STEAM education offers, which are described below: Focus on problem solving, this is promoted through hands-on projects and activities designed to address authentic challenges, which helps students develop critical skills and find innovative and creative solutions; Inclusion of the arts and humanities, their importance in the holistic education of students has begun to be recognized, as greater interaction between STEAM disciplines has been promoted to foster a broad and balanced understanding of the world [2, 4]; Increased use of technology, STEAM education incorporates digital tools and resources in the teaching-learning process, as it enhances experimentation, simulation and collaboration, as well as facilitates access to up-to-date information and data [5, 6]; and focus on gender equality, with the development of specific programs that promote gender equality in the participation of girls and women in the various areas of STEAM [7, 8].

In addition, STEAM education is important at all levels of education [6], especially at the secondary level, where the focus is on preparing students to work in technical fields and develop innovative solutions, as it is essential that they acquire these skills in a society increasingly driven by technology and innovation. Developing critical thinking, problem solving, and creativity is

*This is an expanded version of the article published in the 2022 IEEE Mexican International Conference on Computer Science (ENC)
*Corresponding Author: Mauricio Flores-Nicolás, México, mfloresn90@icloud.com

also important because it enables them to analyze complex situations and develop new ways of approaching problems or projects. Meaningful, hands-on experiences increase their motivation and understanding of abstract concepts and scientific and mathematical principles [4, 9, 10, 11].

Beyond the benefits and importance of STEAM education, it faces several challenges, such as the high cost of materials, the digital literacy of teachers and the technological infrastructure in educational institutions [1, 6, 12, 13], as well as the lack of guidelines or standards that allow for the development and implementation of these types of activities in a satisfactory manner depending on the educational level. For example, at the secondary level [9, 10, 14], they mention that one of the main difficulties teachers face when teaching STEAM activities is time management and the pressure to follow the curriculum. That is, teachers are squeezed between authentic learning, student engagement, and multiple pathways with lesson plan deadlines, especially because a change in the way of teaching would affect the number of ways to assess.

Similarly, teaching basic science concepts at a lower level of education can be valuable in stimulating early interest in science, preparing students for higher education, and encouraging critical thinking. This provides a broader context for how science works in practice and makes the subject relevant; however, it is essential to assess the readiness and maturity of students and to balance and tailor science-based instruction to provide appropriate motivation and support. In addition, it is important to recognize that not all students will be interested in science, so it is important to draw on diverse academic aspirations in the classroom [15, 16].

It is for this reason that the purpose of this paper is to present the GMM in its two stages: first, how to design and build a STEAM activity with the elements described in [1]; and second, to execute such activity effectively and efficiently with secondary school students with a data analysis that proves a significant learning in the students on the topic. It is emphasized that STEAM education is not limited to a specific educational level, therefore the effectiveness of the GMM will be through the topic of logic gates, which is strongly linked to the basic laws of Boolean algebra (disjunction, conjunction, and negation), addition and multiplication (which will serve as a starting point), with the aim of enabling students to develop STEAM skills. The application of this content at the secondary level is of paramount importance because of its potential to fully prepare students for the technological and globalized future. At this stage, young people are eager to discover their passions and talents, and STEAM education offers an interdisciplinary approach that allows them to explore different fields of knowledge in creative and meaningful ways.

This paper discusses a quasi-experimental field study of the implementation of STEAM education in an educational setting using the GMM. The introduction provides an overview of the STEAM approach and highlights its benefits in developing key skills and competencies in students. The materials and methods section describes the design of the study, including the selection of educational institutions and educational resources. The development section describes the pedagogical strategies, practical activities, and classroom implementation. The results present empirical evidence and student testimonies on the impact of STEAM education on the learning process. The discussion analyzes the findings in light of the existing literature, highlighting the strengths and challenges of STEAM implementation in the educational context. Finally, the conclusions synthesize the findings and provide recommendations for effective and sustainable implementation of STEAM education in educational institutions, highlighting its potential for fostering critical thinking and creativity and preparing students for an increasingly technological and changing world.

2 Material and Methodology

The previous research in [1] has provided a valuable starting point, establishing the methodological foundations on which the present study is built. The GMM is designed not only to address current challenges in education, such as the need to effectively integrate technology into the teaching-learning process, but also to promote learning through the design of activities with graded and multidisciplinary approaches. This model draws on several existing pedagogical methodologies and learning strategies, such as constructionism, conceptual change, simulations, gamification, and problem-based learning, but is not limited to these alone, as content adaptation requires simplification, contextualization and careful consideration of students' skills and maturity levels. This includes ensuring that students have a solid foundation of basic concepts related to the topic, motivating them with relevant examples and applications, and providing resources and appropriate instructional support. Similarly, the topic of logic gates in [1] is aimed at undergraduate students, so adaptation is required to understand the characteristics and needs of the target audience, simplify complex information, adapt the language, focus the content, present it clearly, and contextualize it to make it relevant, meaningful, and interesting.

For this reason, this research was carried out in two parts. First, the construction of the activity, which consists in following the triple process of the GMM that starts with a hypothetical learning path (imaginary design or plan that describes the progress and learning process that the individual could follow in a certain subject or area of knowledge [1, 6]), analyzes and adapts the available technological infrastructure (the Minecraft Education software and an IntelNUC) and concentrates the information in an educational planning (it consists in designing and structuring the curricula, contents, methodologies, resources and evaluations to achieve the educational objectives effectively). And second, the evaluation stage, which involves collecting data from secondary school students by observing the learning process and the students' performance while they participate in the activity. The pre and post-test measurement technique was used because by comparing the results it is possible to determine if there has been a significant change in a variable of interest.

In addition, the following research questions will be specifically addressed to provide a deeper and more rigorous understanding of the implementation of GMM in secondary STEAM education: 1)

What is the level of students' mathematical knowledge and competence before and after participating in a STEAM activity; 2) How can STEAM education be effectively and sustainably implemented in educational institutions; and 3) How do students perceive the STEAM learning experience and what are the perceived challenges and benefits? These research questions will serve as a fundamental guide to explore the effectiveness of GMM in teaching logic gates in their deep context (Boolean algebra and electronic symbology), as seen from its representation for high school students using Minecraft.

2.1 STEAM activity design and construction

First of all, an adaptation of the hypothetical learning trajectories proposed in [1, 6] has been carried out, since in these has already been carried out a disciplinary analysis on the subject applied to high school students, it only remains to adapt the presentation for secondary school students, remaining as follows: 1) Science, logical propositions are introduced through philosophical and logical approaches; 2) Technology, a map created in Minecraft Education helps in the exploration and construction of structures with similar function to the main firebrand; 3) Engineering, works intermediate-advanced topics through truth tables, algorithms and real applications; 4) Art, a way to develop creative skills is with the topic of schematic diagrams, allowing students to apply decision-making in the design of electronic circuits; 5) Mathematics, starting with concepts of addition and multiplication and moving to their representation in the context of Boolean algebra.

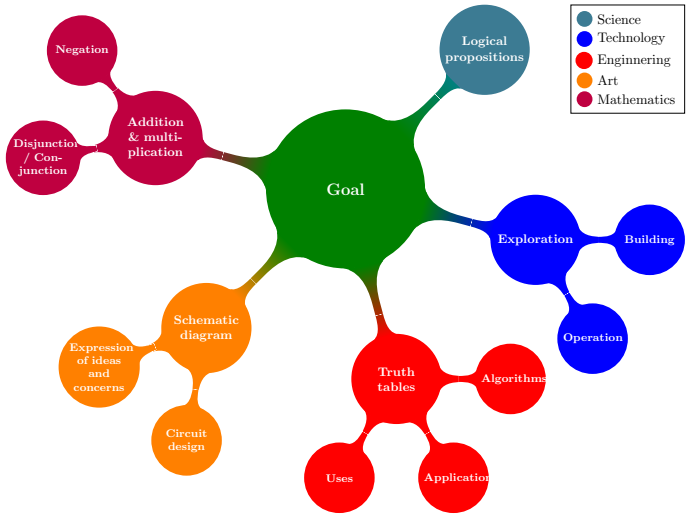


Figure 1: Hypothetical learning trajectory, classified by STEAM discipline and distributed by complexity. Own elaboration based on [1, 6].

The distribution of content is shown in figure 1, with the main theme at the center and branched into different disciplines to facilitate the implementation of STEAM activities. The objective of the second stage of the GMM is to analyze the technological infrastructure in the educational institution and align it with the pedagogical objectives, ensuring that solutions are selected that respond to specific educational needs and goals. It also optimizes their use to enhance teaching and learning and provide meaningful and effective experiences for students.

	Minecraft	Minecraft Education
Price	$6.99 – $39.99, depending on the device	$5.04 per year (institutional management required) or $12 per year (independent educator)
Does it fit in the budget?	Yes, if you have at least one compatible device	Yes, it is appropiate to carry out institutional management for the licenses
Risks	That it becomes a distraction	
Suggestion	Controlling its use	
Collaboration	Yes, in local network and online (requires additional subscriber)	Yes, in local
Trial version	14 days	26 logins
Educational elements	No	Camera, portfolio, tutors, materials (periodic table, control of actions within the map, among others)
Functioning in the institutional computer (IntelNUC)	Best	
Additional requirements	No	Internet connection
Minimum training for its use	Included in the set	
Minimum teaching qualifications	Not available	Minecraft Learn with over 100 lessons
Examples of STEAM educational activities	Not available	Included in-game or in Minecraft Education
Can i share my own game files	Not available	Maps can be imported / exported and .pdf files of the evidence can be generated.
Are there educational resources to facilitate the development of the activity?	Not available	Yes, at: https://bit.ly/3rY76J8 (Official), How to create a map in Minecraft? & How to use Minecraft to teach high school level classes
Controllers	Joystick	Keyboard, mouse and/or joystick
Download	Xbox 360, Xbox One & Xbox series S|X, available in your online store Playstation 3, 4 y 5, available in your online store Nintendo Wii U, Switch, available in your online store Android, available in your online store	PC / macOS / Google Chromebook, Minecraft Education iPhone o iPad, available in your online store Android, available in your online store

Figure 2: Minecraft software analysis report [6].

Some of the advantages of using GMM analysis are that some of the previously created content can be reused, allowing learning to be personalized, resources and tools to be adapted to different learning styles, and leading to a more enriching and relevant education. A report of the software to be used is shown in figure 2, which provides detailed and accurate information about the operation of Minecraft and Minecraft Education IntelNUC computer equipment. Part of the information included in this report is the technical specifications, such as operating systems and compatible devices, examples of activities on similar topics, risks, suggestions, costs, and most importantly, where and how to get trained in this technology if you want to implement it in the classroom.

The third stage of the GMM provides an educational planning format that organizes the ideas, materials, goals, strategies, and pedagogical methods to achieve success during the educational process to achieve significant and lasting results. In figure 3 shows the planning format for activities in STEAM education, which aims to integrate teaching in its five disciplines with a multidisciplinary approach. In this process, teachers can design their activities and hands-on projects that foster critical thinking, problem solving, and creativity by drawing on the results of the previous two stages of the GMM. Clear learning objectives are set, and resources and digital tools are selected to indicate how and when they will intervene in the session in a way that enriches the educational experience. This format promotes collaborative work and meaningful learning,

allowing students to explore, investigate, and apply concepts and skills in real-world situations.

Project	Is it possible to add and multiply in Minecraft?
Goal	Have students build structures with the equivalent operation of logic gates or the fundamental laws of Boolean algebra. Participants can respond by using switches of buttons on the technology resource to see if the practice matches the theory.

ISTE Standards

1.3 Knowledge Builder. Students critically evaluate a variety of resources using digital tools to construct knowledge, produce creative artifacts, and develop meaningful learning experiences for themselves and others.

1.5 Computational Thinker. Students develop and employ strategies to understand and solve problems in ways that harness the power of technological methods to develop and test solutions.
- 1.5.c. Students break problems into component parts, extract key information, and develop descriptive models to understand complex systems of facilitate problem solving.

Level of technology insertion

Substitution (X)	Augmentation (X)	Modification (X)	Redefinition (X)

Pedagogic methodology	Constructionism
Learning strategy	Conceptual change; Problem-based learning; Simulation; Gamification.

How to apply it?

Science	**Technology**	**Enginnering**	**Arts**	**Mathematics**
Use and development of logical-mathematical and logical-computational thinking, basic fundamentals of circuit theory.	To favor the decision-making process with the choice of the technological tool that will allow you to solve the problem.	Know the fundamentals of electronic component manufacturing, such as design, simulation, fabrication and testing.	Designers use their creativity to create complex circuits that solve specific problems, and the circuit design itself is visually appealing process.	Theoretical verification of the operation exposed in the technological tools, used to describe the relationship between inputs and outputs of a digital circuit.

Didactic sequence

Opening (20 minutes)
1. Teacher's presentation to the class.
2. Ask students if they are familiar with Minecraft.
3. Ask students about logic gates.
4. Resolve pre-test.

Development (50 minutes)
1. The teacher presents the objectives of the activity.
2. It briefly explains about Minecraft and the basic logic gates trainer.
3. Definition and importance of logic gates.
4. Theoretical-practical definition and operation of the NOT gate in its Boolean algebra ($\neg$), mathematical ($\bar{A}$) and electronics (IC 74LS04).
5. Solve doubts and errors.
6. Theoretical-practical definition and operation of the AND gate in its Boolean algebra representation ($\wedge$), mathematical (*) and electronics (IC 74LS08).
7. Solve doubts and errors.
8. Theoretical-practical definition and operation of the OR gate in its Boolean algebra representation ($\vee$), mathematical (+) and electronics (IC 74LS32).
9. Solve doubts and errors.
10. Resolve post-test.

Closing (15 minutes)
1. Summarize what was covered in class, highlighting the importance of logic gates for everyday life.
2. Ask about their experience using Minecraft and the basic logic gates trainer in the classroom.
3. Farewell.

Materials
- Diagnostic evaluation (pre-test).
- Final evaluation (post-test).
- IntelNUC with Xbox 360 controller configured and the videogame Minecraft Education (basic logic gates map).
- Basic logic gate trainer (electrical extension, power supply, banana plug male to banana plug male cables).
- Slides.

Evaluation	Data collection through pre-test and post-test, with algebraic and mathematical operations exercises using technological tools for verification.

Figure 3: Planning format. Own elaboration based on [1, 6].

2.2 STEAM activity implementation and evaluation

The evaluation of the STEAM educational activity was carried out on a group of 43 high school sophomores in Mexico through the application of a pre-test designed to measure mathematical reasoning and a post-test designed to measure the students' level of understanding, practical skills, and perception of the learning experience. The data obtained were analyzed using a student's t-test to determine if there were statistically significant differences between the pre and post-test scores.

To design the pre-test measurement instrument, a questionnaire was developed to specifically assess students' logical-mathematical thinking prior to participating in the experience. This questionnaire includes questions on basic mathematical concepts such as addition and multiplication, problem solving skills, and logical reasoning, with the goal of establishing a baseline for the participants' level of mathematical knowledge and competence.

On the other hand, the post-test measurement tool was designed to assess the understanding and impact of the STEAM activity on the students. This questionnaire includes questions that assess the knowledge gained during the activity, the practical skills developed, and the overall perception of the STEAM learning experience. In addition, open-ended questions were included to allow students to express their opinions about the activity, its challenges, and perceived benefits, which will provide more detailed feedback on the teaching process and the level of learning achieved.

Together, these pre and post-test measurement tools provide a comprehensive view of students' progress in mathematical-logical thinking, understanding, and perception of STEAM activities. The combination of these instruments allows for analysis of the impact of the activity on the development of mathematical skills and the learning experience, which facilitates the identification of areas for improvement and informed decision making for future implementation of STEAM activities.

3 Comments on the experience

Some important observations were made while teaching logic gates:

1. Abstract concept. Logic gates are abstract concepts that may be difficult for some students to understand, especially those without a strong background in mathematics or logic. This is due to the use of symbols and logical operations that may seem unfamiliar.

2. Relationship to other disciplines. The topic caused some confusion among students who are not familiar with electronics or computer science concepts. For example, some confused logic gates with physical gates or similar concepts in logic. Also, at first it was not clear to them that the "sum" in the OR logic gate is different when there is 1 + 1 in its truth table, until it was clarified with the existence of binary numbers or simply transferring it to logic disjunction, where true $\vee$ (disjunction) true is true.

3. Supplement with examples. It was useful to provide examples of logic gates in Minecraft because, by using a different context, their translation to scientific knowledge, which is the goal of the activity, was not complex.

4. Encourage logical thinking. Teaching logic gates in secondary school is an opportunity to develop logical thinking and problem solving. Teachers can design activities and exercises that challenge students to apply their logic and reasoning skills.

Figure 5 shows the group of students excited to work together in the school lab. They have the Minecraft Education video game at their disposal, which the entire group can see on a projector. The atmosphere is dynamic and lively as the students collaborate and discuss ideas to solve STEAM activity problems. Likewise, a teacher is also present in the center, providing guidance and support to the students in their learning process. The students' expressions are one of enthusiasm and excitement as they are immersed in a meaningful, hands-on experience that allows them to apply their skills in science, technology, engineering, art, and math.

(a) Secondary school team 1/2.

(b) Secondary school team 2/2

Figure 5: Educational experience orchestrated by the GMM. Own elaboration.

4 Results and discussions

Qualitative interpretation of data from the STEAM activity on basic logic gates in high school students reveals interesting findings. After analyzing the students' responses from the pre-test, post-test, and observations during the activity, different levels of understanding (gradualness) and perception of the concepts related to logic gates could be identified.

Some students showed a solid understanding of how logic gates work and how they relate to the mathematical operations of addition and multiplication. They recognized the AND gate to a multiplication operation in mathematics and the conjunction in Boolean algebra, as well as how the OR gate relates to addition. These students were able to make clear connections between logic gates and mathematical operations, indicating a deep understanding of the concepts. On the other hand, students who remained confused in connecting logic gates to mathematics and algebra may have mixed concepts or made inaccurate comparisons between operations. These findings suggest the need to present some previous fundamentals, such as binary numbers and digital signals, during the teaching of logic gates. In addition, during the activity, it was observed that the level of participation and collaboration among the students was higher than in a traditional classroom, which means that they can work together and solve problems as a group. Overall, the qualitative data from this STEAM activity on basic logic gates provided a more holistic and detailed view of students' understanding and experience with the topic.

Now, to address the quantitative data, the Student's t-test for independent samples will be used, this is applied when there is a data set with two different samples, in this case it is so because the pre-test focuses on measuring the students' mathematical reasoning skills before the intervention, while the post-test measures the knowledge acquired after the intervention, and it is necessary to determine if there is a significant difference between the means of the two samples. The bilateral formulation of the null and alternative hypotheses in this test focuses on two possibilities, which are presented below.

$$H_0: \mu \text{ pre-test} = \mu \text{ post-test}$$

$$H_1: \mu \text{ pre-test} \neq \mu \text{ post-test}$$

The null hypothesis (H_0) states that there is no significant difference between the means, while the alternative hypothesis (H_1) states that there is a significant difference between the means. During the t-test process, a value of t is calculated and compared to a critical value to determine whether the null hypothesis is rejected. If the value of t is greater or less than the critical value, it is concluded that there is a significant difference between the means and the alternative hypothesis is accepted. On the other hand, if the value of t is not sufficiently extreme, the null hypothesis is accepted, indicating that there is no significant difference between the means.

Table 1: Values of Student's t-test assuming equal variances in the high school group. Own elaboration.

	Pre-test	**Post-test**
Mean	6.710963455	2.373285629
Variance	2.092797252	2.069731766
Observations	43	43
Pooled Variance		2.081264509
Hypothesized Mean Diff.		0
Pooled Variance		2.081264509
df		84
t Stat		13.94160851
P($T <= t$) one-tail		$7.22784e^{-24}$
t Critical one-tail		1.663196679
P($T <= t$) two-tail		$1.44557e^{-23}$
t Critical two-tail		1.988609667

The table 1 shows the results of the Student's t-test, describing the values as follows: Mean, in the pre-test, represents the average value of their logical-mathematical skills scores before the intervention, while the post-test represents the results obtained by the participants after the STEAM activity; Variance, indicates how far the individual scores deviate from the mean; Sample, there were 43 students who participated in the activity; Common variance, represents the average variance of both groups combined; Mean difference hypothesis, indicates that the difference between the pre-test and the post-test is equal to zero, meaning that there is no significant difference between the means of the two groups; Degrees

of freedom, represents the number of data that are free to vary after taking into account the restrictions imposed by the study design, in this case it is obtained by subtracting 1 from the total sum of both samples; t-statistic, with a value of 13.94, indicating that the sample means are far apart, indicating a significant difference between the pre and post-test means; P two-tailed value, is $1.44557e^{-23}$, close to zero, indicating a significant difference between the pre and post-test means; the critical two-tailed value is 1.98, which serves as a reference point to compare the t statistic, in this case the statistic value is much greater than the critical value, indicating that the null hypothesis should be rejected, and it is concluded that there is a significant difference between the pre and post-test means.

The results indicate that there is a significantly high difference between the means and there is sufficient statistical evidence to reject the null hypothesis. This indicates that the educational intervention provided by the GMM had a significant impact on the performance and skills of the participants:

1. Specific learning. During the educational intervention, students acquired specific knowledge about logic gates and their mathematical, algebraic, and electronic representation that could improve their future performance.

2. Practical Application. The STEAM activity gave students the opportunity to apply their logical-mathematical reasoning skills to practical contexts and real-world problems, which may have improved their understanding and performance on the post-test.

3. Reinforcement and practice. The educational intervention included examples and exercises that allowed students to reinforce and consolidate their skills and knowledge on the topic, which was reflected in the post-test.

It is important to note that this difference in means between the pre and post-test does not necessarily indicate that the educational intervention was ineffective. On the contrary, if students demonstrate an increase in specific knowledge and understanding related to the STEAM activity on the post-test, this could be considered a positive and desirable outcome in terms of learning. Furthermore, the results may be useful for adapting pedagogical approaches, providing additional support to students who are struggling, and promoting meaningful and effective learning of logic gates and digital logic in secondary education.

On the other hand, in the existing bibliography with the same purpose, such as [17, 18], they do not provide specific information on how these teaching-learning sequences can be adapted to different educational levels. This is done implicitly, inferring that the STEAM strategy that has different knowledge domains automatically has the capacity to adapt the content to different educational levels, which is not correct. It is important to keep in mind that the adaptation of a specific topic such as logic gates or any other topic should not be presented in its complex representation to students who have not developed a certain level of maturity, giving importance to the gradual characteristic of the GMM.

5 Conclusions

It should be noted that this study was conducted in a specific context and with a limited sample of participants. Therefore, the findings may not be generalizable to other populations or educational contexts. Although positive and significant effects of STEAM activities on participants' learning and skill development were found, it is possible that other factors not measured in this study may have contributed to these results, such as student motivation, instructional quality, technology acceptance, and content presentation. Therefore, further research using the GMM is needed to confirm the results of this study and to further explore the factors that contribute to the successful implementation of STEAM education in educational institutions.

Based on the data obtained from the t-Student test, we can affirm that the implemented STEAM activity had a positive and significant effect on the learning and skill development of the participants. The reduction in the mean of the post-test compared to the pre-test suggests that the students acquired new knowledge and skills relevant to the activity, which confirms the effectiveness of the orchestration of the STEAM activity through the GMM and its improvement of the students' performance on the assessed topic.

In addition, it was possible to determine the level of knowledge of the students through the pre-test and post-test to measure the logical-mathematical thinking, level of understanding, practical skills, and perceptions of the students after the STEAM activity. The results of these questionnaires were analyzed to determine if there were statistically significant differences between the scores. Regarding the sustainability of STEAM education in educational institutions, for the time being, further research with the GMM is suggested to further explore the factors that contribute to the success of STEAM education implementation, at least in this work, most of the challenges that have to do with the acquisition of materials can be solved with careful analysis and planning prior to an investment in technology. Finally, students expressed their opinions about the STEAM activity and perceived benefits through some open-ended questions, highlighting a sense of active involvement and relevance in their learning, as these activities are usually practical and meaningful, just as they appreciated the opportunity to apply academic concepts in a real context, which allows them to see the usefulness and concrete application of what they are learning. These attitudes, along with collaboration and team problem solving, also fostered a sense of community and an appreciation of diverse perspectives; overall, the activities orchestrated by the GMM tend to generate a lasting enthusiasm for learning and exploration in these areas, which can have a positive impact on students' future academic and career choices.

Future work Given that the STEAM activity had a positive and significant effect on learning, we make the following recommendations for future work: 1) Expand the sample, increasing the size will allow us to obtain greater representativeness and robustness in the results, providing greater confidence in generalizing the findings to a larger population of students; 2) Conduct long-term follow-up, this will allow us to evaluate whether the positive effects of

the STEAM intervention are maintained over time or if there are long-term learning effects; 3) Conduct more interventions using the GMM; analyzing and comparing different types of STEAM interventions using the GMM would help determine if they are effective with the five disciplines in terms of learning and specific skill development, as well as help identify more efficient and effective pedagogical approaches.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment We are very grateful to the Escuela Secundaria Técnica 34 "Luis V. Massieu" (09DST0034Q), for allowing us to work with the high school group.

References

[1] M. Flores-Nicolás, M. Martínez, "The Graded Multidisciplinary Model: Fostering instructional design for activity development in STEM/STEAM education," in 2022 IEEE Mexican International Conference on Computer Science (ENC), 1–7, 2022, doi:10.1109/ENC56672.2022.9882917.

[2] D. Aguilera, J. Ortiz-Revilla, "STEM vs. STEAM Education and Student Creativity: A Systematic Literature Review," Education Sciences, **11**(7), 2021, doi:10.3390/educsci11070331.

[3] M. Conde, Rodríguez, C. Fernández-Llamas, J. Gonçalves, J. Lima, F. J. García-Peñalvo, "RoboSTEAM Project Systematic Mapping: Challenge Based Learning and Robotics," in 2020 IEEE Global Engineering Education Conference (EDUCON), 214–221, 2020, doi:10.1109/EDUCON45650.2020.9125103.

[4] G. Ozkan, U. Umdu Topsakal, "Exploring the effectiveness of STEAM design processes on middle school students' creativity," International Journal of Technology and Design Education, **31**, 95–116, 2021.

[5] A. Leavy, L. Dick, M. Meletiou-Mavrotheris, E. Paparistodemou, E. Stylianou, "The prevalence and use of emerging technologies in STEAM education: A systematic review of the literature," Journal of Computer Assisted Learning, **39**(4), 1061–1082, 2023, doi:https://doi.org/10.1111/jcal.12806.

[6] M. Flores-Nicolás, M. Martínez-Reyes, "A Theoretical-Practical Case Study on the Graded Multidisciplinary Model: Training of High School Students Through STEAM Education," in Streamlining Organizational Processes Through AI, IoT, Blockchain, and Virtual Environments, 194–218, IGI Global, 2023, doi:10.4018/978-1-6684-8639-9.ch010.

[7] L. Ma, H. Luo, X. Liao, J. Li, "Impact of Gender on STEAM Education in Elementary School: From Individuals to Group Compositions," Behavioral Sciences, **12**(9), 2022, doi:10.3390/bs12090308.

[8] S. Aydın Gürler, O. Kaplan, "Attitudes Towards STEAM, Critical Thinking Disposition and Decision-Making Skills: Mediation and Gender Moderation," Journal of Contemporary Educational Research, **10**(1), 210–223, 2023.

[9] N. Arís, L. Orcos, "Educational Robotics in the Stage of Secondary Education: Empirical Study on Motivation and STEM Skills," Education Sciences, **9**(2), 2019, doi:10.3390/educsci9020073.

[10] Y. Rahmawati, A. Ridwan, T. Hadinugrahaningsih, Soeprijanto, "Developing critical and creative thinking skills through STEAM integration in chemistry learning," Journal of Physics: Conference Series, **1156**(1), 012033, 2019, doi:10.1088/1742-6596/1156/1/012033.

[11] E. Yilmaz Ince, M. Koc, "The consequences of robotics programming education on computational thinking skills: An intervention of the Young Engineer's Workshop (YEW)," Computer Applications in Engineering Education, **29**(1), 191–208, 2021, doi:10.1002/cae.22321.

[12] D. Herro, C. Quigley, "Exploring teachers' perceptions of STEAM teaching through professional development: implications for teacher educators," Professional Development in Education, **43**(3), 416–438, 2017, doi:10.1080/19415257.2016.1205507.

[13] S. Belbase, B. R. Mainali, W. Kasemsukpipat, H. Tairab, M. Gochoo, A. Jarrah, "At the dawn of science, technology, engineering, arts, and mathematics (STEAM) education: prospects, priorities, processes, and problems," International Journal of Mathematical Education in Science and Technology, **53**(11), 2919–2955, 2022, doi:10.1080/0020739X.2021.1922943.

[14] C. F. Quigley, D. Herro, A. Baker, Moving Toward Transdisciplinary Instruction: A Longitudinal Examination of STEAM Teaching Practices, 143–164, Springer International Publishing, Cham, 2019, doi:10.1007/978-3-030-04003-1_8.

[15] Z. Yang, "The challenges of personalized learning and their solutions," Chinese Science Bulletin, **64**(5–6), 493–498, 2018.

[16] M. Á. Conde, F. J. Rodríguez-Sedano, C. Fernández-Llamas, J. Gonçalves, J. Lima, F. J. García-Peñalvo, "Fostering STEAM through challenge-based learning, robotics, and physical devices: A systematic mapping literature review," Computer Applications in Engineering Education, **29**(1), 46–65, 2021.

[17] I. M. Greca, J. Ortiz-Revilla, I. Arriassecq, "Diseño y evaluación de una secuencia de enseñanza-aprendizaje STEAM para Educación Primaria," 2021, doi:10.25267/Rev_Eureka_ensen_divulg_cienc.2021.v18.i1.1802.

[18] M. Queiruga-Dios, E. López-Iñesta, M. Diez-Ojeda, M. Sáiz-Manzanares, J. Vázquez-Dorrío, "Implementation of a STEAM project in compulsory secondary education that creates connections with the environment (Implementación de un proyecto STEAM en Educación Secundaria generando conexiones con el entorno)," Journal for the Study of Education and Development, **44**(4), 871–908, 2021, doi:10.1080/02103702.2021.1925475.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 6, 57-64 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Comparative Study of J48 Decision Tree and CART Algorithm for Liver Cancer Symptom Analysis Using Data from Carnegie Mellon University

Renhe Chi*

Department of Management Information Systems, National Chengchi University, Taipei,116 Taiwan

ARTICLE INFO

Article history:
Received:22 August, 2023
Accepted:22 October, 2023
Online: 30 November, 2023

Keywords:
Liver Cancer
Machine learning
J48(Gain ratio)
CART

ABSTRACT

Liver cancer is a major contributor to cancer-related mortality both in the United States and worldwide. A range of liver diseases, such as chronic liver disease, liver cirrhosis, hepatitis, and liver cancer, play a role in this statistic. Hepatitis, in particular, is the main culprit behind liver cancer. As a consequence, it is decisive to investigate the correlation between hepatitis and symptoms using statistic inspection. In this study, we inspect 155 patient data possessed by CARNEGIE-MELLON UNIVERSITY in 1988 to prognosticate whether an individual died from liver disease using supervised machine learning models for category and connection rules based on 20 different symptom attributes. We compare J48 (Gain Ratio) and CART (Classification and Regression Tree), two decision tree classification algorithms elaborate from ID3 (Iterative Dichotomiser 3), with the Gini index in a Java environment. The data is preprocessed through normalization. Our study demonstrates that J48 outperforms CART, with an average accuracy rate of nearly 87% for the complete specimen, cross-validation, and 66% training data. However, CART has the supreme accurate rate in all samples, with an accuracy rate of 90.3232%. Furthermore, our research indicates that removing the conjunction attribute of the Apriori algorithm does not impact the results. This research showcases the potential for physician and researchers to apply brief machine learning device to attain accurate outcomes and develop treatments based on symptoms.

1. Introduction

This paper is an extension of work originally presented in 4th IEEE Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability 2022 (IEEE ECBIOS 2022) under the name of "Comparison of Decision Tree J48 and CART in Liver Cancer Symptom with CARNEGIE MELLON UNIVERSITY Data" [1].

In the year 2020, liver cancer affected more than 900,000 individuals globally, leading to over 830,000 deaths. It ranks sixth among the top ten cancers worldwide and is the primary cancer in the United States [2]. Accordingly, it is crucial to conduct research on the expression of symptoms in adult liver cancer to facilitate clinical intervention. Previous research on symptoms for Liver Carcinoma and Cancer have provided valuable insights, with physical searches of article references yielding additional findings [3]. Logic-based approach such as supervised models, including linear regression, decision trees, association learning algorithms such as Random Forest and Generalized Boosting Machines, and Support Vector Machines (SVM), are commonly employed in this field. SVMs use a polynomial kernel function and a non-probabilistic approach [4]. Decision trees are an example of a prognostic model that maps entity attributes and values. Each intersection in the tree represents an entity, each bifurcation path represents a potential attribute value, and each leaf intersection correlate with the entity value appear for by the direction from the root junction to the leaf junction. Although decision trees have a single output, multiple trees can be utilized to handle various outputs. This technology is commonly utilized in data mining [5].

In the field of liver cancer, decision tree models can be employed to categorize patients based on their symptoms and forecast the possibility of developing liver cancer or their prognosis post-diagnosis. Decision tree models can also be combined with other machine learning algorithms to enhance precision and credibility [6].

APRIORI algorithms like Random Forest and Generalized Boosting Machines are frequently utilized supervised models for liver cancer prediction [7]. Random Forest is a decision tree-based association learning technique that establish multiple decision

*Corresponding Author: Renhe Chi, 108356503@nccu.edu.tw

https://dx.doi.org/10.25046/aj080607

trees and returns the approach of the class predictions from individual trees [8]. On the other hand, Generalized Boosting Machines is an iterative algorithm that constructs a strong model by combining several weak models, with each weak model aiming to rectify the errors of the previous one [9].

2. Previous work

This section presents an overview of the main machine learning techniques utilized in the analysis of the Hepatitis dataset obtained from CARNEGIE-MELLON University in 1988, as part of the Statlog project. The outcomes of the analysis, performed with and without 10-fold cross-validation, are combined into a consolidated form for the purpose of comparison and assessment.

Table 1: The analysis outcome of hepatitis data provided major from Statlog project

Reference	Year	Method	Accuracy %	With 10 fold
Kemal Polat and Salih Güneş [10]	2007	PCA+AIRS	94.12%	Yes
Statlog project (KG) [8]	1994	21-NN, stand Manhattan	90.3	No
Statlog project (RA) [8]	1994	FSM	90	No
Statlog project (KG) [8]	1994	14-NN, stand. Euclid	89	No
Weiss & K [11]	1990	LDA	86.4	No
Weiss & K [11]	1990	CART (decision tree)	82.7	No
Weiss & K [11]	1990	MLP+backprop	82.1	No
Duch & Grudzinski [12]	1998	Weighted 9-NN	92.9± ?	Yes
Duch & Grudzinski [12]	1998	18-NN, stand. Manhattan	90.2±	Yes
Rafa Adamczak [13]	1995	FSM with rotations	89.7± ?	Yes
Karol Grudzinski [12]	1998	15-NN, stand. Euclidean	89.0±	Yes
Rafa Adamczak [13]	1995	FSM without rotations	88.5	Yes
Stern & Dobnikar [14]	1996	LDA, linear discriminant analysis	86.4	Yes
Stern & Dobnikar [14]	1996	Naive Bayes and Semi-NB	86.3	Yes
Norbert Jankowski [15]	1999	IncNet	86	Yes
Stern & Dobnikar [14]	1996	QDA, quadratic discriminant analysis	85.8	Yes
Stern & Dobnikar, std added by WD [14]	1996	1-NN	85.3± 5.4	Yes
Stern & Dobnikar [14]	1996	ASR	85	Yes
Stern & Dobnikar [14]	1996	Fisher discriminant analysis	84.5	Yes
Stern & Dobnikar [14]	1996	LVQ	83.2	Yes
Stern & Dobnikar [14]	1996	CART (decision tree)	82.7	Yes
Stern & Dobnikar [14]	1996	MLP with BP	82.1	Yes
Stern & Dobnikar [14]	1996	ASI	82	Yes
Stern & Dobnikar [14]	1996	LFC	81.9	Yes
Rafa Adamczak [12]	1995	RBF (Tooldiag)	79	Yes
Rafa Adamczak [12]	1995	MLP+BP (Tooldiag)	77.4	Yes

This study presents a comprehensive comparative analysis of machine learning methods conducted by various authors, including M. Ramassamy, S. Selvaraj, M. Mayilvaganan, and Bascil & Temurtas. Notable methodologies include PCA+AIRS with 94.12% accuracy rate by Kemal Polat and Salih Güneş [10]. Other accuracy rates including 21-NN (stand. Manhattan) with 90.3%, FSM with 90% and 14-NN with 89% by Statlog project, and LDA with 86.4%, CART (decision tree) with 82.7%, and MLP+backprop by Weiss & K with 82.1% [11]. Duch & Grudzinski applied Weighted 9-NN with 92.9%±?, 18-NN (stand. Manhattan) with 90.2±0.7, and 15-NN (stand. Euclidean) with 89.0±0.5% [12], while Rafa Adamczak employed FSM with rotations with 89.7±? and FSM without rotations with 88.5% [13]. Stern & Dobnikar utilized a diverse set of methods, including LDA (linear discriminant analysis) with 86.4% accuracy rate, Naive Bayes and Semi-NB with 86.3%, 1-NN (stand. added by WD) with 85.3%±5.4, ASR with 85%, Fisher discriminant analysis with 84.5%, LVQ with 83.2%, CART (decision tree) with 82.7%, MLP with BP with 82.1%, ASI with 82%, and LFC with 81.9% [14]. Norbert Jankowski implemented IncNet with 86% accuracy rate [15]. These findings collectively contribute valuable insights into the efficacy of distinct machine learning approaches for addressing statistical learning challenges. Within the framework of the Statlog project, machine learning methodologies have demonstrated substantial efficacy in analyzing Hepatitis data, with the PCA+AIRS model outperforming others, achieving an accuracy rate of 94.12%.

Data Mining is getting increasingly important for discovering association patterns for health service innovation and Customer Relationship Management (CRM) etc. Yet, there are deficits of existing data mining techniques. First of all, most of them perform a plain mining based on a predefined schemata through the data warehouse; however, a re-scan must be done whenever new attributes appear. Second, an association rule may be true on a certain granularity but fail on a smaller one and vise verse. Last but not least, they are usually designed to find either frequent or infrequent rules. In this paper, we are going to invent more efficient and accurate approach with novel data structure and multi-dimensional mining algorithm to explore association patterns on different granularities [16] [17].

The paper presents at first the categories of innovative healthcare services as well as the way to find new service patterns. Then, we propose a data mining approach based on Apriori Algorithm for managing such new healthcare services, including a novel data structure and an effective algorithm for multi-dimensional mining association rules on various granularities. It is proved to be very useful for discovering new service patterns, even in-frequent by considering a dimension in a flat level. The advantages of this approach over existing approaches include (1) more comprehensive and easy-to-use (2) more efficient with limited scans (3) more effective with finding rules hold in different granularity levels, e.g. Age={(1-10), (10-20)….} (4) capable of finding frequent patterns and infrequent patterns, for instant we use the algorithm in finding the blood platelet frequently used for the female with age over 60, while the blood platelet infrequently for all the patients. With this method, users can choose the full match and the relaxed match (5) low information loss rate (6) capable of incremental Mining.

3. Data pre-processing

3.1. Field attributes

The majority of the primary dataset comprises symptom information, and therefore, the values in the range are predominantly binary (i.e., negative or positive). The dataset comprises 6 numeric and 14 categorical attributes, amongst the Class attribute is ranked in accordance with the form presented in reference [18].

3.2 Data pre-processing

a. Attribute analysis

Upon importing the raw data, an intrinsic attribute analysis was conducted utilizing the WEKA software to scrutinize the data pertaining to each attribute. [19].

Table 2: Hepatitis symptoms selected by Carnegie Mellon University

	Attributes	Content	Attribute Type	Range
1	Class	Survive and Die	Nominal	(die and live)
2	Age	Age Division	Numeric	(10~80)
3	Sex	Gender Distinction	Nominal	(male and female)
4	Steroid	Steroid	Nominal	(no, yes)
5	Antivirals	Anti-Viral Drug	Nominal	(no, yes)
6	Fatigue	Fatigue	Nominal	(no, yes)
7	Malaise	Depressed	Nominal	(no, yes)
8	Anorexia	Anorexia	Nominal	(no, yes)
9	Liver Big	Enlarged Liver	Nominal	(no, yes)
10	Liver Firm	Liver Currhosis	Nominal	(no, yes)
11	Spleen palpable	Enlarged Spleen	Nominal	(no, yes)
12	Spiders	Arachnoid Membrane(Spider Nervus)	Nominal	(no, yes)
13	Ascites	Ascites	Nominal	(no, yes)
14	Varices	Venous Flexion	Nominal	(no, yes)
15	Bilirubin	Bilirubin	Numeric	0.39, 0.80, 1.20, 2.00, 3.00, 4.00
16	Alk phosphate	Alkaline Phosphatase	Numeric	33, 80, 120, 160, 200, 250
17	Sgot	Aminotransferase	Numeric	13, 100, 200, 300, 400, 500,
18	Albumin	Albumin	Numeri	2.1, 3.0, 3.8, 4.5, 5.0, 6.0

b. Data preprocessing

Prior to data analysis, data preparation is a necessary step which involves data preprocessing and data reduction. The primary goal of data preprocessing is to address impure, incomplete or inconsistent data within the original dataset. Meanwhile, data reduction aims to decrease the volume or dimensionality of the initial data, for the purpose of alleviate the burden of data exploration. The significance of data preparation is demonstrated in the following scenarios: Firstly, the data may contain noise, such as errors or outliers resulting from issues with data collection equipment, human or computer errors during data recording or transmission, etc. Secondly, the data may be incomplete, with some attribute values missing due to reasons such as unnecessary data being excluded during recording or inconsistent records being deleted. Finally, inconsistency may also arise when the same data has multiple conflicting conditions, for instance, when the data is integrated from various sources with different naming conventions.

Data conversion is a critical step in data preprocessing that seeks to transform missing or inaccurate data into a compatible format for the exploration process. This research utilized four distinct preprocessing techniques, namely data discretization, data extreme value handling, data standardization, and data normalization, to enhance the quality of the data.

- Discretization of data

In order to absolve impoverished classification quality, continuous data is discretized to reduce the numerical allocation of the information.

- Data standardization

To normalize attribute data values and bring them into a minor and consistent range with other attribute data, various techniques can be used, such as Min-Max standardization, z-score standardization, and decimal standardization. The diagram below illustrates the attribute analysis of the initial data after standardization.

- Data normalization

Normalization is a data processing technique used to adjust the data values to a common scale or range. This is done to make the data comparable and reduce the impact of different measurement units or scales on data analysis. For instance, when comparing the annual income of customers in Taiwan and the Philippines, it is not appropriate to directly compare the income levels in Taiwan dollars, as the average income in Taiwan is much higher than that in the Philippines. Therefore, normalization is used to redistribute the data into a small and specific range, which allows for objective and meaningful comparisons.

The aforementioned data processing was carried out utilizing the configurations provided by WEKA.

c. Data reduction

Data reduction is a crucial technique that involves reducing the size or dimensions of data without significantly impacting the exploration outcomes. The main object of data reduction is to ease the burden of data exploration, reduce computation time, improve prediction accuracy, and enhance exploration outcomes' quality by removing irrelevant or unnecessary data. Information Gain, Gini Index, and $\chi 2$ independence test are common characteristic selection criteria used in data reduction. In this research, Information Gain is used as the attribute selection standard to remove attributes with the lowest direction gain value to prepare for the J48 data classification method in WEKA (Waikato Environment for Knowledge Analysis). Information Gain measures the difference between the information quantity before

and after a test, represented by the entropy value of the sub-decision tree (Entropy), which is the entropy value of the set produced by a junction with a particular characteristic as the conclusion number. However, Information Gain-based attribute selection may be biased towards attributes with more attributes, resulting in biased decisions. To address this issue, the Gain Ratio method normalizes the Information Gain by dividing it by the number of possible attribute values, preventing bias towards attributes with more qualitative attributes.

By using the conducting analysis, it was observed that attribute 9 and attribute 10 exhibit the lowest information gain and gain ratio. Consequently, these two attributes are eliminated to diminish the data dimensionality. Subsequent to data reduction, the data preprocessing procedures, comprising data discretization, standardization, and regularization, are implemented anew to finalize the data preprocessing stage.

4. Data analysis:

4.1. Classification

4.1.1. Decision tree

In the realm of machine learning, a decision tree is an example used to predict a mapping connection between characteristic and their respective values. Each junction in the tree denotes a task, and every branching path indicates a feasible characteristic value. The terminal nodes correspond to the values of the objects represented by the paths from the root junction to the terminal junctions. A decision tree has a solitary output, and to address multiple outputs, distinct decision trees can be constructed. Decision trees are a frequently employed approach in data mining for the purpose of analyzing and predicting data.

- Categorical decision tree: target variable is categorical

Categorical decision tree examination is a machine learning tactic used when the target variable is categorical in nature, such as predicting the species of a plant or the likelihood of a customer to purchase a product. This approach is implemented using various algorithms, including ID3, C4.5 (J48), and C5.0.

- Regression decision tree: target variable is continuous

Regression decision tree analysis involves the use of decision trees to predict continuous numerical values, such as the temperature or stock price. It is a widely used technique in data analysis and machine learning. Several algorithms can be used to implement regression decision tree analysis, such as CART, CHAID (Chi-Square Test), MP (multivariate polynomial) and C4.5 (Gain Ratio).

a. J48 Algorithm

J48 is a decision tree algorithm that is based on the C4.5 implementation. The creator of C4.5 later upgraded the algorithm to C4.8, which was then implemented in Java by the creators of Weka and named J4.8. The ID3 algorithm must be introduced first because the C4.5 algorithm is an improved version of ID3.

During the construction of the decision tree, the ID3 algorithm uses Information Gain as the criterion to select the attribute with the highest information gain value as the classification attribute. This algorithm is based on the principle of Occam's razor, which states that the smaller the decision tree, the preferable the exhibition. However, the ID3 algorithm is a heuristic algorithm and may not always produce the smallest tree structure. Moreover, one of the issues with ID3 is its bias towards attributes with multiple values. For example, if there is a distinctive recognition characteristic such as an ID, ID3 may choose it as the splitting characteristic. Although this creates a sufficiently clean section, it is nearly futile for classification purposes. To address this problem, the C4.5 algorithm, which is the successor of ID3, employs the gain ratio information obtain extension to reduce this bias.

b. C4.5:

C4.5 is a set of algorithms frequently utilized in machine learning and data mining for classification tasks. Specifically, its purpose is to perform supervised learning, where a dataset contains tuples characterized by attribute values and each tuple belongs to one of several exclusive categories. The aim of C4.5 is to construct a mapping function from the attribute values to categories that can be used to classify new instances with unknown categories.

J. Ross Quinlan proposed C4.5 as an extension of the ID3 algorithm, which is used to construct decision trees. A decision tree is a tree-like structure similar to a flowchart, where each internal node presents an attribute exam, each branch presents an exam outcome, and each leaf node reserves a class label. After the decision tree is constructed, an unclassified tuple can be traversed from the root node to a leaf node, which stores the predicted class label for the tuple. Decision trees are advantageous because they do not postulate any prior estate expertise or guideline settings and are appropriate for investigative comprehension discovery.

C4.5 overcomes the problems of ID3 by adopting the gain ratio of attributes, which normalizes the information gain by computing the break knowledge value of the virtue. In the C4.5 algorithm, the break apart virtue selection process does not solely rely on the virtue with the supreme gain ratio. Instead, it searches for attributes that have a direction obtain higher than the standard level among the candidate separation virtue, and then choice the virtue with the supreme gain ratio. This is because gain ratio tends to favor attributes with smaller values compared to information gain.

C4.5 has several improvements over ID3. Firstly, it can handle continuous attributes. Secondly, it uses gain ratio to overcome ID3's bias towards attributes with many distinct values but little significance. This is because the guidance obtain measure used by ID3 inclines to select virtue with many distinct values, which can lead to the creation of suboptimal decision trees. For example, if

the algorithm divides the data based on a unique attribute like student ID, it would generate numerous branches, each with only one or a few instances, resulting in a high information gain value but a meaningless split.

Table 3: J48 Results Comparison Table

Attribute	Accuracy Rate
Whole samples	90.3226%
Cross Validation	84.5161%
66% Training data	81.1321%

4.1.2. The principle of CART algorithm

As previously stated, the CART algorithm comprises two stages, and in the first stage, a binary tree is constructed recursively. The question then arises: how is the data divided?

In the field of machine learning and data mining, the CART algorithm is often employed for classification tasks, where each data point is assigned to one of several exclusive categories based on a selected attribute. The algorithm divides the multidimensional space recursively into non-overlapping rectangular regions through a process that involves selecting independent variables and partitioning the space based on the values of the selected variable. The procedure is duplicated circularly on each of the resulting areas until the entire space is covered by non-overlapping regions.

The standard for dividing the space is an important consideration in the CART algorithm. For variable attributes, the dividing point is typically determined as the middle between a pair of endless variable virtue values. The amount of adulteration that can be reduced by dividing on each attribute is then calculated and used to sort the attributes. The decrease of adulteration is explicit as the aggregate of the amount of adulteration before separates minus the amount of adulteration at each node after division. The Gini index is routinely accustomed to as a method for measuring adulteration. Gini impurity measures the probability that a given node represents a certain class, and it is minimized when all samples in the node belong to the same class.

The core concept of the CART algorithm is to recursively classify data based on a minimum distance-based Niki index estimation function. One of the main benefits of the algorithm is its simple and easy-to-understand rule extraction process. Moreover, the CART algorithm is robust against issues such as missing values and a large number of variables, making it a widely used and effective tool in machine learning and data mining.

Algorithm limitations: attribute selection is restricted to generating only two child nodes; error rate may increase rapidly with a large number of categories.

Application domains: identification of information distortion, identification of potential customers in the telecommunications industry, prediction of loan risks, and others.

c. J48 VS CART

The fundamental distinction between CART and J48 algorithms lies in the criterion used to split the nodes. CART adopts the GINI index to measure the purity of data partitions or training datasets when choosing a splitting attribute. The GINI index quantifies the purity of a sample based on the likelihood of it belonging to a specific category. Consequently, the attribute that results in the minimum GINI index is selected for division.

Table 4: CART Results Comparison Table

Attribute	Accuracy Rate
All samples	90.3226%
Cross-Validation	84.5161%
66% of Samples	81.1321%

Table 5: Classification results analysis

Algorithm	Pre-process	Test mode	Accuracy Rate
J48	Normalization	Whole sample	87.0968%
J48	Normalization	Cross Validation	85.1613%
J48	Normalization	66% Training data	88.6792%
CART	Normalization	Whole sample	90.3226%
CART	Normalization	Cross Validation	94.5161%
CART	Normalization	66% Training Data	81.1321%

4.2. Association Law (APRIORI)

Within a vast bibliography, interrelationships among specific purposes exist, commonly referred to as Market Basket Analysis, which originated from analyzing the extent of conjunction of sizeable itemsets in merchandise case data. This analysis utilizes the law of association, frequently used in the study of shopping baskets, to examine the correlation between purchased products in customer acquire data recorded by the POS system. An exemplary illustration of the law of association is the well-known paradigm of the correlation between beer and diapers.

Association rules can be conveyed in the form of X => Y [Support, Confidence], where T={t1, t2, …, tm} represents the set of all items, X⊂T, Y⊂T, and X∩Y=Φ. Here, X and Y denote unique data item sets in the transaction set T, indicating that if X emerge, Y may also emerge concomitantly. The assist of an association rule X Y in T, performed by support (X Y) = P(X Y) = s, is the ratio of agreement including X∩Y to all agreement in T. The accreditation of an association rule X Y in T, served by

confidence (X Y) = P (Y|X) = c, is the ratio of agreements including X∩Y to agreements including X in T. The values of support and confidence range between 0 and 1.

An item set (itemset) refers to a collection of distinct items, such as {A, B, C} in a record, which can produce the item group {A}, {B}, {C}, {A, B}, {B, C}, {A, C}, {A, B, C}. If there are n items in the item set, the item group comprises 2n-1 items. Therefore, when managing a large number of items, the item group can be extensive.

The support (Support) of an item set in the database is the frequency of the item set in the database, typically denoted by Support (X), where X is itemset. For instance, if a database has 100 transaction records, and 40 of them indicate the purchase of milk, then the support of this item set is 40/100=40%. The higher the support, the more crucial the item set is for further exploration.

Confidence (Confidence) signifies the level of trust between two itemsets and is represented by the conditional probability that Y will appear under the probability of X appearing, usually expressed as Support (X∩Y) /Support (X), where X and Y are itemsets.

In the realm of data mining, a robust association rule is characterized by X => Y, a rule that can be established in the transaction set T if it satisfies two conditions: support (A => B) >= min_sup and confidence (A => B) >= min_conf, where min_sup and min_conf indicate the minimum thresholds for support and confidence, respectively. When both of these criteria are met, the rule X => Y is deemed a robust association rule within the transaction set T.

In evaluating association rules, two critical criteria must be met:

The rule should identify unanticipated and unintended associations.

The rule should be capable of making an impact.

One well-known algorithm for generating association rules is Apriori. It employs a bottom-up, iterative approach to identify high-frequency item sets by breading and examination applicant item sets. From these high-frequency item sets, the algorithm identifies useful association rules.

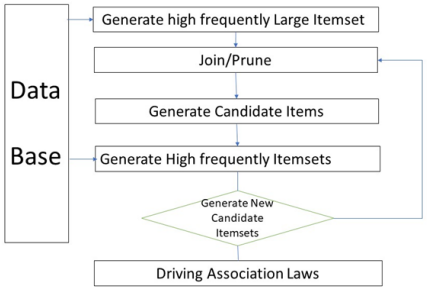


Figure 1: The Apriori algorithm

The Apriori algorithm involves two main steps:

Discovering Large itemsets from the transaction items in the database

The goal is to identify frequent Large itemsets, which requires repeated searches of the database. As Large itemsets have the property that all of their subsets are also frequent, the algorithm generates new sub-itemsets using join and prune operations.

Generating association rules based on the discovered Large itemsets

The Large itemsets obtained in step one are used to derive meaningful association rules. A rule is considered meaningful only if its confidence exceeds the minimum confidence threshold (Min Confidence).

Apriori algorithm process:

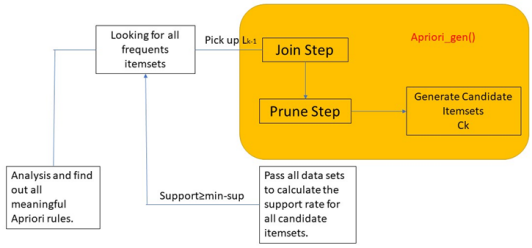


Figure 2: The Apriori process diagram

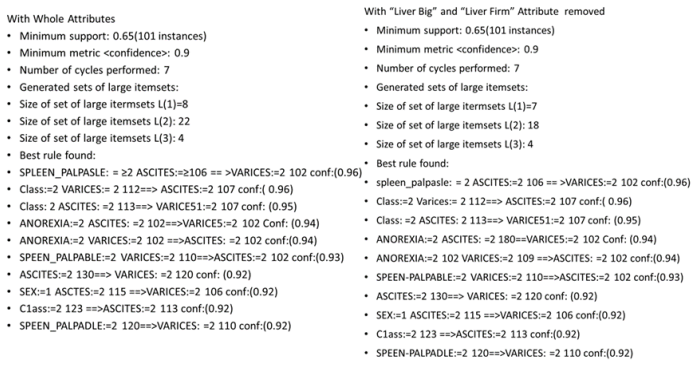


Figure 3: The Apriori result before and after "Liver big" and "Liver Firm" attribute removed.

In the APRIORI algorithm, the first pass through the database is employed to determine the Large 1-itemsets.

For subsequent passes, the algorithm is composed of two stages:

- In the first stage, the Apriori-gen function is utilized to generate new candidate itemsets Ck from the previously discovered Large itemsets Lk-1.

- In the second stage, the database is examined to calculate the Support value of the candidate itemsets in Ck.

The following is the result of the APRIORI algorithm based on our data, before removing “Liver Big” and “Liver Firm” in the APRIORI attribute.

Based on the two datasets provided and the outcomes of the association analysis, it is discernible that a substantive correlation between the "Liver Big" and "Liver Firm" attributes and other pertinent attributes appears to be lacking. The findings are expounded as follows:

In the initial dataset, encompassing both the "Liver Big" and "Liver Firm" attributes, the derived association analysis results are expounded as follows:

The cardinality of the generated large itemsets: L(1)=8, L(2)=22, L(3)=4

Optimal rules identified: Diverse rules, exemplified by instances such as spleen_palpable≥2, Ascites≥106 → Varices=2 conf:(0.96), and comparable formulations.

Conversely, upon the exclusion of the "Liver Big" and "Liver Firm" attributes from the dataset, the ensuing association analysis outcomes are delineated as follows:

The cardinality of the generated large itemsets: L(1)=7, L(2)=18, L(3)=4

Optimal rules identified: Analogous to those observed in the primary dataset, including instances like spleen_palpable=2, Ascites=2 → Varices=2 conf:(0.96), alongside other commensurate rules.

In light of these results, the following rationales can be adduced:

Scarcity of Substantive Rules: In both datasets, conspicuous absence of significant rules directly associating the "Liver Big" and "Liver Firm" attributes with other attributes is noticeable. This indicates a limited propensity for these two attributes to interact significantly with the remaining attributes in the datasets.

Attribute Sparse Occurrence: The rare occurrence of instances wherein the "Liver Big" and "Liver Firm" attributes co-occur with other attributes might be attributed to data scarcity. This scarcity may engender challenges in discerning robust associations between these attributes.

Threshold Specification: The stipulated thresholds for minimum support and confidence, set at 0.65 and 0.9 respectively, might inadvertently sift out associations characterized by lower frequencies and confidence levels. Given the presumed low-level associations of "Liver Big" and "Liver Firm" attributes, adherence to the specified thresholds could preclude their inclusion in the derived association rules.

Data Profile Dynamics: The outcomes are also liable to be influenced by data profile intricacies and distribution patterns. In instances where the "Liver Big" and "Liver Firm" attributes do not manifest as prominent co-occurring features within the dataset, the association analysis might struggle to identify substantial relationships.

In summation, predicated on the proffered datasets and the contextual framework of the association analysis, the dearth of observable significant associations between the "Liver Big" and "Liver Firm" attributes and other pertinent attributes is discernible. This, however, does not conclusively imply a universal lack of connection; rather, it underscores the paucity of apparent associations within the existing conditions and dataset parameters.

5. Discussion and future study

This investigation is fundamentally grounded in the amelioration of machine learning techniques, as opposed to adopting traditional statistical methods. Additionally, it has been observed that mixed methods generally yield higher accuracy levels, thereby substantiating the selection of J48 and the Gini index-based CART algorithm as apt methodologies for this particular study.

It is imperative to acknowledge that both machine learning and AI are continuously evolving fields, and with access to an augmented sample size and the elucidation of additional attributes, there is a potential for even more exemplary performance and a more meticulous analysis.

The Apriori analysis conducted revealed a low correlation between the attributes "Liver Big" and "Liver Firm," indicating that their removal does not impact the final results significantly. For analyzing relationships such as the variations in age groups, we recommend employing our Multi-dimensional Multi-granularities Data Mining based on the Apriori Algorithm. This approach enables the segmentation of patient ages into various granularities, specifically {(10-20), (20-30), …, (70-80)}. Subsequently, we can mine for association patterns within these defined segments, ensuring that phenomena pertinent to children do not get erroneously associated with adults. However, upon constructing data cubes for age ranges (10-20), (60-70), and (70-80), we may uncover associations within these specific segment combinations or granularities.

6. Conclusion

The consolidation of machine learning with the therapeutic realm presents numerous advantages, such as an improved understanding of disease characteristics and the potential to aid healthcare providers in developing more efficient treatment strategies for patients. Machine learning finds application in diverse areas within the medical sector, not just limited to the employment of qualitative and quantitative material categorization to draw inferences, and association rules to establish links between manifestation. For example, in the domain of oncology, machine learning is utilized in supervised therapeutic photo and quantitative data-based congregate to determine if a tumor is hostile. Furthermore, deep learning and computer vision technologies aid in detecting brain tumors. These advancements are indicative of the maturing machine learning

applications in medical treatment. With easily accessible tools, physicians and researchers can obtain precise results and prescribe appropriate medication for symptom management, while the principle population can adopt this knowledge to prevent and improve recognize diseases. The medical field anticipates the emergence of additional machine learning and data mining utilization in the future, extending beyond the treatment of hepatitis.

References

[1] J. K. Chiang and R. Chi, "Comparison of Decision Tree J48 and CART in Liver Cancer Symptom with CARNEGIE-MELLON UNIVERSITY Data," 2022 IEEE 4th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS), Tainan, Taiwan, 28-31, 2022, doi: 10.1109/ECBIOS54627.2022.9945039.

[2] Liver cancer deaths in 2020 approaching incidence, https://www.cn-healthcare.com/articlewm/20210115/content-1180778.html.

[3] M. E. Cooley, "Symptoms in adults with lung cancer: A systematic research review," Jounal of Pain and Symptom Management, **19**(2), February, 2000.

[4] C. M. Lynch, "Prediction of lung cancer patient survival via supervised machine learning MARK classification techniques," International Journal of Medical Informatics 108, 1-8, 2017.

[5] Z. Mahmoodabai, S. S. Tabrizi, "A new ICA-Based algorithm for diagnosis of coronary artery disease," Intelligent Computing, Communication and Devices, **2**, 415-427, 2014.

[6] Datasets used for classification comparison of results. https://www.is.umk.pl/~duch/projects/projects/datasets.html#Hepatitis

[7] M. Hegland, The APRIORI Algorithm—A Tutorial, https://www.worldscientific.com/doi/abs/10.1142/9789812709066_0006

[8] D. Michie, D.J. Spiegelhalter, C.C. Taylor, "Machine Learning, Neural and Statistical Classification," Ellis Horwood Series in Artificial Intelligence: New York, NY, USA, **13**, 1994.

[9] S. Touzani, J. Granderson, S. Fernandes, "Gradient boosting machine for modeling the energy consumption of commercial buildings," Energy and Buildings, 158(1533-1543), 2018, doi: 10.1016/j.enbuild.2017.11.039

[10] K. Polat, S. Güneş, "Hybrid prediction model with missing value imputation for medical data, Expert Systems with Applications," **42**(13), 5621-5631, 2015

[11] S.M. Weiss, I. Kapouleas, "An empirical comparison of pattern recognition, neural nets and machine learning classification methods," Department of Computer Science, Rutgers University, New Brunswick, NJ 08903, 1989

[12] W. Duch, K. Grudzi´nski, "Weighting and selection of features," Intelligent Information Systems VIII, Proceedings of the Workshop held in Ustro´n, Poland, 1999

[13] N Jankowski, A Naud, R Adamczak, "Feature Space Mapping: a neurofuzzy network for system identification,", Department of Computer Methods, Nicholas Copernicus University, Poland, 1995

[14] B. Stern and A. Dobnikar, "Neural networks in medical diagnosis: Comparison with other methods," Proceedings of the International Conference EANN, **96**, 427-430, 1996.

[15] Norbert Jankowski, "Approximation and Classification in Medicine with IncNet Neural Networks," Department of Computer Methods Nicholas Copernicus University ul. Grudziądzka 5, 87-100, Toruń, Poland, 1999

[16] J. K. Chiang, C. C. Chu, "Multi-dimensional multi-granularities data mining for discovering innovative healthcare services," Journal of Biomedical Engineering and Medical Imaging, **1**(3), 214, DOI: 10.14738/jbemi.13.243

[17] J. K. Chiang, C. C. Chu, "Multidimensional multi-granularities data mining for discover association rule," Transactions on Machine Learning and Artificial Intelligence, **2**(3), 2014.

[18] Hepatitis Data Set. https://archive.ics.uci.edu/ml/datasets/Hepatitis

[19] Weka website. https://www.cs.waikato.ac.nz/~ml/weka/.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 8, No. 5, 62-69 (2023)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Modeling Control Agents in Social Media Networks using Reinforcement Learning

Mohamed Nayef Zareer*, Rastko Selmic

Department of Electrical and Computer Engineering, Concordia University, Montreal, H3G1M8, Canada

ARTICLE INFO

Article history:
Received: 01 August, 2023
Accepted: 29 September, 2023
Online: 30 October, 2023

Keywords:
Markov Decision Process
Multi-agent Systems
Q-Learning
Opinion Dynamics
Social Network Analysis

ABSTRACT

Designing efficient control strategies for opinion dynamics is a challenging task. Understanding how individuals change their opinions in social networks is essential to countering malicious actors and fake news and mitigating their effect on the network. In many applications such as marketing design, product launches, etc., corporations often post curated news or feeds on social media to steer the users' opinions in a desired way. We call such scenarios opinion shaping or opinion control whereby a few selected users, called control users, post opinionated messages to drive the others' opinions to reach a given state. In this paper, we are interested in the control of opinion dynamics in social media using a combination of multi-agent systems and Q-learning. The social media environment is modeled with flexible multi-agent opinion dynamics that can capture the interaction between individuals in social media networks using a two-state updating mechanism. The environment is formulated as a partially observable multi-agent Markov decision process. We propose using intelligent reinforcement learning (RL) agents to control and shape the social network's opinions. We present the social media network as an environment with different kinds of individuals and connections and the influencing agent as an RL agent to influence the network.

1 Introduction

Social media has received widespread attention due to its rise as one of the most essential tools for societal interaction and communication. Social media platforms have significantly contributed to the rapid dispersion of news and information and the facilitation of communication between different groups worldwide. Nevertheless, in the past few years, there has been a notable surge in the spread of misinformation, a trend propelled and magnified by the influence of social media platforms such as Facebook and Twitter. Due to this, it has become increasingly vital to understand opinion dynamics in social networks to minimize the detrimental effects of malicious agents, fake news, and other polarizing factors. This work builds on the work presented in [1], where we discussed how updating dynamics on social media and competition for influence affect the overall opinion and how intelligent agents decrease polarization and disagreement in social media networks.

Extensive research has been conducted to examine and understand this issue of misinformation in social media networks [2]–[4]. The prevailing trends to combat this issue are automated tools for detecting fake news and misleading posts [5]. However, this presents multiple challenges, such as defining what is fake and true. For example, simply defining what is true and false can generate a lot of discussion and controversy. This is compounded by the availability of artificial intelligence tools that can mimic a human's voice or generate fake video recordings. Additionally, misinformation in social media can result in polarization, economic impacts, time and resource wastage, and cybersecurity concerns.

Furthermore, understanding the effect of RL on influencing opinions in a social media environment is vital for several avenues such as public opinion management [6], policy implementation [7], commercial interests [8], platform design [9], and Ethical Considerations [10].

This paper proposes an RL-based method for influencing and controlling opinions in a social media environment. Unlike what is presented in the literature, where the model or algorithm is static and does not react to changes in the social network, our method proposes that each control agent can react and take action using Deep Q-Networks (DQN) to influence and persuade the other agents in the network to adopt a similar opinion. This paper expands the work in [1] with the following contributions:

1. The control agents can change their expressed opinions to influence the social media network.

2. The environment is a two-state expressed and private opinion dynamics model with asynchronous and synchronous updat-

*Corresponding Author: Mohamed Zareer, Canada, 514-3782524 & mngasem1990@gmail.com

ing dynamics that emulate the interactions of individuals in social media networks.

3. The agents work to influence the opinions of others in the network to a set goal.

4. Simulation results that demonstrate our approach for multiple social media networks.

The rest of the paper is organized as follows. Related works and preliminaries are provided in Section 2. We introduce the Markov decision process (MDP), partially hidden MDPS, and RL in Section 3. In Section 4, we formulate the problem and present the paper's main results. The experimental results are given in Section 5. Concluding remarks are given in Section 6.

2 Related Work

Social media networks have become an integral part of our society. This resulted in widespread attention to opinion dynamics and social network analysis. Opinion dynamics studies how opinions, beliefs, and attitudes form, evolve, and interact within social networks and communities. Opinion dynamics encompasses a variety of disciplines, such as sociology, political science, complex systems analysis, psychology, and multi-agent systems. One of the essential assumptions in opinion dynamics is that opinions in social networks are mainly influenced by others.

Many opinion dynamics have been proposed to study the evolution of opinions in social networks [11]–[13]. One of the most popular opinion dynamics models is agent-based models [14]. In agent-based models, individuals are depicted as agents, and their opinions on a specific topic are captured as evolving real values over time. The underlying communication network in agent-based models is represented by a graph, where a node represents an agent, and an edge represents communication between two individuals.

One of the earliest agent-based models is the French-DeGroot model, commonly referred to as the DeGroot model [15]. The model assumes that the opinion of each individual evolves as a result of integrating the opinions of their connected neighbors with agents' own opinions using weighted averaging (modeled using a differential or a difference equation). The DeGroot model was experimentally validated in [16, 17]. An extension of the DeGroot model is the Friedkin-Johnsen model, which simulates strong diversity due to stubborn agents by introducing a variable that measures the agent's susceptibility to social influence [18]. This model has been experimentally verified for small and medium-sized networks [19]–[21]. A model that encodes the effects of social pressure on the agents in the network, termed the expressed and private opinion (EPO) dynamics model, was introduced in [22]. The EPO model is based on Asch's conformity experiments and Prentice and Miller's field experiments on pluralistic ignorance [23, 24]. An extension of the EPO model is the asynchronous and synchronous expressed and private opinion dynamics model (EPOAS) [25]. This model introduces different updating dynamics for the agent's expressed and private opinions, which emulates the interaction of individuals in social media networks.

These models tend to reach a consensus if all the agents in the network agree on one opinion on a specific topic. However, it can be observed that in most social networks, the presence of stubborn (who insist on their own opinion) or controlling agents has a significant impact on the opinions of the other agents in the network [26]. This trend is evident in social media marketing, economics, and political campaigns [27]. Moreover, numerous political and economic entities utilize data mining techniques and social science principles to strategically engage specific individuals within social networks, aiming to enhance profit margins [28].

There have been many attempts to study and simulate opinion control in social networks. The control of the DeGroot model under the influence of a leader was investigated in [29]–[31]. In [32], the authors study the optimal placement of control agents in a social network to influence other individuals to reach a consensus where the control agents have a fixed common state. Another control approach using noise to affect the opinions of individuals in the network was investigated in [33]. A control strategy based on the degree of connection each agent has shown that it is possible to drive the overall opinion toward a desired state even if we control only a suitable fraction of the nodes was presented in [34].

In [35], the control of public opinion using social bots was investigated. In this approach, an agent's opinion is modeled as a static value, based on the approach described by Sohn and Geidner [36]. The study demonstrates that, depending on the density and position within the network, a mere $2\% - 4\%$ of bots are sufficient to influence all the opinions in the network.

3 Markov Decision Process

An MDP is a mathematical framework designed to model decision-making where sequential actions are involved, and the outcomes of each action are partially random and partially under the control of the decision-making agent [37]. An MDP was developed to model decision-making problems where the outputs are probabilistic and are affected by the agent's actions. An MDP is modeled by a tuple (S, A, P_a, R_a) where:

1. States (S) encapsulate all potential scenarios the agent could encounter at any given time step.

2. Actions (A) embody all the options or decisions accessible to the agent in any given state. Agents select an action to transition from one state to another.

3. Transition probabilities (P_a) specify the likelihood of transitioning to a new state. These probabilities represent the dynamics of the environment and determine how the agent's actions influence the next state.

4. Rewards (R_a) quantify the desirability of taking a specific action in the current state. The agent's goal is to maximize the rewards it receives to reach its desired goal efficiently.

The goal of an MDP is to have the agent learn a good policy for decision-making. A policy $\pi(S)$ can be defined as a strategy that maps states to actions, indicating the action the agent should choose in each state to maximize the agent's reward and help the agent reach their goal.

3.1 Partially Observable MDP

Partially Observable Markov Decision Processes (POMDPs), Figure 1, provide a robust framework for effectively modeling and resolving decision-making challenges in contexts marked by uncertainty and partial observability. A POMDP represents a decision-making scenario for an agent. In this scenario, it is assumed that the system's behavior is governed by a MDP. However, the agent is unable to perceive the inherent state of the system directly.

Decision-making in real-world environments often involves inherent uncertainty and partial observability, where agents lack complete information about the underlying states and face uncertain outcomes from their actions. The framework of POMDPs is sufficiently versatile to represent a wide array of sequential decision-making scenarios encountered in the real world.

In POMDPs, an agent makes decisions based on a belief state (a probability distribution over all possible states) rather than the actual state. The agent's belief state is updated based on the actions it takes and the observations it receives. The agent's goal is to choose actions over time to maximize its expected cumulative reward.

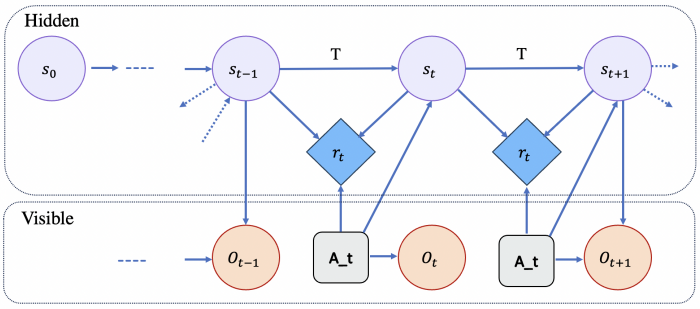


Figure 1: Partially Observable Markov Decision Processes (POMDPs).

3.2 Reinforcement Learning (RL)

One of the most powerful and widely used machine learning tools in data science is RL. The ability of RL algorithms to learn about the environment and generate a suitable policy, which can be improved through trial and error, has made them a popular choice for tasks such as game playing, autonomous driving, robotics, and resource management, among others. One of the most essential RL algorithms is the Q-learning algorithm. This can be attributed to the algorithm's versatility, simplicity, and robustness.

Q-learning is a model-free RL algorithm, meaning it does not require knowledge of the environment's dynamics to generate an efficient policy for solving the problem. This characteristic makes the algorithm particularly suitable for problems where a model of the environment is challenging to obtain or the environment is non-deterministic.

The essence of Q-learning is learning the action-value function (Q-function). The Q-function evaluates the value of taking a specific action in a given state

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma\, max_a Q(s',a') - Q(s,a)], \qquad (1)$$

where s represents the agent's present state, s' represents the agent's next state, and a is the action taken by the agent in the current time step. The Q-value of the action a for the current state s is denoted by $Q(s,a)$, which is the state-action value. The reward the agent receives at each time step is represented by r, and γ is the discount factor that reduces the value of future rewards over time.

The value $Q(s,a)$, for the current state s, is updated at every time step based on a blend of the existing value and the equation that identifies the optimal action in the current state. Initially, the Q-value table is randomly populated for each state and potential action. The Q-learning process continues by updating the Q-value for each state using (1). The policy is then updated using the highest Q-values for each state-action pair. After the agent performs an action a in state s, it transitions to the next state s'. This procedure is repeated multiple times until the overall Q-values reach a point of convergence [38]. The algorithm for Q-learning is described below.

Algorithm 1: Q-learning for estimating $\pi \approx \pi^*$

Algorithm parameters: step size $l_r \in (0,1]$ ε, γ;
Initialization: $Q(s,a)$ *for all* $s \in S, a \in A$
for *episode* $\leftarrow$ *episodes* **do**
  initialize s
  while *not done* **do**
    Choose a from s using policy derived from Q (e.g., $\varepsilon - greedy$) ;
    Take action a and observe the reward r, and next state s';
    $Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma\, max_a Q(s',a') - Q(s,a)]$,
    $s \leftarrow s'$
  end
end

The learning rate l_r needs to be fine-tuned over time to solidify the learned policy. Discount factor γ is a numerical value between 0 and 1 that determines the importance of future rewards compared to immediate ones. If γ is close to 0, then the agent will prioritize immediate rewards and largely ignore future rewards. However, if γ is close to 1, then the agent will consider future rewards almost as important as immediate ones.

4 Problem formulation

Consider a group of n agents interacting in a social media environment, where agents form and evolve their opinions at each time step by interacting and exchanging information with other connected agents.

An independent agent or a group of agents is arbitrarily selected to control and influence the network to a desired outcome. The agents can have the same goal and work together to achieve that goal or they can have different goals and compete for influence in the network.

4.1 Modeling the Agents

We model n agents as nodes interacting in a multi-agent opinion dynamics environment. The agents are interconnected via an underlying graph $\mathcal{G}[W]$, which maps out the relationships between these agents. The underlying graph $\mathcal{G}[W]$ is a directed graph that encodes the flow of information from one agent to another. Each individual

i in the network has two states: one state represents the individual's expressed opinion $\hat{y}_i(t)$, and the second state is the individual's private opinion $y_i(t)$. Each agent within the network starts with an initial private opinion, denoted as $y_i(0)$, and an initially expressed opinion, represented as $\hat{y}_i(0)$ (for the remainder of this paper, the terms 'state' and 'opinion' will be used interchangeably). The initial private and expressed opinions of each agent can either coincide ($y_i(0) = \hat{y}_i(0)$) or differ ($y_i(0) \neq \hat{y}_i(0)$). This mirrors real-world scenarios where individuals may choose to voice their genuine opinions or, due to external influences, conceal their private views by expressing differing opinions. At every time step t, each individual i in the network has two choices encoded by the variable α. The individual can express his/her opinion ($\alpha_i = 1$) or hide the opinion ($\alpha_i = 0$). If an individual decides to hide his/her opinion, then the last expressed opinion made by that individual will appear as his/her current expressed opinion to his/her neighbors. Note that for control agents $\alpha(t) = 1 \forall t$. At every time step, all individuals in the network, except control agents, update their private opinions using the following dynamics

$$y_i(t+1) = \lambda_i w_{ii} y_i(t) + \lambda_i \sum_{j \neq i} w_{ij} \hat{y}_j(k) + (1-\lambda_i) y_i(0), \tag{2}$$

where $\lambda_i \in [0,1]$ represents the coefficient of susceptibility of agent i, $w_{ij} \geq 0$ represent the weights assigned by agent i to agent j, $w_{ii} \geq 0$ represents the weight assigned by agent i to his/her opinion.

The expressed opinion is updated asynchronously, utilizing the information in the vector $\alpha(t) \in [0, N]$. This vector encodes the decisions made by each agent at the time step t. If agent i opts to express their opinion, then the update of their expressed opinion adheres to the subsequent dynamics

$$\hat{y}_i(k+1) = \begin{cases} \phi_i y_i(t) + (1-\phi_i)\hat{y}_{avg}(t) & \alpha_i = 1 \\ \hat{y}_i(t) & \alpha_i = 0 \end{cases}, \tag{3}$$

where ϕ represents the resilience of the agent to social pressure and $\hat{y}_{avg}(t)$ represents social pressure or the prevailing opinion (*public opinion*) throughout the network. The average opinion, as observed by agent i, is given by the following dynamic

$$\hat{y}_{i,avg}(t) = \sum_{j \neq i} m_{ij} \hat{y}_j(t), \tag{4}$$

where $m_{ij} \geq 0$ satisfy $\sum_{j=1}^{n} m_{ij} = 1$. In many instances, w_{ij} and m_{ij} are not identical. This occurs as an individual's viewpoint can be shaped and influenced by a certain group of individuals, all while concurrently feeling the urge to align with the expectations of the overall network.

Note that control agents do not follow the same updating dynamics and do not have a private opinion. The expressed opinions of the control agents are controlled by the Q-learning dynamics based on the optimal policy learned by the agent.

4.2 Communication Topology

The communication topology for n individuals (agents) can be represented as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}[W], W)$, where $\mathcal{V} = \{v_1, \ldots, v_n\}$ is the set of nodes (which represent agents or individuals in the network), and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of ordered edges, $\mathcal{E} = \{e_1, \ldots, e_n\}$. Each edge in the network is denoted as $e_{ij} = (v_i, v_j) \in \mathcal{E}$. With N_i we denote the neighbor set for the agent i (set of agents connected to the agent i).

The relative influence matrix of the network is modeled by W. The influence matrix W encodes how much trust or weight each agent has in the opinions of his/her connected neighbors. It is assumed that the influence matrix is static and connected. However, the communication network changes at each time step based on the updating dynamics of the agents in the network.

Each agent can only observe the expressed opinions of other connected agents. This means that the Q-table depends on the number of individuals connected to the control agents rather than the size of the network. The agent can have an idea of the overall opinion of the network based on the M weight matrix in (4). However, in this work, we assume that $M = W$. The observation space depends on the connections (neighbors) of the control agents. For example, if the control agent has three connections, the agent will have an observation space of three expressed opinions, as shown in Figure 2.

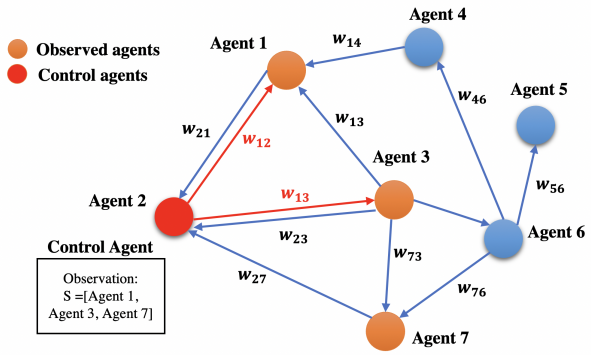


Figure 2: Observation of a control agent.

Some key information about the agent can be summed up in the following table.

Table 1: Agent information.

Action Space	Discrete(5)
Observation Shape	(number of neighbors,)
Observation High	1.0
Observation Low	0.0

4.3 The Environment Dynamics

The environment is based on the asynchronous and synchronous expressed and private opinion dynamics model. Consider a connected network of n agents. Let $y(t) = [y_1, y_2, \ldots, y_n]^T \in \mathbb{R}^n$, and $\hat{y}(t) = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n]^T \in \mathbb{R}^n$ be a vector representing the private, and expressed opinions (states) of all the agents in the network, respectively. Next, we define $Y(t) = [y(t)^T, \hat{y}(t)^T]^T$, as the vector of all opinions (states) among the individuals at time step t. The vector of initial private opinions (prejudices) is defined as $u = [u_1, u_2, \ldots, u_n]^T \in \mathbb{R}^n$. We define M as a matrix that encodes the weights for calculating social pressure. This matrix can be similar to the influence matrix $M = W$ (which is followed in this work)

or it can be different depending on the source of the social pressure (prevailing opinion in the network).

We use $W = [w_{ij}]$ as the influence matrix of the network which is a stochastic matrix. We define $\tilde{W} = \text{diag}(w_{ii})$ as the diagonal matrix containing the diagonal values of W which represents the self-confidence of each agent in the network, and $\hat{W}$ as the exact matrix as W with 0 in its diagonal ($\hat{w}_{ij} = w_{ij}$ for all $j \neq i$). The influence matrix can be rewritten as $W = \tilde{W} + \hat{W}$.

Additionally, we define $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_n]$ as a vector that encodes the agent's susceptibility to social influence, while $\phi = [\phi_1, \phi_2, \ldots, \phi_n]$ is a vector that encodes the agent's resilience to social pressure. The matrix $\Lambda = \text{diag}(\lambda)$ is a susceptibility matrix that encodes the susceptibility of the agents in its diagonal, and the matrix $\Phi = \text{diag}(\phi)$ is a resilience matrix encoding the resilience of the agents in its diagonal.

We designate $\alpha(t) = [\alpha_1(t), \alpha_2(t), \ldots, \alpha_n(t)]$ as a choice vector of zeros and ones that encode the actions agents have chosen by the agents at time t, where a zero value indicates that a particular agent has chosen not to express their opinion at the time step t, and a value of one signifies the agent's choice to share their opinion with their neighbors. Let $E = \text{diag}(\alpha(t))$ be a zero matrix with the choices of the agents encoded in its diagonal, and $T = I_n - E$ be an identity matrix where the value of the activated agents are set to zero.

To demonstrate the dynamics of the system we define $P_{\alpha(t)}$ and C as follows

$$P_{\alpha(t)} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} \Lambda\tilde{W} & \Lambda\hat{W} \\ E\Phi & T + E(I_n - \Phi)M \end{bmatrix} \in \mathbb{R}^{2n\times 2n}, \quad (5)$$

$$C = \begin{bmatrix} (I_n - \Lambda) \\ \mathbf{0}_{nxn} \end{bmatrix} \in \mathbb{R}^{2n\times n}, \quad (6)$$

where $\mathbf{0}_{nxn}$ is a square zero matrix, and I_n is the identity matrix. The dynamics of the environment are updated using the following equation

$$\begin{bmatrix} y(t+1) \\ \hat{y}(t+1) \end{bmatrix} = P_{\alpha(t)} \begin{bmatrix} y(t) \\ \hat{y}(t) \end{bmatrix} + \begin{bmatrix} (I_n - \Lambda) \\ \mathbf{0}_{nxn} \end{bmatrix} u. \quad (7)$$

This equation can be rewritten as follows

$$Y(t+1) = P_{\alpha(t)} Y(t) + Cu. \quad (8)$$

The matrix $P_{\alpha(t)}$ changes with each time step t based on the changes that occur in the choice vector $\alpha(t)$. Each control agent in the environment has the following available actions.

Table 2: Action space

Action	Opinion value	Meaning
0	0.1	Strong disagreement
1	0.3	Slight disagreement
2	0.5	Neutral
3	0.7	Slight agreement
5	0.9	Strong agreement

The following algorithm is used to update the dynamics of the system at each time step.

Algorithm 2: Step function in system dynamics

```
time step = time step +1
Agents_Activation = random(size=number of agents)
E = diag(Agents_Activation)

for i in range(length(control agents) do
    E[control agents[i]] = 1
    Agents_Current_Exp_Opinions[control agents[i]] =
      action values[actions[i]]

    T = I_n − E
    P11 = Λ W_wave
    P12 = Λ W_hat
    P21 = E Φ
    P22 = T + E(I_n − Φ)M
    c = (I_n − Λ) initial_private_states)

    Ypvt = P11 Agents_Current_Pvt_Opinions + P12
      Agents_Current_Exp_Opinions + c

    Yexp = P21 Agents_Current_Pvt_Opinions + P22
      Agents_Current_Exp_Opinions

    Agents_Current_Pvt_Opinions = Ypvt
    Agents_Current_Exp_Opinions = Yexp
    controlling agent observation = []

    for agent in control agents do
        a = []
        for i in control agent observations[agent] do
            a.append(Agents_Current_Exp_Opinions[i])
        end
    end

    controlling agent observation.append(a)
    check reward(Agents_Current_Exp_Opinions)
    check terminal(Agents_Current_Exp_Opinions)

    return controlling agent observation, reward, terminal
end
```

4.4 Reward Function

Every control agent in the network aims to influence the individuals' opinions to a specific value (opinion). The opinions of the agents in the network range from 0, which represents strong disagreement, to 1, which represents strong agreement. Let ref_i be the opinion goal or reference of the control agent i.

We define the disagreement $d(i, j)$ between two individuals i, j as the squared difference between their opinions at equilibrium: $d(i, j) = w_{ij}(y_i^* - y_j^*)^2$, and the total disagreement is defined as $D_G = \sum_{(i,j)\in\mathcal{V}} d(i, j)$, [39].

The overall goal of the control agents is to influence the network to a predetermined opinion. To encourage the agents to reach their goal within the desired number of steps, we define $O_i(t)$ as the vector containing the expressed opinions of the agents connected to agent i at time step t. The reward function for each control agent is given

by

$$r_i(t) = -\sum_{j \in N_i} (ref_i - O_{ij}(t))^2. \quad (9)$$

The control agent is given an increasingly negative reward for each connected agent that does conform to the goal opinion to encourage it to influence as many individuals as possible.

5 Experimental Results

We tested the effectiveness of the RL agent in influencing and controlling the opinions of individuals in social media networks. A graph was randomly generated using the Erdos-Renyi random graphs model, preferential attachment model, or the stochastic block model. The weights agents assigned to themselves and their neighbors were generated randomly.

The susceptibility and resilience values were randomly generated to ensure the agent learned under different conditions. Then one or more control agents were randomly selected to influence the rest of the network to one specific opinion. At each time step, the agent observed the expressed state of their connected neighbors and took action to influence their opinion. The reward was calculated at each time step to motivate the agent to complete the task in a timely manner. Note that the agents start with no knowledge of the underlying network. In addition, the agents do not have any prior knowledge about other agents in the network.

The agents started with a very high exploration factor ε that decayed gradually until the value reached 0.01, which means that the agents reverted to an almost purely greedy algorithm when choosing their action. The factor ε is decayed by a factor of 0.0001 every episode. The discount factor for the agents was selected as 0.99, which means that the agents developed long-term planning to achieve their goal rather than focus on immediate rewards.

The agent was trained for 500, 000 episodes, and the reward was logged for every 1, 000-th episode. Additionally, for every 1, 000-th episode, we log the episode's length to measure how fast the agent completes the assigned task. Figure 3 shows the evolution of an agent's learning process, starting with a randomly defined knowledge base and no experience over 250, 000 iterations. The rewards (orange line) show the 500 rolling average of the rewards for a more clear illustration.

Figure 3 shows an agent's learning in a random network. The agent starts to receive high rewards after 150, 000 iterations. However, there are still episodes where the control agents struggle to efficiently control the network due to the difference in agent personality (susceptibility and resilience) or opinions.

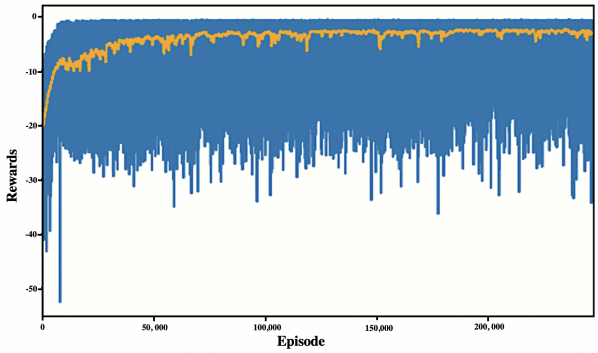


Figure 3: Training reward of a control agent over 250,000 training episodes with data collected every 1,000 episodes.

Figure 4 shows the reward over 250, 000 training episodes with the reward logged every 1, 000-th training episode and a 500 rolling average. The agent starts with a large negative reward when they are exploring ways to influence the network, then the agent starts getting consistent rewards close to 0, which is the optimal reward, after 150, 000 training iterations. The figure shows that the agent is able to influence the system where the reward of the network reaches 0 meaning that the opinions of the agents in the network converge to the opinion desired by the control agent. However, increasing the training time after 250, 000 iterations does not have much effect on the reward received by the agent.

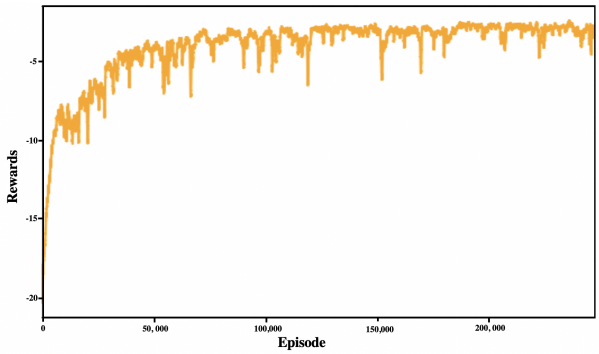


Figure 4: Training reward of a control agent over 250,000 training episodes (with 500 rolling average).

Figure 5 shows the length of every episode and how long the agent takes to reach an efficient policy. It can be seen that with more training, the agent can influence the individual's opinion in a much smaller period of time. It can also be observed that for the first 90 training episodes, the agent cannot influence the network in less than 30 time steps. However, this improves as the agent learns more techniques to gather influence and control the network more efficiently. Any training over 150, 000 episodes did not generate drastically different results.

Figure 6 shows a random group of 4 connected agents interacting in a social media environment with one control agent (Agent 1). The control agent can observe the expressed opinion of the other three individuals in the network. However, the agent can only influence two of these individuals (the size of the nodes indicates the level of influence each individual has). The network shows that all other agents have more influence than the control agent. The

agents in the network randomly expressed their opinions at each time step.

The evolution of opinions of the agents in Figure 6 is observed in Figure 7. The control agent aims to have all others in the network reach an opinion value of 0.9 (strong agreement) on the discussed topic. The control agent changes their opinion to influence the rest of the group. At the end of the discussion, the other agents in the network adopt an opinion similar to the goal of the control agent. Additionally, we can see that occasionally the control agent changes their opinion to match those of his/her neighbors in order to increase their overall reward.

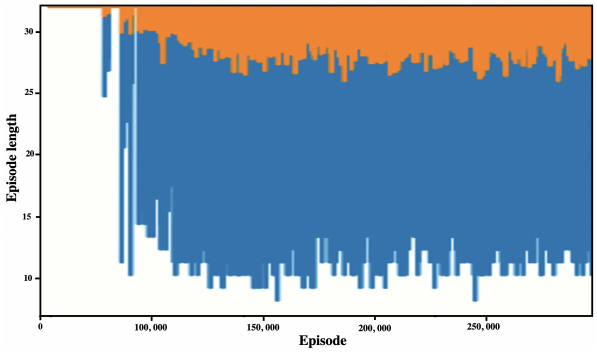


Figure 5: Episode length of a control agent over 250,000 training episodes.

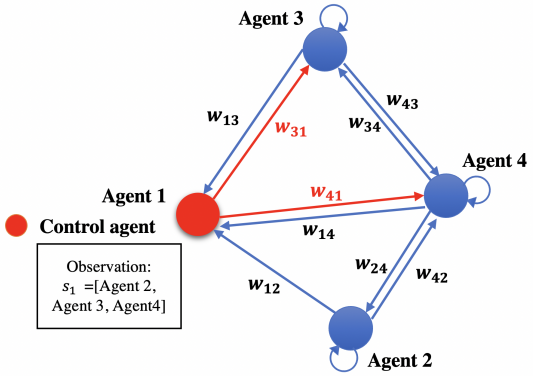


Figure 6: A network of 4 agents with one control agent (red) (size of the agent indicates their connections).

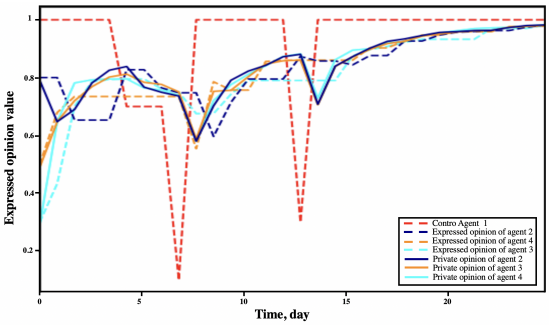


Figure 7: The evolution of private and expressed opinions of a network of 4 agents with one control agent (red).

6 Conclusion and Future Work

In this paper, we proposed an approach based on RL to solve the problem of disagreement between agents in multi-agent systems as well as social network control. To do so we used the expressed and private opinion dynamics model with asynchronous and synchronous updating dynamics to create an environment that closely resembles a social media network. Additionally, we used independent RL control agents to influence and control the network to a desired output. The method was validated using a random social media environment where agents interact and update their opinions randomly at every time step.

For future work, we are planning to investigate the interaction of control agents that are working cooperatively in very large networks. Additionally, we will explore the performance of controlling agents on evolving networks, and on networks with stubborn agents.

References

[1] M. N. Zareer, R. R. Selmic, "Modeling Competing Agents in Social Media Networks," in 2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV), 474–480, 2022, doi:10.1109/ICARCV57592.2022.10004247.

[2] H. Allcott, M. Gentzkow, "Social media and fake news in the 2016 election," Journal of economic perspectives, **31**(2), 211–236, 2017.

[3] A. Fourney, M. Z. Racz, G. Ranade, M. Mobius, E. Horvitz, "Geographic and Temporal Trends in Fake News Consumption During the 2016 US Presidential Election." in CIKM, volume 17, 6–10, 2017, doi:10.1145/3132847.3133147.

[4] A. Guess, B. Nyhan, J. Reifler, "Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign," 2018.

[5] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, L. De Alfaro, "Some like it hoax: Automated fake news detection in social networks," arXiv preprint arXiv:1704.07506, 2017, doi:10.48550/arXiv.1704.07506.

[6] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, et al., "The science of fake news," Science, **359**(6380), 1094–1096, 2018.

[7] M. Howlett, Designing public policies: Principles and instruments, Routledge, 2019.

[8] P. Kotler, H. Kartajaya, I. Setiawan, Marketing 4.0: moving from Traditional to Digital, John Wiley & Sons, 2016.

[9] J. Van Dijck, T. Poell, M. De Waal, The platform society: Public values in a connective world, Oxford University Press, 2018.

[10] L. Floridi, J. Cowls, "A unified framework of five principles for AI in society," Machine learning and the city: Applications in architecture and urban design, 535–545, 2022, doi:10.1002/9781119815075.ch45.

[11] M. Saburov, "Reaching a consensus in multi-agent systems: a time invariant nonlinear rule," Journal of Education and Vocational Research, **4**(5), 130–133, 2013, doi:10.22610/jevr.v4i5.110.

[12] Y. Dong, W. Liu, F. Chiclana, G. Kou, E. Herrera-Viedma, "Are incomplete and self-confident preference relations better in multicriteria decision making? A simulation-based investigation," Information Sciences, **492**, 40–57, 2019, doi:10.1016/j.ins.2019.04.015.

[13] S. E. Parsegov, A. V. Proskurnikov, R. Tempo, N. E. Friedkin, "Novel multidimensional models of opinion dynamics in social networks," IEEE Transactions on Automatic Control, **62**(5), 2270–2285, 2016, doi:10.1109/TAC.2016.2613905.

[14] B. D. Anderson, M. Ye, "Recent advances in the modelling and analysis of opinion dynamics on influence networks," International Journal of Automation and Computing, **16**(2), 129–149, 2019, doi:10.1007/s11633-019-1169-8.

[15] M. H. DeGroot, "Reaching a consensus," Journal of the American Statistical association, **69**(345), 118–121, 1974.

[16] J. Becker, D. Brackbill, D. Centola, "Network dynamics of social influence in the wisdom of crowds," Proceedings of the national academy of sciences, **114**(26), E5070–E5076, 2017, doi:10.1073/pnas.1615978114.

[17] A. G. Chandrasekhar, H. Larreguy, J. P. Xandri, "Testing models of social learning on networks: Evidence from two experiments," Econometrica, **88**(1), 1–32, 2020, doi:10.3982/ECTA14407.

[18] E. Johnsen, "Social influence and opinions," J. Math. Sociology, **15**(3-4), 193–205, 1990.

[19] N. E. Friedkin, E. C. Johnsen, Social influence network theory: A sociological examination of small group dynamics, volume 33, Cambridge University Press, 2011.

[20] H. Kreft, W. Jetz, "Global patterns and determinants of vascular plant diversity," Proceedings of the National Academy of Sciences, **104**(14), 5925–5930, 2007.

[21] C. C. Childress, N. E. Friedkin, "Cultural reception and production: The social construction of meaning in book clubs," American Sociological Review, **77**(1), 45–68, 2012, doi:10.1177/0003122411428153.

[22] M. Ye, Y. Qin, A. Govaert, B. D. Anderson, M. Cao, "An influence network model to study discrepancies in expressed and private opinions," Automatica, **107**, 371–381, 2019, doi:10.1016/j.automatica.2019.05.059.

[23] S. E. Asch, "Effects of group pressure upon the modification and distortion of judgments," Groups, leadership, and men, 177–190, 1951.

[24] D. A. Prentice, D. T. Miller, "Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm." Journal of personality and social psychology, **64**(2), 243, 1993.

[25] M. N. Zareer, R. R. Selmic, "Expressed and private opinions model with asynchronous and synchronous updating," in 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2840–2846, IEEE, 2021, doi:10.1109/SMC52423.2021.9659269.

[26] S. Moscovici, M. Zavalloni, "The group as a polarizer of attitudes." Journal of personality and social psychology, **12**(2), 125, 1969.

[27] R. Felix, P. A. Rauschnabel, C. Hinsch, "Elements of strategic social media marketing: A holistic framework," Journal of business research, **70**, 118–126, 2017, doi:10.1016/j.jbusres.2016.05.001.

[28] S. Issenberg, "Cruz-connected data miner aims to get inside US voters' heads," Bloomberg, https://www. bloomberg. com/news/features/2015-11-12/is-the-republicanparty-s-killer-data-app-for-real, 2015.

[29] J. Shao, J. Qin, A. N. Bishop, T.-Z. Huang, W. X. Zheng, "A novel analysis on the efficiency of hierarchy among leader-following systems," Automatica, **73**, 215–222, 2016, doi:10.1016/j.automatica.2016.07.007.

[30] F. Dietrich, S. Martin, M. Jungers, "Control via leadership of opinion dynamics with state and time-dependent interactions," IEEE Transactions on Automatic Control, **63**(4), 1200–1207, 2017, doi:10.1109/TAC.2017.2742139.

[31] J. Shao, W. X. Zheng, T.-Z. Huang, A. N. Bishop, "On leader–follower consensus with switching topologies: An analysis inspired by pigeon hierarchies," IEEE Transactions on Automatic Control, **63**(10), 3588–3593, 2018, doi:10.1109/TAC.2018.2797205.

[32] A. Clark, B. Alomair, L. Bushnell, R. Poovendran, "Leader selection in multi-agent systems for smooth convergence via fast mixing," in 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), 818–824, IEEE, 2012, doi:10.1109/CDC.2012.6426323.

[33] W. Su, X. Chen, Y. Yu, G. Chen, "Noise-Based Control of Opinion Dynamics," IEEE Transactions on Automatic Control, **67**(6), 3134–3140, 2022, doi:10.1109/TAC.2021.3095455.

[34] G. Albi, L. Pareschi, M. Zanella, "On the optimal control of opinion dynamics on evolving networks," in System Modeling and Optimization: 27th IFIP TC 7 Conference, CSMO 2015, Sophia Antipolis, France, June 29-July 3, 2015, Revised Selected Papers 27, 58–67, Springer, 2016, doi:10.1007/978-3-319-55795-3_4.

[35] B. Ross, L. Pilz, B. Cabrera, F. Brachten, G. Neubaum, S. Stieglitz, "Are social bots a real threat? An agent-based model of the spiral of silence to analyse the impact of manipulative actors in social networks," European Journal of Information Systems, **28**(4), 394–412, 2019, doi:10.1080/0960085X.2018.1560920.

[36] D. Sohn, N. Geidner, "Collective dynamics of the spiral of silence: The role of ego-network size," International Journal of Public Opinion Research, **28**(1), 25–45, 2016, doi:10.1093/ijpor/edv005.

[37] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction, MIT press, 2018.

[38] M. Langlois, R. H. Sloan, "Reinforcement learning via approximation of the Q-function," Journal of Experimental & Theoretical Artificial Intelligence, **22**(3), 219–235, 2010, doi:10.1080/09528130903157377.

[39] C. Musco, C. Musco, C. E. Tsourakakis, "Minimizing polarization and disagreement in social networks," in Proceedings of the 2018 world wide web conference, 369–378, 2018, doi:10.1145/3178876.3186103.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 9, No. 1, 67-78 (2024)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Enhancing the Network Anomaly Detection using CNN-Bidirectional LSTM Hybrid Model and Sampling Strategies for Imbalanced Network Traffic Data

Toya Acharya*, Annamalai Annamalai, Mohamed F Chouikha

Electrical and Computer Engineering, Prairie View A & M University, Prairie View, Texas,77446, USA

ARTICLE INFO

Article history:
Received: 15 November, 2023
Revised: 17 January, 2024
Accepted: 17 January, 2024
Online: 20 January, 2024

Keywords:
Random Under Sampling
Random Over Sampling
Network Anomaly Detection
UNSW-NB15
NSL-KDD
Data imbalance
Convolutional Neural Network
CNN-BLSTM
Data Imbalance

ABSTRACT

The cybercriminal utilized the skills and freely available tools to breach the networks of internet-connected devices by exploiting confidentiality, integrity, and availability. Network anomaly detection is crucial for ensuring the security of information resources. Detecting abnormal network behavior poses challenges because of the extensive data, imbalanced attack class nature, and the abundance of features in the dataset. Conventional machine learning approaches need more efficiency in addressing these issues. Deep learning has demonstrated greater effectiveness in identifying network anomalies. Specifically, a recurrent neural network model is created to recognize the serial data patterns for prediction. We optimized the hybrid model, the convolutional neural network combined with Bidirectional Long-Short Term Memory (BLSTM), to examine optimizers (Adam, Nadam, Adamax, RMSprop, SGD, Adagrad, Ftrl), number of epochs, size of the batch, learning rate, and the Neural Network (NN) architecture. Examining these hyperparameters yielded the highest accuracy in anomaly detection, reaching 98.27% for the binary class NSL-KDD and 99.87% for the binary class UNSW-NB15. Furthermore, recognizing the inherent class imbalance in network-based anomaly detection datasets, we explore the sampling techniques to address this issue and improve the model's overall performance. The data imbalance problem for the multiclass network anomaly detection dataset is addressed by using the sampling technique during the data preprocessing, where the random over-sampling methods combined with the CNN-based BLSTM model outperformed by producing the highest performance metrics, i.e., detection accuracy for multiclass NSL-KDD and multiclass UNSW-NB15 of 99.83% and 99.99% respectively. Evaluation of performance, considering accuracy and F1-score, indicated that the proposed CNN BLSTM hybrid network-based anomaly detection outperformed other existing methods for network traffic anomaly detection. Hence, this research contributes valuable insights into selecting hyperparameters of deep learning techniques for anomaly detection in imbalanced network datasets, providing practical guidance on choosing appropriate hyperparameters and sampling strategies to enhance model robustness in real-world scenarios.

1. Introduction

This study is extension version of the presented conference paper titled "Efficacy of CNN-Bidirectional LSTM Hybrid Model for Network-Based Anomaly Detection" [1] at the 2023 IEEE 13th Symposium on Computer Applications & Industrial Electronics (ISCAIE).

As technology undergoes rapid advancements, the transmission of information has transformed significantly, adopting various methods such as wired, wireless, or guided networks. This evolution in network technology is pivotal to people's daily activities. Whether it's communicating with others, accessing online resources, or sharing information, the efficiency and security of these interactions depend heavily on the underlying network infrastructure. A system attains security when it effectively maintains the three essential notions of computer information security: Availability, Integrity, and Confidentiality (CIA). In essence, information security involves safeguarding information from unauthorized entities and protecting against

*Corresponding Author: Toya Acharya, tacharya@pvamu.edu

illegal access, use, disclosure, reformation, recording, or destruction of data. Confidentiality guarantees that the information is accessible only to individuals or systems. In information network technology, encryption methods and access controls prevent unauthorized users from gaining access to sensitive data during transmission. Information integrity guarantees that data remains unaltered during transmission. In the context of information network technology, this involves implementing mechanisms to detect and prevent unauthorized modifications to data, ensuring that the information received is the same as what was sent. Availability ensures that information and resources are available and accessible when needed. In an information resources and security environment, availability involves designing robust and reliable systems that can withstand potential disruptions, whether they are due to technical failures or malicious attacks. The overarching goal of information security is to safeguard information from unauthorized access and malicious activities. This includes preventing unauthorized individuals or systems from gaining access to sensitive data, ensuring that information remains unchanged and reliable during transmission, and guaranteeing that information and resources are available when needed.

Measures to achieve information security encompass a range of strategies, including encryption to protect data confidentiality, checksums or digital signatures to ensure data integrity, and redundancy and fault-tolerant systems to enhance availability. Additionally, access controls, firewalls, and intrusion detection are commonly utilized to fortify the security posture of networked systems, mitigating the risks associated with information resources.

A traditional network cannot be fully protected by relying solely on a firewall and antivirus software. These security measures identify predefined anomalous activities and establish the rule to prevent those unusual events by the cyber expert. In anomaly detection, outliers and anomalies are occasionally employed interchangeably. This approach finds extensive use across diverse domains, such as commercial, network attack detection, health systems monitoring, credit card fraud transaction detection, and identifying faults in mission-critical infrastructure systems. Anomaly detection is crucial in cybersecurity, providing robust protection against cyber adversaries. Ensuring safeguard network resources is essential to safeguard the organization from cyber threats.

Anomalies are categorized into point, contextual, and collective types based on the results generated by the detection method [2]. Point anomalies occur when a specific activity diverges from the typical rules or patterns. Contextual anomalies involve unusual patterns within a particular circumstance that consistently differ from numerous normal activities. Collective anomalies occur when a group of related instances exhibit anomalous behavior compared to the normal activity dataset.

Intrusion detection techniques can be broadly classed into two main types: Signature-based Intrusion Detection System (SIDS) and Anomaly-based Intrusion Detection System (AIDS). Anomaly detections, in contrast, are classified according to their origins, resulting in network-based and host-based intrusion/anomaly detection systems. Detecting anomalies in data is facilitated by employing labels to differentiate between normal and abnormal occurrences. There are three fundamental approaches to detecting anomalies: supervised, semi-unsupervised, and unsupervised methods. In the supervised approach, the system is trained on labeled data, distinguishing between normal and anomalous instances. On the other hand, unsupervised methods detect anomalies without prior labeling, relying on deviations from established patterns. Semi-supervised techniques combine elements of both, using labeled and unlabeled data for training. AIDS overcomes the drawbacks of SIDS by utilizing ML, statistical-based, or knowledge-based methods to model normal behaviors. However, it's worth noting that anomaly-based detection may produce false results due to alterations in user habits.

Numerous traditional machine learning algorithms favor shallow learning methodologies, giving significant importance to feature engineering designed for smaller data. The feature engineering phase needs more processing time and domain expertise to create pertinent features and eliminate unrelated ones from anomaly detection algorithms. The effectiveness of anomaly detection is intricately tied to feature engineering and data preprocessing implementation. Traditional machine learning methods, characterized by simplicity, low resource consumption, and subpar performance in areas like vision, language processing, and image translations, underscore the limitations of these approaches.

CNN is predominantly employed for image signals, leveraging its architecture to effectively capture and analyze visual information. Individual neurons play a key role in reducing the dimensionality of the network's features in the lower layers of a CNN. These neurons are adept at identifying essential small-scale features within the images, including boundaries, corners, and variations in intensity. The CNN network links lower-level features to produce more complicated features in the upper layers, encompassing fundamental shapes, structures, and partial objects. The ultimate layer of the network amalgamates these lower features to generate the output results.

The functioning of a long short-term memory differs from that of a CNN due to its specific design to safeguard long-range info within a sequential order. Unlike CNNs, LSTMs are crafted to remember and store information over extended sequences, avoiding the loss of crucial details. In the case of BLSTM, an additional LSTM layer is incorporated, introducing a reversal in the information flow direction. This architectural enhancement addresses challenges related to vanishing gradients, ensuring more effective training by considering information from both forward and backward directions in the sequence.

Data imbalance, including network anomaly detection, is a common challenge in ML applications. In network traffic anomaly detection, data imbalance refers to the unequal distribution of normal and anomalous instances in the dataset for training the detection model. Anomalies in network traffic are typically rare incidents compared to normal activities, leading to imbalanced data.

The deep learning approach addresses issues found in conventional machine learning. The effectiveness of the deep learning-based anomaly detection algorithm relies on factors such as the NN architecture, #hidden layers, activation functions, batch

size, and the number of epochs utilized during DL model testing, training, and validation. The careful selection of these factors, including hyperparameters and the architecture of NN in deep learning, is crucial for enhancing the detection accuracy of network traffic anomaly detection. The essential selection of ML or DL models overcomes the class imbalance problem. The ensemble method, which combines more different individual models, requires longer training time and consumes more resources. The sampling method generates random data or deletes the random data based on the implemented sampling methods to create the balanced form of the final dataset, which is efficient in dealing with the imbalanced dataset.

2. Literature Review

The rapid increase of information and technology has led to widespread connectivity of numerous end terminals to the internet and networks. Those smart terminals contribute to generating substantial volumes of data, commonly referred to as big data. This huge flood of data is a valuable resource for analysis and insights. Machine learning and deep learning algorithms come into play to extract meaningful information from this vast data pool. The daily growth of big data presents difficulties for conventional machine learning algorithms, necessitating thorough feature extraction and discovery. DL substantially increases anomaly detection and model performance. Nevertheless, the dataset's attributes and features, hyperparameters in deep neural networks, and the structure of neural networks are pivotal elements that impact the efficacy of identifying anomalies in network-based IDS.

Conventional machine learning strongly relies on intricate and time-consuming feature engineering, often impractical for real-time applications. In the [3] study, the authors proposed an approach for payload classification utilizing CNN and RNN to detect attacks, achieving detection accuracies of 99.36% and 99.98% on the DARPA98 network data, respectively. CNN methods discern specific grouping patterns through convolution around input neighborhoods, while RNN works on sequences by calculating correlations between previous and current states. In another[4] study, class imbalance was handled utilizing a CNN with a Gated Recurrent Unit (GRU) hybrid model. To address the data class imbalance and feature redundancy, they used a hybrid sampling technique that integrates Pearson Correlation Analysis (PCA), repeated edited nearest neighbors, Random Forest (RF), and adaptive synthetic sampling. With the detection accuracies of 99.69%, 86.25%, 99.69%, and 99.65% on the NSL-KDD, UNSW_NB15, and CIC-IDS2017 datasets, respectively, their CNN-GRU model performed better.

The research authors [5] proposed using an Adaptive Synthetic Sampling (ADASYN) technique in a DL-based network intrusion detection system to overcome dataset imbalance. On the NSL-KDD network data, they used an autoencoder to reduce dimensionality. The CNN-BLSTM hybrid DL method obtained the greatest F1 score (89.65%) and accuracy (90.73%). To address problems resulting from data in class imbalance and heterogeneous data distribution across various information sources, the research [6] used convolutional neural networks with federal transfer learning. The UNSW-NB15 multiclass network dataset produced an average detection accuracy was 86.85% for the model.

In [7], the researcher addressed data class imbalance on network datasets: NSL-KDD, KDD99, and UNSW-NB15 datasets using heterogeneous ensemble-assisted ML methods for binary and multi-class network intrusion detection. Using the NSL-KDD dataset, the model showed a 96.2% AUC and a true positive rate (TPR) of 94.5%.

The authors of [8]discovered that ML classifier performance increased with the decrease in target classes. Conventional ML approaches, such as Naïve Bayes, Random Forest, J48, Bagging, Adaboost, and BayesianNet, were used to investigate this idea on three network traffic-based intrusion datasets: KDD99, UNSW-NB15, and CIC-IDS2017_Thursday.

In a study [9], the authors suggested a method for achieving network intrusion classification with low computing cost, creating a group of target classes based on the nature of network traffic. They created cluster characteristics for each group using K-means on the KDD99 network dataset, resulting in a detection accuracy of 98.84%. However, the intrusion detection model accuracy for user2root (U2R) is notably low at 21.92%, impacting overall performance. In [10], authors employed a hybrid method, combining CNN and LSTM, to enhance model classification accuracy, achieving 96.7% and 98.1% on CIC-IDS2017 and NSL-KDD network data, respectively.

In the study[11], CNN and LSTM combined to create a hybrid model was proposed to enhance network intrusion detection model facilities for advanced metering infrastructure through cross-layer features combination. This method achieved the highest network intrusion detection accuracy of 99.79% on NSL-KDD and 99.95% on KDD Cup99 but with limited user2root (U2R) detection capabilities. Similarly, in [12], authors employed a hybrid method of combining CNN and LSTM to improve model network intrusion detection by capturing additional network traffic data's spatial and temporal features.

In[13], the researchers implemented a hybrid technique based on the mean control of the CNN and BLSTM to address issues of conventional data pre-processing and imbalanced numerical distribution of class instances in the NSL-KDD, achieving the optimal detection accuracy of 99.10%. However, the accuracy for the minority traffic data class remains suboptimal. Using a different methodology, the authors [14] created a DL model that combined CNN and BLSTM to learn temporal and spatial characteristics. Accuracy levels on the binary class UNSW-NB15 were 93.84% and binary NSL-KDD of 99.30%.

Data was preprocessed using one-hot encoding and min-max normalization by authors in [15], which achieved an accuracy of 96.3% on CNN and Bi-LSTM hybrid methods on the multiclass NSL-KDD dataset. Using preprocessed on given NSL-KDD data, researchers in[16] applied the hybrid model using CNN and BLSTM algorithm with a 95.4% accuracy rate. A bidirectional LSTM model was used by the authors in their study [17] for the binary NSL-KDD dataset with the highest accuracy of 98.52%. Using a Bidirectional LSTM deep learning model, authors[18] got 99% accuracy on UNSW-NB15 and KDDCUP-99, which is an exceptional achievement. But a lot of the models that are now in use need help effectively identifying uncommon (rare) attack types, especially user2root (U2R) and remote2local (R2L) attacks,

which frequently have poorer detection accuracy as compared with other network attack types.

To overcome the difficulties found in the above literature review, authors in [19] presented a Bi-LSTM-based network intrusion detection system on the NSL-KDD dataset, which offered a binary classification accuracy of 94.26%. Furthermore, the authors proposed a Bi-directional GAN-based method [20] for the NSL-KDD and CIC-DDoS2019 datasets. The bidirectional GAN model demonstrated strong performance with an f1 score and detection accuracy of 92.68% and 91.12%, respectively, on the unbalanced NSL-KDD dataset.

In the research study [1], [17] the Authors used the hyperparameters tunning to obtain the best model performance on network intrusion detection datasets, including NSL-KDD and UNSW-NB15. In [21], the Authors implement the BLSTM model combined with random over-sampling strategies, which produces a high anomaly detection accuracy of 99.83% for multiclass imbalance network anomaly datasets NSL-KDD dataset.

The deep learning model discussed in [3] and [4] overcomes challenges traditional machine learning encounters in anomaly detection. While the CNN standalone model is unsuitable for sequential data preprocessing, and RNN requires complex data preprocessing, this model effectively addresses these issues. Data imbalance problems are tackled in [5]-[8]. Feature engineering emerges as a critical factor in enhancing the accuracy of both ML and DL models. Much research has been conducted on feature engineering, with studies focusing on attribute grouping found in [9]-[12]. The BLSTM, which brings together two distinct LSTMs to allow input processing in both directions (from the past to the future and vice versa), is implemented in [13]-[20] to improve the accuracy of network anomaly detection models.

Most of the researchers mentioned above concentrate on enhancing the detection accuracy of conventional or ML DL models and employ ensemble methods for feature engineering to address data imbalance. However, there needs to be more emphasis on exploring hyperparameter selection in DL-based models, determining the train-test split ratio, and defining the architecture of DNN. Some researchers need to elaborate on adopting these values in their studies. Subsequently, this research addressed these limitations in network traffic anomaly detection systems. We experimented using binary and multiclass versions of the UNSW-NB15 and NSL-KDD. Our focus includes investigating the performance comparison between random under-sampling and over-sampling to identify superior methods for imbalanced network data.

The contributions of our research effort in the area of network anomaly detection and imbalanced datasets are listed as:

a. Examining the impact of CNN and BLSTM neural network architecture and performance for binary/multi-class datasets, specifically NSL-KDD and UNSW-NB15.

b. Exploring the model performance of hyperparameters on binary and multi-class network datasets, namely UNSW-NB15 and NSL-KDD.

c. Exploring the enhancement of CNN Bi-LSTM by varying memory elements and numbers of layers of NN.

d. This study's interest is developing and implementing a CNN Bi-LSTM hybrid model for network anomaly detection, achieving high accuracy rates of 98.27% on NSL-KDD binary data and 99.87% on UNSW-NB15 binary data.

e. Exploring the network anomaly detection model based on CNN Bi-LSTM using UNSW-NB15.

f. Investigating the random sampling methods for imbalanced data with detection accuracy greater than 99.83% for NSL-KDD multiclass data and 99.99% for the UNSW-NB15 multiclass dataset.

The rest of the paper unfolds: Section 3 delineates the system model and individual blocks comprising our CNN Bi-LSTM hybrid approach. Section 4 elucidates the experimental setup, experimental results, and discussion of the findings, and section 5 encapsulates the conclusion of this research.

3. Network Anomaly Detection Model Description

The complete proposed model comprises the following steps:

1. Network traffic-based data collection
2. Data pre-processing and cleaning
3. Training and testing data preparation
4. CNN BLSTM model preparation
5. Train and test model
6. Evaluation of CNN BLSTM model
7. Compare the model and decision-making

The CNN BLSTM-based model's entire implementation schematic is displayed in Figure 1. The ensuing sections offer a thorough explanation of the approaches mentioned previously. The components of CNN and BLSTM layers and the intricate architecture of neural networks are seen in Figure 2.

3.1. Network traffic-based data collection

Numerous datasets are accessible for research in network intrusion detection systems. Examples include the KDD Cup99, Kyoto 2006+, NSL-KDD, CICIDS2017, UNSW-NB15, and several others, providing valuable resources for intrusion detection research. During this research, the UNSW-NB15 and NSL-KDD datasets are specifically employed.

NSL-KDD KDDTrain+ [22] originates from the DARPA KDD99 dataset, with the elimination of noise and undesired data. This dataset encompasses the complete NSL-KDD training set, including labels denoting attack types and difficulty levels. Comprising 41 features, it delineates five different attack classes: "Normal," "DoS," "Probe," "R2L", and "U2R".

NSL-KDD represents a refined form of the KDD99 data, free from duplicate records in the training set and the test sets. Each entry in the dataset consists of 42 attributes, with 41 of them related to the input traffic and the final label indicating whether the traffic

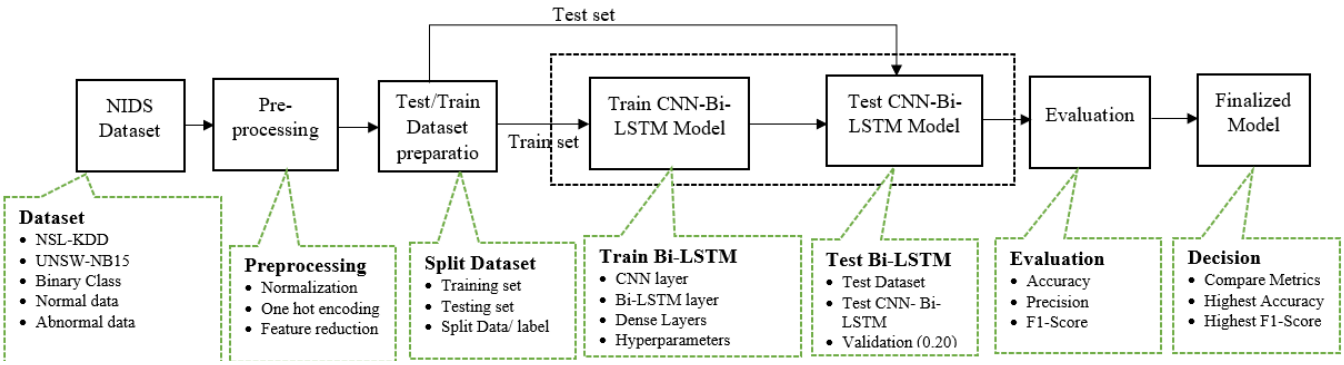


Figure 1: Pipeline for CNN BLSTM hybrid model

is normal or abnormal (target). The KDDTrain+ dataset encompasses 125,973 data entries, while the KDDTest+ dataset consists of 22,544 data entries utilized in this research work. Table 1 documents the detailed information regarding the traffic and data information [23].

Table 1: Details of NSL-KDD data

Traffic	KDDTrain+	KDDTest+
R2L	995	2,885
U2R	52	67
DoS	45,927	7,460
Normal	67,343	9,711
Probe	11,656	2,421
Total	125,973	22,544

Similarly, The Australian Centre for Cyber Security (ACCS) cybersecurity research team constructed the UNSW-NB15 dataset [24], unlike KDD99 and NSL-KDD, which is a recently developed network intrusion dataset created by IXIA PerfectStrom tools within the Cyber Range Lab of the ACCS, this dataset consists of approximately 100GB of PCAP files capturing raw network traffic flows between two hosts either server to client or vice versa. The Argus and Bro_IDS tools and 12 other algorithms generated 49 features accompanied by class labels. Numerous records were utilized to construct the training and testing sets, where UNSW_NB15_training-set and UNSW_NB15_testing-set were used during this research work. The training set comprises 175,341 records, while the testing set comprises 82,332 records, encompassing various attacks and normal network activity. Table 2 shows detailed information regarding the attacks and normal traffic.

Table 2: Details of UNSW-NB15 data

Network Traffic	testing-set.csv	training-set.csv
Exploits	11,132	33,393
Generic	18,871	40,000
Worms	44	130
Fuzzers	6,062	18,184
DoS	4,089	12,264
Reconnaissance	3,496	10,491
Analysis	677	2,000
Backdoor	583	1,746
Shellcode	378	1,133
Normal	37,000	56,000
Total	82,332	175,341

The KDDTrain+ and KDDTest+ subsets of the NSL-KDD dataset were employed in our research experiment—likewise, experiments involved using training-set.csv and testing-set.csv from the UNSW-NB15 dataset.

3.2. Data pre-processing and cleaning

NSL_KDD data is an improved version of the KDD99 dataset; minimum work is required for data preprocessing. The downloaded separate data files are used to test and train the model. The target class is initially isolated from the training and testing datasets to create the class labels. From the remaining attributes, numerical features and three categorical features—"protocol_type", "service", and "flag" are extracted. The categorical features undergo conversion into numerical values using dummy one-hot encoding techniques, while the numerical attributes are standardized using standard Scalar methods. Afterwards, both types of feature sets are combined into a unified data frame, yielding the final data sets for training and testing. One hot encoding generates one binary variable for each individual categorial value. The dummy encoding is similar to one hot encoding and converts the categorical values into numeric binary values. The dummy encoding represents N categories using N-1 binary variable. Let's say we have three categories of traffic "protocol_type," "service," and "flag" that are going to be dummy encoded as [1 0], [0 1], and [0 0], respectively. The standard scalar converts the numeric values so that the data standard deviations become 1.

Since there are different types of services present in the KDDTrain+ dataset and KDDTest+ dataset, the one hot encoding produces unequal numbers of features. The KDDTrain+ dataset contains 126 features, while the KDDTest+ includes a total of 120 features after the implementation of one hot encoding. Those additional features "service_aol," "service_harvest," "service_http_2784", "service_http_8001", "service_red_i," and "service_urh_i" are inserted into the KDDTest+ dataset after finding the exact location where those features reside into the KDDTrain+ dataset. We preserved the attacks_types and difficulty_level features because those features are highly relevant to the target class and increase the model's efficiency.

The UNSW-NB15 dataset was divided into two sets for training and testing purposes: UNSW_NB15_training-set and UNSW_NB15_testing-set. The UNSW_NB15_training-set comprises 175,341 entries, while the UNSW-NB15_testing-set contains 82,332 entries, encompassing various attacks and normal data. Initially, the features on this dataset are 49. First, those categorical attributes are changed into numeric using dummy one hot encoding. All numerical attributes are applied to the standard scalar normalization method. After preprocessing the numeric and categorical features, 192 features for UNSW_NB15_testing-set

data and 196 features for UNSW_NB15_training-set data were generated. Again, here we are taking two sets of data: one we can use for training and the other for testing or vice versa. The categorical values of data entries are not the same for both datasets; hence, the one hot encoding produces unequal numbers of features on both data sets after preprocessing.

Some features generated from one hot encoding, such as state_ACC and state_CLO, are not included in the UNSW-NB15_training-set. Similarly, proto_icmp, proto_rtp, state_ECO, state_PAR, state_URN, and state_no features are not included on UNSW_NB15_testing-set. The empty features columns are added in the exact column location of those missing features on the respective dataset, generating 198 features plus one target class.

3.3. Training and testing data preparation

In experiments concerning the binary NSL-KDD dataset, the training and testing datasets were created using a split ratio. The train-test split approach assesses the performance of machine learning algorithms in making predictions from data that wasn't part of the training set. We opted for a 70:30 split ratio to generate the train and test dataset. For the CNN BLSTM hybrid model, 70% of KDDTrain+ was used to train, and the remaining data was used to test the model for binary NSL-KDD data.

A similar split percentage was employed in the binary class UNSW-NB15, using the "UNSW_NB15_training-set". In the case of multiclass experiments for UNSW-NB15 and NSL-KDD, two distinct files were selected—one subset for training the CNN-BLSTM model and another for testing. Detailed information regarding this split is provided in the respective experimental sections.

3.4. CNN BLSTM model

CNN is a forward DNN designed for image signal and classification. CNN comprises three primary layers: the convolutional, the pooling, and the fully connected layers. The convolutional layer is the main component of CNN and uses the convolutional operation to grab the various features from the image signal. Then, the number of pooling layers extracts features, and a fully connected layer employs the output from the preceding layer for classification. Combining convolutional layers with pooling layers is responsible for feature extraction, while the final fully connected dense layer is utilized for classification purposes. CNN also involves various hyperparameters, including the number of filters, stride, zero-padding, pooling layers, and others.

An RNN is an artificial NN designed to manage sequential data by integrating feedback loops into its structure. Diverging from conventional feedforward neural networks that linearly handle input data, RNNs feature connections forming loops, enabling them to retain a memory of past inputs and utilize that information to impact the current output. The memory in an RNN serves as a short-term storage, allowing the network to retain information about past events and use it to make predictions about future events. This is especially valuable in applications where context and temporal relationships are essential. Machine learning issues, including speech recognition, language processing, and picture categorization, have been resolved with RNN.

Yet, traditional RNNs encounter challenges, notably needing help with learning long-term dependencies attributed to the vanishing or exploding gradient problem. Advanced RNN versions such as gated recurrent units (GRUs) and long short-term memory (LSTM) networks have been devised in response to these constraints. These architectures include mechanisms for selectively storing and retrieving information across extended sequences, enhancing their effectiveness in tasks that demand capturing long-term dependencies.

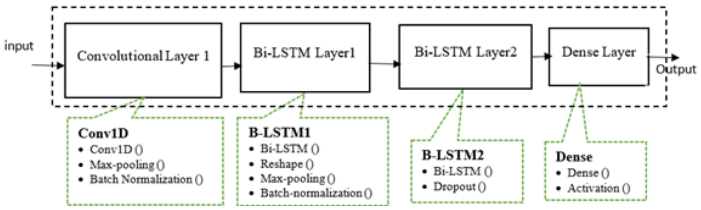


Figure 2: CNN BLSTM layer architecture

LSTM handles the vanishing gradient in RNN. There is a memory block and three multiplicative units in LSTM. The input corresponds to the write operation, output to read and forget gates corresponding to the reset operations for cells that make up the LSTM architecture. By allowing LSTM memory cells to keep and access data for longer periods. Those multiplicative gates mitigate the vanishing gradient.

To process input in both directions—from the future to the past and from the past to the future—bidirectional RNN combines two independent RNNs. Both forward and backward LSTM networks make up the Bi-LSTM. The features extracted by the forward LSTM hidden layer point forward, whereas those extracted by the reverse LSTM hidden layer point backward. By taking finite sequences into account about earlier and later items, the bidirectional LSTM can anticipate or tag the sequence of each element. Two LSTMs processed in series—one from left to right and the other from right to left—produce this. The CNN and BLSTM hybrid models have several layers, each with a set of hyperparameters. Figure 2 shows the CNN BLSTM's architectural layout.

3.5. CNN BLSTM model training

The CNN BLSTM model's neural network architecture is prepared for training. The datasets consist of two sets: one for training and the other for testing, or vice versa. The split percentage determines how much data is allocated for training and testing when a single data set is present. The selection of hyperparameters for model training is conducted through various experiments involving fine-tuning epochs and batch size to enhance detection efficiency. Within the training data, 20% is designated for validating the CNN Bi-LSTM model.

3.6. Test the CNN BLSTM hybrid model and evaluation.

Deep learning (DL) and machine learning (ML) models offer performance consistency. After the CNN BLSTM model is built, the model is trained using the training dataset with specified hyperparameter values. These chosen hyperparameter values influence the training duration. Following training, the model can assess the unseen dataset to evaluate its performance.

Hyperparameter selection lacks a predefined rule, allowing for random selection and subsequent fine-tuning through various experiments. After the model testing, performance metrics are determined based on the type of ML model employed. In the case of the supervised machine learning model, ground truth values are utilized to measure the performance metrics on the test dataset. Various metrics, such as detection accuracy, precision, F1-Score, recall, program execution time, and Area under the ROC, are available to compare the model efficiency. Confusion metrics from Karas generate True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values. In the context of a classification report, the terms "weighted" and "macro" refer to different strategies for computing metrics such as precision, recall, and F1-score across multiple classes. Macro-averaging computes the metric for every class separately before averaging them. This means that each class is treated equally in the computation, regardless of size. Macro-averaging gives the same weight to each class, which can be useful when all classes are considered equally. Weighted averaging, on the other hand, takes the average of the metrics, but it weights each class's contribution based on its proportion in the dataset. In other words, classes with more samples have a greater impact on the average. W weighted averaging is especially helpful when working with unbalanced datasets—where certain classes may have substantially more instances than others.

The classification report provides a thorough summary of the model's performance metrics for the specified training and testing data sets. Lastly, to assess the performance of our CNN BLSTM hybrid model, the performance metrics are compared with the findings of earlier research publications.

3.7. Compare models and decision-making.

Several sets of experiments were conducted to find the hyperparameter settings that yielded the best results. Following model testing and evaluation, choosing the best model pipeline from various options is part of the cognitive process of comparison and decision-making. Throughout this study, several sets of experiments are carried out to determine values for various CNN Bi-LSTM model hyperparameters to enhance the model's performance. To create an effective Bi-LSTM pipeline, it is necessary to decide on the hyperparameters, which include optimizers, number of epochs, batch, NN design, class size, and techniques of raw data preprocessing. This is achieved by evaluating performance metrics across multiple sets of experiments. The performance metrics of the Bi-LSTM model are then juxtaposed with previously published results for the binary/numerous class UNSW-NB15 and NSL-KDD. The class imbalance problem in the multiclass version of both NSL-KDD and UNSW-NB15 datasets was exposed with sampling data during the preprocessing stages. The sampling methods randomly deleted on down-sampling and randomly generated data samples in over-sampling. This resulted in the balanced form of datasets to compare the CNN Bi-LSTM model performance.

4. Results and Discussion

To detect anomalies, intrusion detection uses a mix of DL and ML methods. The implementation of a network anomaly detection model is implied using Python script. Python has specialized packages for building machine learning models, including NumPy, Pandas, Keras, and Scikit-learn. Additionally, commonly used tools like Java, C#, WEKA, Visual C++, and MATLAB play vital roles in network anomaly detection systems. On the Jupyter Notebook platform, seed values are fixed to guarantee consistency in outcomes over several runs. Plots and tables representing the results of experiments are analyzed using the Microsoft Office suite. Every experiment is run on a Windows machine with an i7 processor and 16GB of RAM.

Python and the packages it is linked with keep version information used in all experiments. For example, TensorFlow 2.9.1, Keras 2.6.0, and Python 3.7.12 are used. Hyperparameters will be determined, performance will be evaluated across class sizes, and the efficacy of various sampling approaches will be assessed about the CNN BLSTM model for the multi-class and binary-class UNSW-NB15 and NSL-KDD. Detailed explanations of these experiments are provided in subsequent sections.

The architecture shown in Figure 2 consists of a single 16-unit convolution layer that uses batch normalization and max-pooling. BLSTM neural network layer 1 contains 50 memory units; batch normalization, max-pooling, and reshaping come next. Bi-LSTM neural network layer 2 with 100 memory units and dropout is also available. The dense layer consists of a sigmoid activation, and the final output is obtained. The detection accuracy of the model is evaluated through a series of tests involving the adjustment of optimizers, learning rate (LR), number of epochs, batch size, and dropout rate. As explained below, the UNSW-NB15 and NSL-KDD binary/multiclass network traffic datasets are used for these investigations.

4.1. Experiment: Model performance Vs. Optimizers

In the context of ML and DL, an optimizer is an algorithm or method used to adjust the parameters of a model to minimize or maximize a certain objective function. The performance of an optimizer is crucial in training machine learning models because it determines how well the model learns from the data. Choosing the optimizer is essential during the training of the CNN BLSTM model, as it significantly contributes to expediting results for the machine learning/deep learning model.

TensorFlow offers nine optimizers (Ftrl, Nadam, Adam, Adadelta, Adagrad, gradient descent, Adamax, RMSprop, and Stochastic Gradient Descent (SGD)) based on the optimizer's methods. The choice of optimizer can significantly impact the training performance of an ML model. Optimizers may converge at different rates or achieve different final accuracies on a given task. An optimizer's performance may be influenced by the model's architecture, the dataset, and the hyperparameters employed.

It is common practice to experiment with several optimizers to determine which combination of optimizers and hyperparameters is optimum for a given task. Additionally, some optimizers may perform better on certain types of neural network architectures or for specific data types. In summary, the relationship between the optimizer and machine learning performance is crucial, and choosing the right optimizer is an important part of the model training process. It often involves experimentation and tuning to find the optimal combination for a given task.

The model used in the experiment comparing Optimizers versus Accuracy has a 20% dropout rate and the Relu activation

function. To determine the best optimizer for our CNN-BLSTM model, seven optimizers, including Nadam, Ftrl, SGD, Adam, RMSprop, Adagrad, and Adamax, were tested. Based on the model performance metrics for UNSW-NB15 and NSL-KDD binary data, which are shown in Table 3, it was found that the Nadam optimizer performed best for NSL-KDD. In contrast, the Adam optimizer produced the best accuracy for the UNSW-NB15 dataset. Interestingly, although both optimizers used the same model architecture, they performed differently for both Network Intrusion Detection System (NIDS) datasets.

Table 3: Model performance Vs. Optimizer

Number of epochs = 10, Batch = 256, NSL-KDD_C2 and UNSW-NB15_C2				
Optimizer	**ACC-NSL**	**F1-NSL**	**ACC-UN**	**F1-UN**
Ftrl	53.47	69.68	80.99	80.99
RMSprop	97.87	98.01	97.93	98.46
Adamax	97.65	97.78	95.33	96.51
Adam	98.02	98.16	**99.15**	**99.38**
Adagrad	96.98	97.21	94.04	95.62
SGD	97.74	97.91	99.14	99.37
Nadam	**98.13**	**98.26**	99.11	99.34
ACC: Accuracy in %, F1: F1Score in %, NSL: KDD-NSL, UN: UNSW_NB				

Table 4: CNN BLSTM performance Vs. Optimizer on NSL-KDD

Model Performance Vs. Optimizer on NSL-KDD Multiclass Datasets			
Epochs=10, Batch_size= 512, Training_data = KDDTrain+, Testing_data = KDDTest+, Multiclass=5			
Optimizers	**Accuracy %**	**wt_Precision %**	**wt_F1score %**
Adam	**88.46**	**88.87**	**88.23**
RMSprop	85.49	87.15	82.84
Nadam	84.79	86.97	82.45
SGD	82.86	84.99	77.12
Adamax	82.6	86.99	82.72
Adagrad	75.65	67.01	69.94
Ftrl	43.08	18.56	25.94

Table 5: CNN BLSTM performance Vs. optimizer on UNSW-NB15

CNN Bi-LSTM Performance Vs. Optimizer on UNSW-NB15 Multiclass Datasets			
Epochs=15, Batch_size= 512, Training data=UNSW-NB15Train82332, Testing_data = UNSW-NB15Test175341, Multiclass=10			
Optimizers	**Accuracy %**	**wt_Precision %**	**wt_F1score %**
SGD	**89.84**	**87.49**	**88.01**
Adam	87.21	87.47	85.96
Nadam	84.59	84.48	83.38
RMSprop	79.3	75.71	76.85
Adamax	76.84	76.37	74.74
Adagrad	70.82	63.28	62.05
Ftrl	31.94	10.20	15.46

The selection of the optimizers depends on the combination of the different hyperparameters and NN architecture of the CNN BLSTM model. Popular optimization algorithm Adam combines concepts from RMSprop and momentum. It adapts the learning rates of individual parameters and is widely used in deep learning. An Adam extension that uses the Nesterov Accelerated Gradient (NAG). NAG involves looking ahead in the direction of the momentum before computing the gradient that combines the benefits of Adam and Nesterov momentum. Figure 3. shows the accuracy comparison for NSL-KDD and UNSW15.

Tables 4 and 5 shows the comparative performance metrics of the multi-class NSL-KDD and UNSW-NB15. The same optimizer does not provide the same performance for a similar dataset. The hyperparameters and datasets used to test and train the CNN-based BLSTM model are provided in Tables 4. and 5. For the NSL-KDD multiclass dataset, Adam performed better than SGD, whereas for the UNSW-NB15 multiclass dataset, SGD performed better than other optimizers.

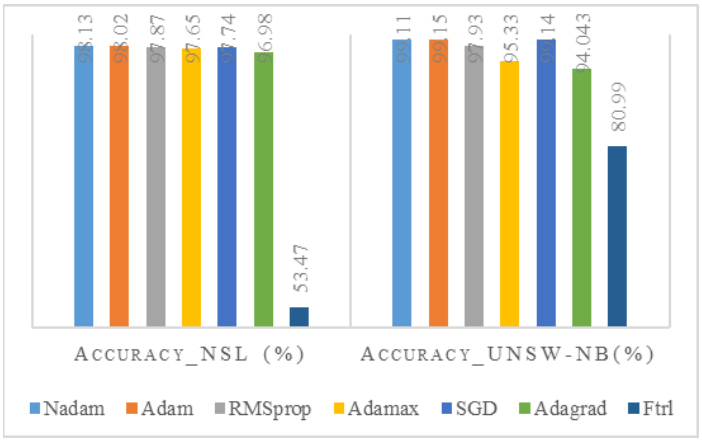


Figure 3: Optimizer Vs. Accuracy

4.2. Experiment: Learning rate Vs. model performance

The learning rate, a positive scalar multiplied by gradient descent gradient, controls the step size in parameter space. A higher rate facilitates faster convergence but raises the risk of overshooting and oscillation. On the other hand, a lower rate ensures stability but may demand more iterations for convergence.

With optimizers chosen from the preceding Experiment 4.1, the same CNN BLSTM model neural network architecture is used to determine the ideal learning rate to enhance the model performance. The NSL-KDD binary data is preprocessed from the subset of the KDDTrain+ dataset, and the split ratio splits the data for training and testing. The learning rate determines the rate at which new weights are added to neural network models. The other hyperparameters remain constant throughout this experiment while the learning rates are adjusted to optimize the model's accuracy. Table 6 displays a comparison of learning rate with CNN BLSTM model performance. The model performs best on the UNSW-NB15 binary data and the NSL-KDD binary dataset, achieving a learning rate of 0.01 and 0.0002, respectively. The same learning rate provides different model performances.

Table 6: CNN BLSTM model Learning rate Vs. Performance metrics

Epochs size = 10, Batch = 256, KDD_C2 (Nadam), UNSW-NB15_C2 (adam)				
LR	**ACC-NSL**	**F1-NSL**	**ACC-UN**	**F1-UN**
0.01	97.49	97.67	**99.67**	**99.76**
0.001	98.16	98.29	99.54	99.66
0.0001	98.06	98.20	95.81	96.85
0.0002	**98.18**	**98.3**	97.9	98.44
0.0003	98.14	98.27	98.44	98.86
0.0004	97.97	98.11	99.13	99.35
0.0005	98.11	98.25	99.09	99.32
LR: Learning rate, ACC: Accuracy in %, F1: F1Score in %, UN: UNSW_NB				

4.3. Experiment: Model dropout rate Vs. model performance

The phrase "dropout rate" in machine learning usually refers to a regularization method that neural networks employ to avoid overfitting. When a model becomes overfit, it can have poor generalization on new, unknown data because it has learned the training set too well, including its noise and outliers. During training, randomly selected neurons (units) in the neural network

are "dropped out" or omitted temporarily. This means these neurons do not contribute to the forward or backward pass during a particular iteration of training. The probability of a neuron being dropped out is called the dropout rate. The dropout rate is one hyperparameter that must be determined before training the model.

The CNN BLSTM model was tested and trained for both datasets using a batch size of 256 and 10 epochs. Different dropout rate values were used to evaluate the efficiency of the model. The model performed better than the others, with a 30% dropout rate on the UNSW-NB15 dataset; however, a 60% dropout rate worked better for the NSL-KDD. The hyperparameter values, dropout rates, and corresponding performance metrics are presented in Table 7. The experimental results highlight the varying dropout rates for distinct datasets despite the similarity between the two datasets.

Table 7: Dropout rate Vs. model performance

Epochs size = 10, batch = 256, KDD_C2 (madam), UNSW-NB15_C2 (adam)				
DropOut %	**ACC-NSL**	**F1-NSL**	**ACC-UN**	**F1-UN**
0.1	98.10	98.24	97.44	98.15
0.2	98.02	98.16	98.98	99.25
0.3	98.16	98.29	**99.87**	**99.9**
0.4	98.04	98.17	99.27	99.47
0.5	97.93	98.09	99.47	99.61
0.6	**98.21**	**98.33**	99.81	99.86
0.7	98.01	98.15	99.58	99.69
0.8	98.04	98.18	98.57	98.94
ACC: Accuracy in %, F1: F1Score in %, NSL: KDD-NSL, UN: UNSW_NB KDDTrain+, UNSW-NB15 training.csv binary with test-train split				

The batch size is a hyperparameter in machine learning that determines how many samples are used in a training iteration. The batch size represents the number of samples used in a single training iteration. Using a smaller batch size incorporates a limited number of data samples and results in a longer training time for the CNN Bi-LSTM model compared to a larger batch size. Throughout experimentation (Experiment A-C), the batch size is altered while maintaining other hyperparameters, such as a fixed number of epochs is 5, the learning rate of the optimizer, and the dropout rate values assigned to the model based on previous findings with the respective datasets.

Table 8: Model performance Vs. batch size

Number of Epochs = 5, KDD_C2 (Nadam), UNSW-NB15_C2(adam)				
Batch	**ACC-NSL %**	**F1-NSL %**	**ACC-UN %**	**F1-UN %**
32	97.89	98.04	**99.40**	**99.55**
64	97.95	98.10	99.35	99.52
128	**98.06**	**98.20**	99.33	99.50
256	97.64	97.79	96.36	97.26
512	97.92	98.08	96.90	97.70

The dataset size, the amount of computing power available, and the specifics of the optimization issue can all influence the batch size decision. Experimenting with various batch sizes is a frequent way to determine which is most effective for a certain task.

The experimental result in Table 8 demonstrates how the neural network's hyperparameter combinations affect performance. In this experiment, batch sizes of 128 for the binary NSL-KDD datasets and 32 for the binary UNSW-NB15 datasets for epochs 5 demonstrated the best performance of the CNN BLSTM model.

4.4. Experiment: Epochs Vs. model performance

An "epoch" in machine learning is one whole iteration through the training dataset a model goes through while training. The learning method processes the complete dataset throughout each epoch, modifying the neural network weights and parameters to reduce the error or loss function. A hyperparameter called epoch count determines an algorithm's running frequency over the full training dataset. The integer between one to infinity can be used as the epoch. Selecting smaller epoch values results in a longer training time for the model and vice versa. Underfitting, the ML model cannot identify the original patterns in the data, which can be caused by using too few epochs. However, an excessive number of epochs might cause overfitting, in which case the model becomes inattentive to new data and underperforms on previously unknown data.

The CNN BLSTM hybrid model performance for binary KDD-NSL and binary UNSW-NB15 with the different values of epochs are documented in Table 9. The performance increases with large values of epochs but is different for a while. After 75 epochs, the model performance decreases. The amount of data utilized for training and testing, the size of the output class, and other hyperparameter combinations affect the epochs and performance of the machine learning/deep learning models.

Table 9: Epochs Vs. model performance

Batch size = 256, NSL- KDD_C2 (Nadam)		
Number of Epochs	**Accuracy-NSL %**	**F1Score-NSL %**
2	95.48	95.94
10	98.13	98.26
25	98.21	98.33
50	98.20	98.33
75	**98.27**	**98.39**
100	98.26	98.39

The selection of epoch size to produce a superior performance on an imbalanced dataset is challenging. The binary dataset is more balanced than the multiclass network-based intrusion dataset. The experimental results in Table 9 are not the determining experiment for the number of epochs on multiclass NSL-KDD and UNSW-NB15 datasets. Hence, we experimented with and documented multi-class experimental results to determine the values of epochs where we can produce higher accuracy on the provided dataset. Tables 10 and 11 show the experimental results for multiclass datasets to investigate the values of epochs to make superior detection accuracy. In summary, while epoch size and class size are conceptually different, they can influence each other indirectly, especially when dealing with imbalanced datasets. Selecting the right number of epochs for a given problem is crucial, as is keeping an eye on how class sizes affect model performance.

Table 10: Model performance Vs. Epochs on UNSW-NB15 multiclass data

Batch=512, Optimizer=SGD, Training=UNSW-NB15Train.csv82332 testing data= UNSW-NB15Test.csv175341, Multiclass=10				
Epochs	**ACC**	**wt_Prec**	**wt_F1Score**	**Prg_exe_time**
10	**93.10**	**91.10**	**91.94**	**0.64**
25	83.09	79.94	79.69	1.17
50	86.4	82.47	83.35	2.24
75	87.1	86.46	85.24	3.12
100	90.04	88.36	88.01	4.24
150	82.23	81.49	80.82	6.36

200	81.13	78.65	78.84	7.95
ACC: Accuracy in %, wt_Prec: weighted Precision in %, wt_F1Score: weighted F1Score in %, Prg_exe_time: Program script run time in hr.				

Program execution time is the sum of the model's training and testing phases. The program execution time depends on various factors, such as the neural network's architecture, training and testing data size, class size, and combination of hyperparameters. We documented the performance of the CNN BLSTM hybrid model along with the program execution time. The higher the epochs result, the longer the program execution time. Table 10 and Table 11 show the multi-class NSL-KDD run for almost 8 Hrs. to complete training testing and evaluate the model for 200 epochs. A similar scenario for multiclass UNSW-NB15 dataset. Hence, selecting epoch and batch size is the trade-off with the model training, testing, and evaluation time. We found in Table 10 and Table 11 the different epoch sizes for NSL-KDD (outperform at epoch size 10) and UNSW-NB15 (outperform at epoch 100) during multi-class model performance.

Table 11: Performance Vs. epochs on NSL-KDD multiclass data

Batch=512, Optimizer= Adam, Data=KDDTrain+_125973, KDDTest+ 22544, Multiclass = 5				
Epochs	**ACC**	**wt_Prec**	**wt_F1Score**	**Prg_exe_time**
10	86.21	88.13	83.85	0.45
25	86.64	88.01	84.1	1.08
50	86.64	88.7	84.78	2.15
75	87.11	88.39	84.84	3.31
100	**87.63**	**89.85**	**86.44**	**4.53**
150	87.22	89.66	85.81	7.03
200	86.84	90.61	86.41	8.76
ACC: Accuracy in %, wt_Prec: weighted Precision in %, wt_F1Score: weighted F1Score in %, Prg_exe_time: Program script run time in hr.				

4.5. Experiment: Imbalance data sampling Vs. performance

This experiment investigates the sampling techniques for imbalanced data to provide a high detection rate. The researcher employed various techniques, such as a critical selection of ML and DL algorithms, ensemble methods, data sampling, etc., to address the issue of data imbalance because there are fewer attacks than typical traffic data in the provided network intrusion detection dataset.

Sampling methods generate or delete random data from the dataset based on class data distribution. Random under-sampling and random over-sampling are two techniques used in imbalanced classification problems, where one class (usually the minority traffic class) is significantly under-represented compared to the other class(es). These methods are utilized to tackle class imbalance and enhance the efficacy of machine learning models. Random Under Sampling (RUS) involves randomly removing instances from the majority class until the distribution between the majority and minority classes is more evenly distributed. However, random over-sampling produces an equal distribution by randomly duplicating minority class instances or creating synthetic instances to increase the number of minority class instances.

Tables 12 and 13 provide the hyperparameter information and performance of this experiment's CNN BLSTM hybrid model. Table 12 compares the NSL-KDD multiclass dataset's performance when random over- and under-sampling is applied. After preprocessing multi-class NSL-KDD data, the training and testing datasets merge into a single file. Sampling is implemented on merged data, and a 70:30 split ratio is used to split data into train and test datasets.

Table 12: CNN BLSTM performance Vs. Sampling on NSL-KDD

CNN BILSTM, Epochs= 25, Batch_size= 512, Data = combine (KDDTrain+KDDTest+) (sampling)				
Class	**Recall_RUS**	**F1_RUS**	**Recall_ROS**	**F1_ROS**
DoS	0	0	99.86	99.92
Probe	100	36.69	99.89	99.91
U2R	10	18.18	100	99.73
R2L	0	0	99.45	99.63
Normal	0	0	99.93	99.93
Wt_Average →	22.15	10.07	99.83	99.83
Macro_Avg→	22	10.97	99.83	99.83
Accuracy %	22.15		**99.83**	
F1:F1Score, RUS: Random Under Sampling, ROS: Random Over Sampling] %				

Similarly, preprocessed training data and testing data files are merged into a single file to implement the sampling method on the UNSW-NB15 multiclass dataset. The sampled data is then split into training and testing datasets using a 70:30 train-test split ratio During Random under-sampling, data instances are randomly deleted from the majority class, resulting in significant information loss. Deleting samples from the majority class results in a smaller sample, unsuitable for the deep learning model and worsens the model performance, which is found in the experiment and documented in Tables 12 and Table 13. Random over-sampling (ROS) helps prevent information loss, as none of the minority class instances are removed. It can be more effective when the amount of data in the minority class is limited.

Table 13: Model performance Vs. sampling on UNSW-NB15

CNN_BLSTM Epochs=25, Batch_size=512, data = combine (UNSW-NB15training-set 175341+UNSW-NB15testing-set 82332) sampling				
Class	Recall_RUS	F1_RUS	Recall_ROS	F1_ROS
Analysis	0	0	100	100
Backdoor	0	0	100	100
DoS	0	0	100	99.95
Exploits	100	17.51	99.91	99.95
Fuzzers	0	0	99.98	99.99
Generic	0	0	99.98	99.98
Normal	0	0	100	100
Reconnaissance	0	0	100	100
Shellcode	0	0	100	100
Worms	0	0	100	100
Weighted_Average	9.2	1.61	99.99	99.99
Macro Average	10	1.75	99.99	99.99
Accuracy (%)	9.20		**99.99**	
F1:F1Score, RUS: Random Under Sampling, ROS: Random Over Sampling] %				

This method generates random data based on the data distribution in the dataset. The huge amount of data is always suitable for deep learning models. Regarding detection accuracy, our suggested CNN BLSTM hybrid model performs better than the random over-sampling technique, offering over 99%. Tables 12, 13, and Figure 4 above detailed the CNN BLSTM hybrid model's performance for the UNSW-NB15 imbalance dataset and the multiclass NSL-KDD.

5. Conclusion

The previous research from the literature reviews shows that while the detection accuracy for rarely occurring attack classes (U2R, R2L) is low, the average model accuracy for normal traffic in the UNSW-NB15 and NSL-KDD is roughly 99%. Regardless

of the type of attack, each poses a threat to network machines equally. To provide a comparative analysis, we juxtapose our results with existing findings of 91.12% [20] and 90.83% [5] detection accuracy for NSL-KDD binary, and 99.70% [18], 82.08% [14], 82.08% for UNSW-NB15 binary datasets. Our experiments enhance accuracy to 98.27% on NSL-KDD and 99.87% on UNSW-NB15 binary datasets by carefully selecting hyperparameters and conducting various experiments. We explored the CNN BLSTM hybrid model's hyperparameters (dropout. epochs, batch size, learning rate, and optimizer) to maximize detection accuracy for the binary NSL-KDD and UNSW-NB15.

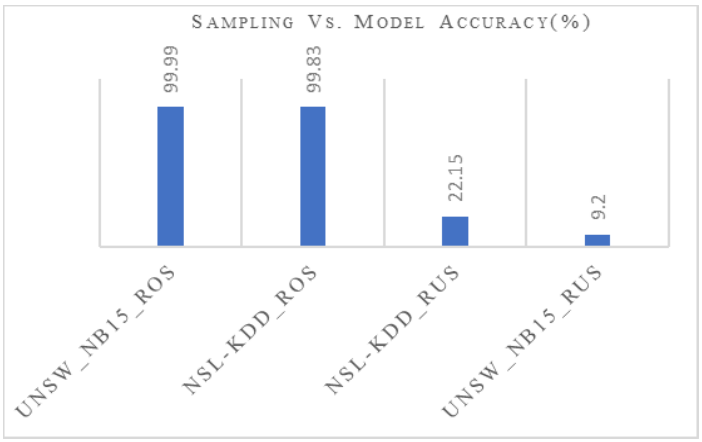


Figure 4: Sampling Vs. multiclass model accuracy (%)

The model performance depends on the combination of hyperparameters, the size of the dataset used to train/test the model, and the selection of the machine learning/ deep learning model. Our research provides information about the data size used during the experiments and the choice of hyperparameters. The suggested model uses random over-sampling techniques on a single set of data to provide 99.99% and 99.83% model accuracy for the multiclass UNSW-NB15 and NSL-KDD datasets, respectively (train and test data merge into a single file before sampling).

Selecting random over-sampling or under-sampling relies on the particulars of the dataset and the issue at hand. To achieve a balance, combining the two methods, a practice known as hybrid sampling may occasionally be necessary. It's crucial to remember that there are more sophisticated methods for dealing with class imbalance, such as SMOTE (Synthetic Minority Over-sampling Technique), which creates synthetic instances for the minority class instead of merely copying real instances. Thoroughly examining those approaches in various network intrusion detection multiclass datasets extends this research effort. The proper use of hyperparameters of neural networks, size of dataset used to train the model, and sampling methods for CNN BLSTM network anomaly model provide the highest detection accuracy for imbalance network data.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The National Science Foundation (NSF) and Scholarship for Service CyberCorps (SFS CyberCorps) programs support this research study. The award information of NSF and CyberCorps are #1910868 and #2219611, respectively. This cannot be completed without the continuous support of advisors and the Electrical and Computer Engineering Departments of Prairie View A&M University

References

[1] T. Acharya, A. Annamalai, M.F. Chouikha, "Efficacy of CNN-Bidirectional LSTM Hybrid Model for Network-Based Anomaly Detection," in 13th IEEE Symposium on Computer Applications and Industrial Electronics, ISCAIE 2023, Institute of Electrical and Electronics Engineers Inc.: 348–353, 2023, doi:10.1109/ISCAIE57739.2023.10165088.

[2] N. Moustafa, J. Hu, J. Slay, "A holistic review of Network Anomaly Detection Systems: A comprehensive survey," Journal of Network and Computer Applications, **128**, 33–55, 2019, doi:10.1016/j.jnca.2018.12.006.

[3] H. Liu, B. Lang, M. Liu, H. Yan, "CNN and RNN based payload classification methods for attack detection," Knowledge-Based Systems, **163**, 332–341, 2019, doi:10.1016/j.knosys.2018.08.036.

[4] B. Cao, C. Li, Y. Song, Y. Qin, C. Chen, "Network Intrusion Detection Model Based on CNN and GRU," Applied Sciences (Switzerland), **12**(9), 2022, doi:10.3390/app12094184.

[5] Y. Fu, Y. Du, Z. Cao, Q. Li, W. Xiang, "A Deep Learning Model for Network Intrusion Detection with Imbalanced Data," Electronics (Switzerland), **11**(6), 2022, doi:10.3390/electronics11060898.

[6] X. Ji, H. Zhang, X. Ma, "A Novel Method of Intrusion Detection Based on Federated Transfer Learning and Convolutional Neural Network," in IEEE Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Institute of Electrical and Electronics Engineers Inc.: 338–343, 2022, doi:10.1109/ITAIC54216.2022.9836871.

[7] T. Acharya, I. Khatri, A. Annamalai, M.F. Chouikha, "Efficacy of Heterogeneous Ensemble Assisted Machine Learning Model for Binary and Multi-Class Network Intrusion Detection," in 2021 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc.: 408–413, 2021, doi:10.1109/I2CACIS52118.2021.9495864.

[8] T. Acharya, I. Khatri, A. Annamalai, M.F. Chouikha, "Efficacy of Machine Learning-Based Classifiers for Binary and Multi-Class Network Intrusion Detection," in 2021 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc.: 402–407, 2021, doi:10.1109/I2CACIS52118.2021.9495877.

[9] M. Xiong, H. Ma, Z. Fang, D. Wang, Q. Wang, X. Wang, "Bi-LSTM: Finding Network Anomaly Based on Feature Grouping Clustering," in ACM International Conference Proceeding Series, Association for Computing Machinery: 88–94, 2020, doi:10.1145/3426826.3426843.

[10] S.N. Pakanzad, H. Monkaresi, "Providing a hybrid approach for detecting malicious traffic on the computer networks using convolutional neural networks," in 2020 28th Iranian Conference on Electrical Engineering, ICEE 2020, Institute of Electrical and Electronics Engineers Inc., 2020, doi:10.1109/ICEE50131.2020.9260686.

[11] R. Yao, N. Wang, Z. Liu, P. Chen, X. Sheng, "Intrusion detection system in the advanced metering infrastructure: A cross-layer feature-fusion CNN-LSTM-based approach," Sensors (Switzerland), **21**(2), 1–17, 2021, doi:10.3390/s21020626.

[12] P. Sun, P. Liu, Q. Li, C. Liu, X. Lu, R. Hao, J. Chen, "DL-IDS: Extracting features using CNN-LSTM hybrid network for intrusion detection system," Security and Communication Networks, **2020**, 2020, doi:10.1155/2020/8890306.

[13] L. Zhang, J. Huang, Y. Zhang, G. Zhang, "Intrusion Detection Model of CNN-BiLSTM Algorithm Based on Mean Control," in Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS, IEEE Computer Society: 22–27, 2020, doi:10.1109/ICSESS49938.2020.9237656.

[14] J. Sinha, M. Manollas, "Efficient Deep CNN-BiLSTM Model for Network Intrusion Detection," in ACM International Conference Proceeding Series,

Association for Computing Machinery: 223–231, 2020, doi:10.1145/3430199.3430224.
[15] A. Li, S. Yi, "Intelligent Intrusion Detection Method of Industrial Internet of Things Based on CNN-BiLSTM," Security and Communication Networks, **2022**, 2022, doi:10.1155/2022/5448647.
[16] J. Gao, "Network Intrusion Detection Method Combining CNN and BiLSTM in Cloud Computing Environment," Computational Intelligence and Neuroscience, **2022**, 2022, doi:10.1155/2022/7272479.
[17] T. Acharya, A. Annamalai, M.F. Chouikha, "Efficacy of Bidirectional LSTM Model for Network-Based Anomaly Detection," in 13th IEEE Symposium on Computer Applications and Industrial Electronics, ISCAIE 2023, Institute of Electrical and Electronics Engineers Inc.: 336–341, 2023, doi:10.1109/ISCAIE57739.2023.10165336.
[18] P. TS, P. Shrinivasacharya, "Evaluating neural networks using Bi-Directional LSTM for network IDS (intrusion detection systems) in cyber security," Global Transitions Proceedings, **2**(2), 448–454, 2021, doi:10.1016/j.gltp.2021.08.017.
[19] Y. Imrana, Y. Xiang, L. Ali, Z. Abdul-Rauf, "A bidirectional LSTM deep learning approach for intrusion detection," Expert Systems with Applications, **185**, 2021, doi:10.1016/j.eswa.2021.115524.
[20] W. Xu, J. Jang-Jaccard, T. Liu, F. Sabrina, J. Kwak, "Improved Bidirectional GAN-Based Approach for Network Intrusion Detection Using One-Class Classifier," Computers, **11**(6), 2022, doi:10.3390/computers11060085.
[21] T. Acharya, A. Annamalai, M.F. Chouikha, "Optimizing the Performance of Network Anomaly Detection Using Bidirectional Long Short-Term Memory (Bi-LSTM) and Over-sampling for Imbalance Network Traffic Data," Advances in Science, Technology and Engineering Systems Journal, **8**(6), 144–154, 2023, doi:10.25046/aj080614.
[22] M. and B.E. and L.W. and G.A.A. Tavallaee, "A detailed analysis of the KDD CUP 99 data set," in 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, IEEE, 2009, doi:{10.1109/CISDA.2009.5356528}.
[23] L. Dhanabal, S.P. Shantharajah, "A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms," International Journal of Advanced Research in Computer and Communication Engineering, **4**, 2015, doi:10.17148/IJARCCE.2015.4696.
[24] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, ACT, Australia**,** **2015**, 1-6, doi: 10.1109/MilCIS.2015.7348942**.**

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No.5, 70-76 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Resonance Coil Design for a Novel Battery Cell Balancing with using Near-Field Coupling

Juhyeon Jeon, Dongho Lee*

Department of Electrical and Control Engineering, Mokpo National University, Naju, 58277, Republic of Korea

ARTICLE INFO

Article history:
Received:26 June, 2023
Accepted:18 October, 2023
Online: 30 October, 2023

Keywords:
Battery cell balancing
Near-field coupling
Wireless power transfer
Coil design

ABSTRACT

In this paper, we delve into the pressing necessity for proficient battery cell balancing, an imperative in the context of the escalating adoption of renewable energy and electric vehicles. While traditional methodologies, including the passive technique, offer a straightforward and cost-effective solution, they compromise on efficiency. The active technique, though superior in efficiency, is hindered by its innate restriction of transferring energy solely to proximate cells, thus prolonging the balancing process. To address these limitations, we introduce a novel near-field coupling method centered on a meticulously designed resonant coil with an emphasis on achieving a larger Q-factor, a pivotal factor for enhanced battery cell balancing. This augmented Q-factor not only propels our approach past the passive method in efficiency but also catalyzes rapid balancing by enabling wireless energy transfer to cells regardless of their relative positioning. Validating our theoretical insights, we developed physical coil prototypes and adopted a Series-Parallel circuit configuration, steered by the resonant coil's Q-factor. Preliminary experiments with three batteries substantiate our claim, showcasing that our proposed technique achieves cell balancing with approximately double the speed of conventional strategies.

1. Introduction

This paper is an extension of a work originally presented in ITC-CSCC 2022 [1]. Through the Paris Climate Change Agreement, major countries around the world aim to achieve Net-zero, setting their carbon emissions to '0' by 2050. Among the various policies being pursued by South Korea to achieve these goals, the expansion of renewable energy and electric vehicles plays a major role [2-3]. Renewable energy serves as an alternative to fossil fuels and nuclear power, with solar and wind power at its core. However, such energy generation is intermittent depending on weather conditions, necessitating the use of an Energy Storage System (ESS). The ESS is composed of large-capacity battery packs[4], and each of these battery cells demonstrates slight discrepancies in the State of Charge (SOC) during manufacturing. This deviation tends to increase during charging and discharging processes, leading to problems in battery life and efficiency[5-6]. Battery cell balancing is a technique used to rectify these issues. On the other hand, as the proportion of electric vehicles continues to increase globally, a new environmental issue, namely the 'problem of battery disposal,' has been highlighted [7]. One solution to this problem that is actively researched involves recycling spent batteries to create ESS [8-9]. In this context, the battery cell balancing technology is recognized as playing a crucial role in maximizing the efficiency and lifespan of ESS.

However, the conventional battery cell balancing, as discussed in Section 2, has limitations. As the proportion of renewable energy and electric vehicles continues to expand, there is a growing need for faster and more efficient battery cell balancing techniques. Therefore, in this paper, we propose a novel battery cell balancing method and evaluate its performance.

2. Conventional Battery Cell Balancing

Traditional cell balancing can be divided into passive battery cell balancing and active battery cell balancing. The passive type connects resistors to the battery to expend energy for balancing, while the active type connects capacitors, inductors, or transformers to the battery to transfer energy to adjacent batteries for balancing [10-13].

2.1. Passive Battery Cell Balancing

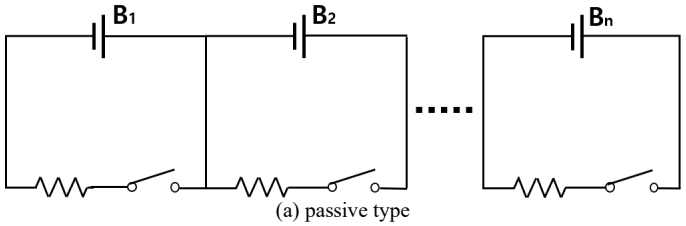
The passive type, which exhausts energy from the batteries with resistors, is based on the battery with the lowest State of Charge (SOC). This method is widely used due to its simple circuit and low cost. However, it has several drawbacks: it wastes energy leading to low efficiency, it generates heat making it prone to fire,

*Corresponding Author: Dongho Lee, dongho.lee864@gmail.com

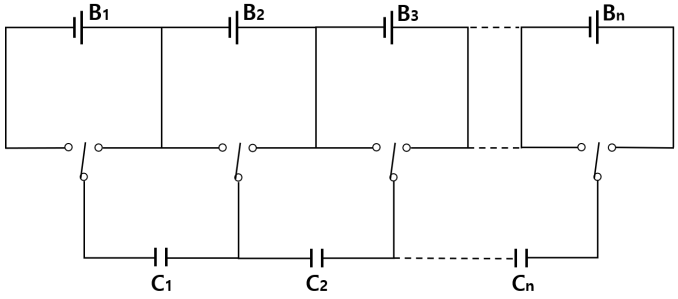
and it can't perform balancing during discharge operation. Therefore, it is not suitable for high-efficiency battery cell balancing. Figure 1 (a) illustrates a switching shunting resistor method of the passive type, where all switches are controlled identically or individually depending on the SOCs of the batteries.

2.2. Active Battery Cell Balancing

The active type transfers energy from cells with a relatively high SOC to those with a lower SOC. While it's more efficient than the passive type, its circuit is complex, control is difficult, and it only allows balancing between adjacent cells, which slows down the balancing speed. Hence, it is not suitable for high-speed battery cell balancing. Figure 1 (b) shows a switching capacitor method of the active type. This method is a simpler and cheaper option within the active type, but if the cells with the highest and lowest SOCs are assumed to be at the ends of the circuit, more switches need to be controlled, leading to a longer time needed for balancing. Other methods to overcome this issue have complicated circuit configurations or high costs.



(a) passive type



(b) active type

Figure 1: Conventional battery cell balancing

2.3. Proposal Battery Cell Balancing

As such, conventional battery cell balancing methods are not suitable for high-speed, high-efficiency applications due to their respective disadvantages. Table 1 summarizes the characteristics, advantages, and disadvantages of typical conventional battery cell balancing methods. It shows that passive type is not suitable for high efficiency, and active type is not suitable for high speed. Therefore, a novel battery cell balancing method, which is more efficient than the passive type and faster than the active type, is needed to overcome the disadvantages of conventional methods. This paper proposes a novel battery cell balancing method that uses near-field coupling for balancing.

Table 1: Summary of typical conventional battery cell balancing

category	passive type		active type		
method	shunt resistor	switching resistor	switching capacitor	inductor	Proposal
devices used	n resistors	n switches and resistors	n-1 capacitors and 2n switches	PWM	N+1 coils and N Rectifiers
advantage	simple circuit and low cost		simple circuit and low cost	reasonable balancing speed	Fast balancing & Reasonable efficiency
disadvantage	low efficiency and safety issue		slow balancing speed	low efficiency & high cost	high cost

3. Near-Field Coupling Battery Cell Balancing

The new battery cell balancing method proposed in this paper utilizes 'Near-Field Coupling'. This method enables selective cell balancing by transmitting energy to specific cells. 'Near-Field Coupling' is a method of transmitting energy between objects in close proximity using a magnetic resonance method. The resonance frequency selected for this method is 13.56MHz in the ISM band (Industrial Scientific Medical Band). Figure 2 illustrates the circuit of the proposed method. When the State of Charge (SOC) of battery B1 is lower than that of B2 and B3, the transmitter's capacitor is variably adjusted to achieve impedance matching with the receiver of B1's battery. The energy transmission continues until the SOC of B2 and B3 matches that of B1.

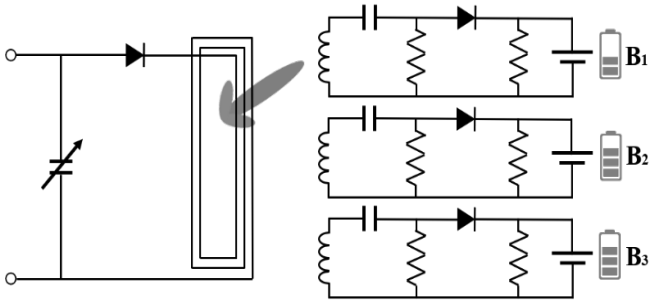


Figure 2: Proposed battery cell balancing with using near-field coupling

This proposed method is more efficient than the conventional passive type. The passive type consumes energy, leading to low efficiency, particularly because the balancing is based on the battery with the lowest SOC. However, the proposed method is more efficient because it supplements energy rather than wasting it. The energy efficiency of wireless power transmission can be divided into transmitter efficiency, receiver efficiency, and coil efficiency. With the transmitter efficiency being less than 80%, receiver efficiency at 95%, and coil efficiency less than 85%, the combined overall efficiency can reach approximately 64.6%[14-15]. This makes the proposed method more suitable for high efficiency than the passive type.

Additionally, the proposed method is faster in balancing speed than the active type. The primary reason for the slower balancing speed of the active type is because it can only transfer energy to adjacent batteries. On the other hand, our proposed method allows energy transfer even to relatively distant cells by variably adjusting the transmitter's capacitor to achieve impedance matching with a specific receiver. This results in a faster balancing speed than the active type.

4. Impedance Matching

Impedance serves as a measure of electrical flow resistance, and impedance matching refers to all methods aimed at reducing reflection caused by differences in impedance between the input and output terminals. That is, the more closely the impedances of both terminals are matched, the more the reflection due to the impedance difference reduces, leading to less energy loss and higher efficiency[16]. Such impedance matching techniques enable selective cell balancing. By designing so that the impedance matches with a specific battery and differs from other batteries, it is possible to deliver energy only to the desired cells. The proposed method to implement this involves differentiating the mutual inductance and coupling coefficient between the transmitter and each receiver depending on the location of the receivers, based on the transmitter. This leads to unique impedance differences, enabling the selection of specific receiver cells.

4.1. Mutual Inductance

Mutual inductance is a value that represents the magnetic field interaction between two coils. As shown in Figure 3, When the current flowing through a coil changes, the magnetic field passing through that coil also changes.

This change induces an electrical change in an adjacent coil, and this interaction is referred to as mutual inductance. This varies according to distance, coil orientation, and position, leading to each battery cell having a unique value.

4.2. Coupling coefficient

The coupling coefficient is calculated using the mutual inductance and the inductance of the coil itself, and represents the efficiency of energy transfer between two coils. The higher the coupling coefficient, the higher the energy transmission efficiency, so it's crucial to adjust the coupling coefficient appropriately. This also varies according to cell distance, similar to mutual inductance.

$$k = \frac{M}{\sqrt{L_1 L_2}} \tag{1}$$

Through such impedance matching, the energy transmission efficiency between battery cells can be increased, and based on this, efficient cell balancing can be implemented.

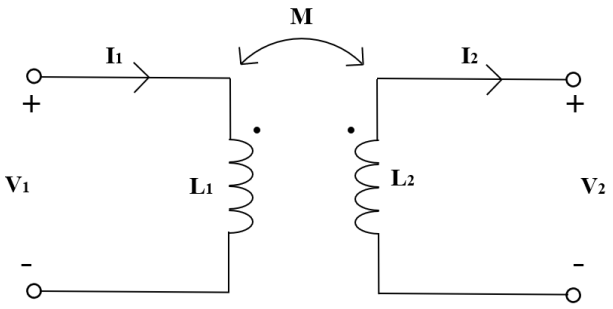


Figure 3: Mutual inductor circuit

5. Transmitter and Receiver Coil Design

Coil patterns can be designed in various forms such as square, circular, or hexagonal. However, in this study, we used the square pattern for easy modeling, based on the previous research which states that the difference between patterns is not significant at the frequency of 13.56MHz utilized in this research[1]. Furthermore, we decided on an LC circuit configuration suitable for selective battery cell balancing according to the Q-factor.

5.1. Q-factor

As shown in Figure 4, the Q-factor, defined by equation (2), represents the ratio of the resonance frequency ω_0 to the 3dB bandwidth, defined as the difference between the higher frequency ω_h and the lower frequency ω_l, situated 3dB below ω_0. For effective cell balancing application, a sufficiently large Q-factor is indispensable, signifying reduced insertion loss at specific frequencies and significant increase in insertion loss with minor frequency deviations. This narrow bandwidth impedance matching is vital for the proposed selective cell balancing approach. Analyzing equations (3) and (4), it's evident that for a consistent resonance frequency, a series LC circuit demands a larger inductance for a heightened Q, while a parallel LC circuit requires a smaller inductance for the same effect.

Given the physical constraints like battery size and coil diameter, the receiver coil was designed with a parallel configuration, where a reduced inductance yields a higher Q. Conversely, to ensure efficient energy transmission via impedance matching, the transmitter coil, designed in a series configuration, incorporates a larger inductance for an enhanced Q. This culminated in the LC circuits for both transmission and reception units being structured as Series-Parallel (SP).

$$Q = \frac{\omega_0}{\omega_h - \omega_l} \tag{2}$$

$$Q(series\ LC) = \omega_0 \frac{2W_m}{P_{loss}} = \frac{\omega_0 L}{R} = \frac{1}{\omega_0 RC} \tag{3}$$

$$Q(shunt\ LC) = \omega_0 \frac{2W_m}{P_{loss}} = \frac{R}{\omega_0 L} = \omega_0 RC \tag{4}$$

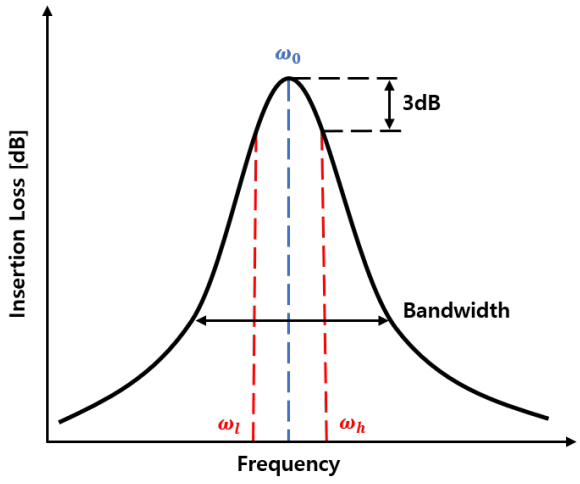


Figure 4: Q-factor in resonance circuit

Building on the crucial role of the Q-factor, the transmitter was crafted with an increased inductance, embodying a series LC circuit, fine-tuned for selective cell balancing. In parallel, the receiver was architected with diminished inductance, aligning with

a parallel LC circuit, to cater to the same balancing purpose. Factors like pattern spacing, internal and external diameters, and thickness play pivotal roles in determining coil inductance. For instance, in Figure 5, pattern spacings for the transmitter and receiver coils were fixed at 1mm and 2mm respectively, echoing past research indicating augmented inductance with reduced spacing. Coil measurements in this study were not merely based on theoretical derivations; instead, they were corroborated through actual coil measurements and simulations, ensuring accuracy by accounting for potential production discrepancies.

5.2. Coil Design

By leveraging the Q-factor, it was ascertained that the inductance of the transmitter coil can be designed to be larger, ensuring that the serial LC circuit configuration is suitable for selective cell balancing. Conversely, by designing the inductance of the receiver coil to be smaller, it was confirmed that the parallel LC circuit configuration is suitable for selective cell balancing. Figure 6 presents the blueprint of the transmitter and receiver coils designed using 3D EM Simulation, as well as the fabricated PCB. (a) Represents the transmitter coil, and (b) represents the receiver coil.

The inductance of the coil is determined by various variables as shown in Figure 6, and the inductance changes even if only one of these conditions changes. The smaller the pattern spacing 's' of the transmitter and receiver coil, the larger the inductance; therefore, the transmitter was designed with 1mm and the receiver with 2mm. Similarly, the smaller the internal diameter 'Di', the larger the inductance; thus, the transmitter was designed with 1mm and the receiver with 5mm. As can be seen from Figure 6, considering the size of the battery, the receiver's outer diameter 'Do' was 20x19mm, while the transmitter, considering the size of the receiver, was 81x31mm. The thickness 'w' was designed to be the same at 1mm. There are various papers that infer inductance values using these variables[17-18], but in this paper, the inductance is inferred by comparing the actual measurement results of the coil with the simulation.

5.3. De-embedding

Circuits that use RF (Radio Frequency), such as 13.56MHz, cannot be measured for voltage and current like normal circuits. This is because as the frequency increases, the size of the wavelength decreases, and the voltage and current measurement values change depending on the circuit location. Therefore, RF circuits should use S-parameters, which interpret the circuit as a ratio of output voltage to input voltage, and in this paper, the proportion of the input signal reaching the output port was measured for efficiency verification.

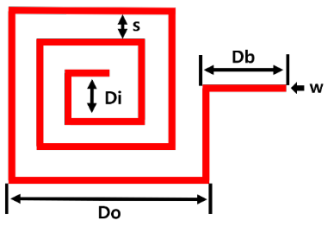
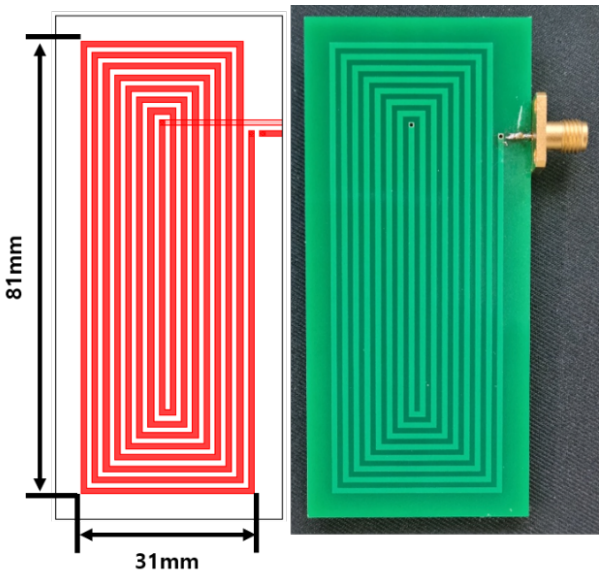
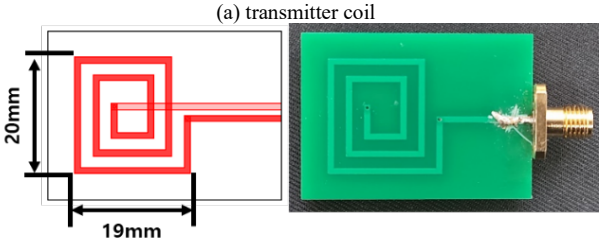


Figure 5: Coil variables



(a) transmitter coil



(b) receiver coil

Figure 6: Transmitter and receiver coil designed using 3D EM Simulation, as well as the fabricated PCB

To measure S-parameters, an NA(Network Analyzer) was used. For the accuracy of the measurement results, the NA's own calibration function is used to remove error components, and there are standardized methods such as SOLT(Short, Open, Load, Through), TRL(Through, Reflect, Line). In this paper, an SOLT calibration kit was used, and components such as phase delay that are not removed by general correction should be further corrected through de-embedding. De-embedding is not a standardized method like ordinary correction, so it must be conducted according to the research environment. When the circuit in the open state is represented as a Smith chart, it is indicated in the center to the right like the red dot in Figure 7, and the degree of phase delay can be understood by arbitrarily opening the fabricated transmitter and receiver PCB coil and comparing it with the measurement. The transmitter and receiver PCB coil data measured with the NA was adjusted in the simulation so that the electrical length of the ideal Transmission Line came out as in Figure 7. Figure 8 shows the result, with the blue plot being the result measured with the NA and the red plot being the result adjusted for electrical length. Figure 9 is the result of the transmitter and receiver de-embedding, and including up to the transmission line, it becomes an inductor model with phase delay removed. (a) is the result of transmitter de-embedding with an electrical length of –17.4° at 1GHz. (b) is the result of receiver de-embedding, changing the electrical length to –18.5°. Therefore, by applying the electrical length obtained through the de-embedding of the transmitter and receiver to the measurement results, accurate results with phase delay removed can be obtained.

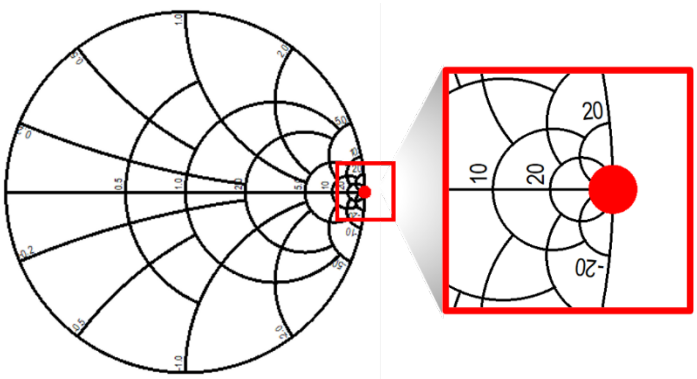
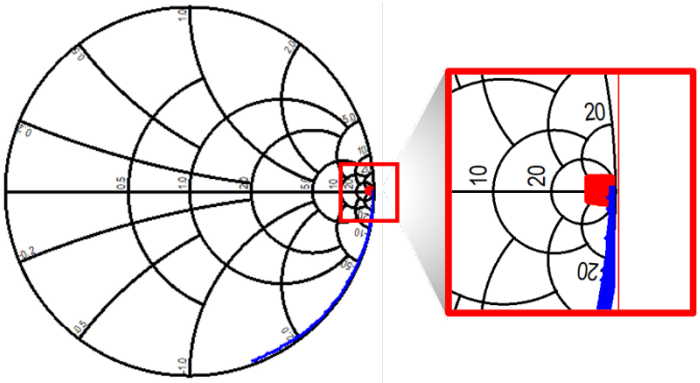


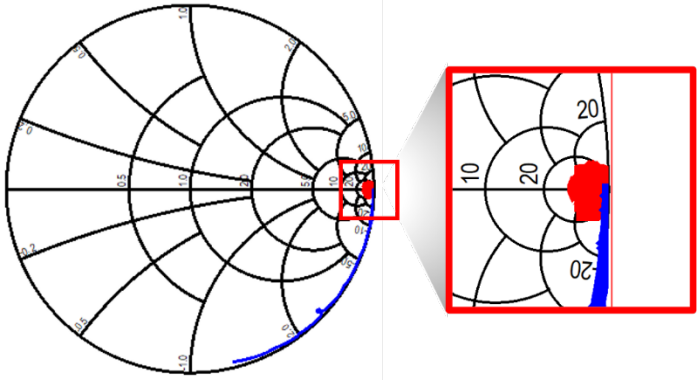
Figure 7: Open circuit on Smith chart

5.4. Inductance inference

Figure 10 shows the results of comparing the corrected measurement results through the transmitter's de-embedding, 3D EM Simulation, and Schematic 2600nH. By comparing the phase using the s-parameter and Smith chart, there was an error of less than 3% at 13.56MHz, which makes it possible to represent the coil with one inductor of the Schematic. The receiver was compared with 130nH, and the results were similar to the transmitter. Figure 10 shows the results of comparing the measurement results of the LC circuit determined through modeling of the transmitter and receiver and the inferred LC circuit. The schematic circuit of the resonant circuit in this application is shown in Figure 11.

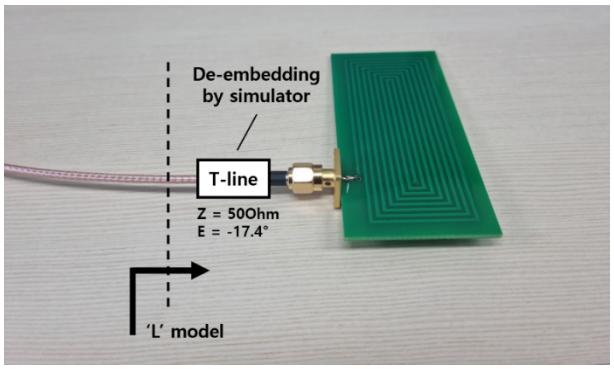


(a) impedance of transmitter on Smith chart

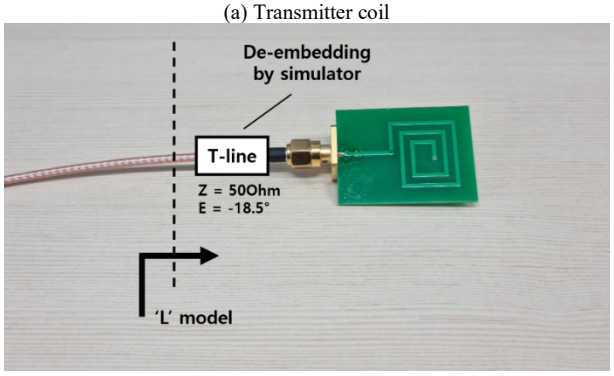


(b) Impedance of receiver on Smith chart

Figure 8: Transmitter and receiver coil open state measurement results(blue) electrical length adjusted results(red)



(a) Transmitter coil



(b) Receiver coil

Figure 9: De-embedding of transmitter/receiver coil

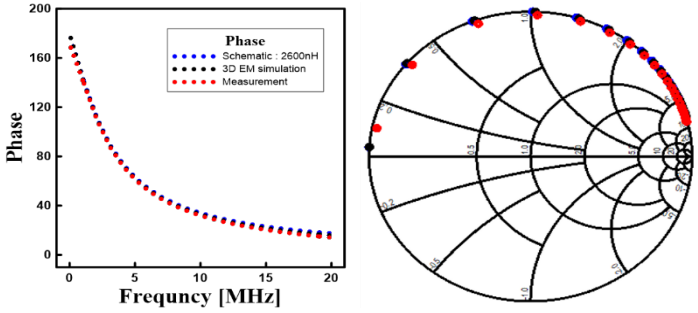


Figure 10: Transmitter de-embedding, 3D EM Simulation and Schematic 2600nH comparison results, error of less than 3% at 13.56MHz

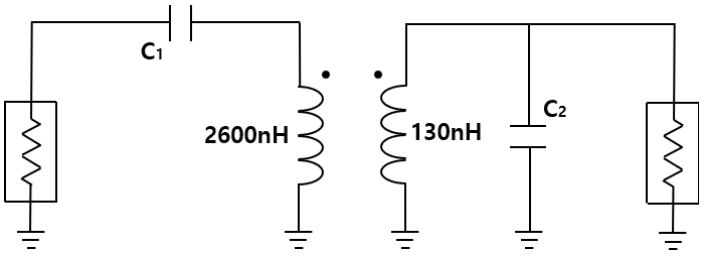


Figure 11: The schematic of measured resonance coils

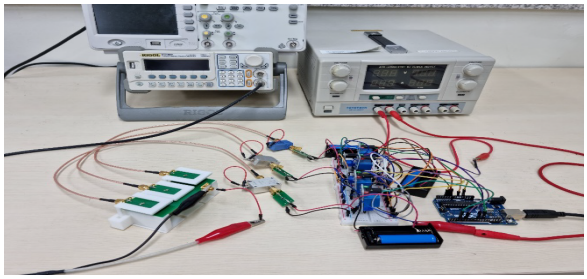


Figure 12: Cell balancing measurement set-up

5.5. Cell balancing measurement

Figure 12 shows an experiment with three battery cells with the proposed near-field coupling. One of the three cells has less charge than the other two. Figure 13 compares the regular balancing results with the balancing results obtained by implementing Near-Field Coupling using the coils and inferred capacitance modeled through this paper. The X-axis is time and the Y-axis is the voltage difference with the reference battery. The red graph represents the regular balancing results, and the blue graph represents the balancing results with added Near-field. Assuming that balancing was successfully achieved when the voltage difference with the reference battery decreased to less than 1%, the regular balancing took about 248 minutes, while the balancing with added Near-Field took about 121 minutes, nearly twice as fast as the former.

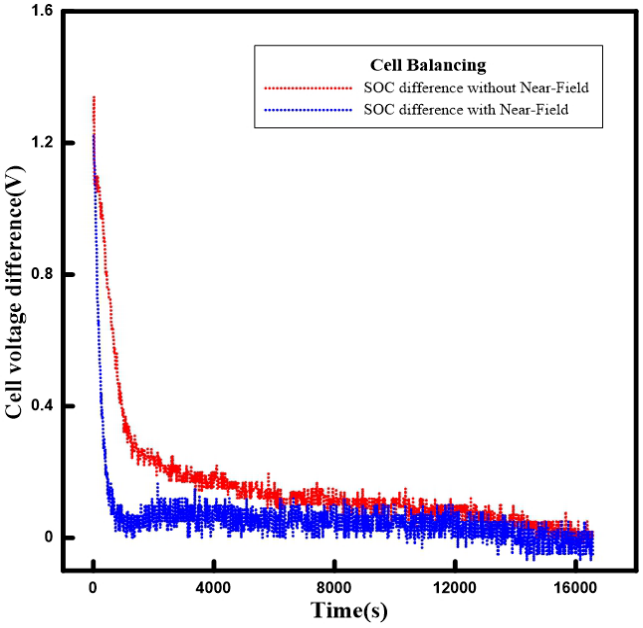


Figure 13: Cell voltage difference with and without near-field coupling

6. Conclusion

Our study confirmed that configuring the transmitter with a serial LC circuit and the receiver with a parallel LC circuit, as determined through the Q-factor, is more suitable for selective battery cell balancing. The coil of the receiver was constructed taking into account the battery size, which consequently determined the size of the transmitter coil. After the coil construction, uncorrected parasitic components were eliminated through additional correction processes such as de-embedding. Additionally, the electrical length of the transmitter coil's t-line was adjusted to -17.4° at 1GHz and the receiver coil to -18.5° to remove phase delay.

A comparison of the more accurate measurement results obtained through de-embedding, 3D EM Simulation, and Schematic with the s-parameter and Smith chart revealed less than 3% error at 13.56MHz. As the Schematic was compared using a single inductor, it was confirmed that the coil to be produced could be represented by a single inductor. Accordingly, we inferred that the transmitter coil was 2600nH and the receiver coil was 130nH. A simple test of cell balancing with three battery cells using the proposed method confirms that cell balancing is achieved compared to no cell balancing. The results of this experiment confirm that there is not enough unwanted coupling between the coils

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This Research was supported by Research Funds of Mokpo National University in 2022

References

[1] J. Jeon, W. Park, S. Pyo and D. Lee, "Coil design and de-embedding for a novel cell balancing circuit using near-field coupling," 2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), 546-548, 2022, doi: 10.1109/ITC-CSCC55581.2022.9895073.

[2] A. Qazi et al., "Towards Sustainable Energy: A Systematic Review of Renewable Energy Sources, Technologies, and Public Opinions," in IEEE Access, **7**, 63837-63851, 2019, doi: 10.1109/ACCESS.2019.2906402.

[3] S. Thangavel, D. Mohanraj, T. Girijaprasanna, S. Raju, C. Dhanamjayulu and S. M. Muyeen, "A Comprehensive Review on Electric Vehicle: Battery Management System, Charging Station, Traction Motors," in IEEE Access, **11**, 20994-21019, 2023, doi: 10.1109/ACCESS.2023.3250221.

[4] X. Xu, M. Bishop, D.G. Oikarinen and C. Hao, "Application and modeling of battery energy storage in power systems," CSEE journal of power and energy system, **2**(3), 82-89, 2016, doi: 10.17775/CSEEJPES.2016.00039

[5] J. Shen, S. Dusmez and A. Khaligh, "Optimization of Sizing and Battery Cycle Life in Battery/Ultracapacitor Hybrid Energy Storage Systems for Electric Vehicle Applications," in IEEE Transactions on Industrial Informatics, **10**(4), 2112-2121, 2014, doi: 10.1109/TII.2014.2334233.

[6] Q. Ouyang, W. Han, C. Zou, G. Xu and Z. Wang, "Cell Balancing Control For Lithium-Ion Battery Packs: A Hierarchical Optimal Approach," in IEEE Transactions on Industrial Informatics, **16**(8), 5065-5075, 2020, doi: 10.1109/TII.2019.2950818.

[7] D. L. Thompson, J. M. Hartley, S. M. Lambert, M. Shiref, G. D. J. Harper, E. Kendrick, P. Anderson, K. S. Ryder, L. Gaines and A. P. Abbott, "The importance of design in lithium ion battery recycling – a critical review" Green Chem, **22**, 7582-7584, 2020, doi: 10.1039/D0GC02745F

[8] M. A. Hannan, M. M. Hoque, A. Hussain, Y. Yusof and P. J. Ker, "State-of-the-Art and Energy Management System of Lithium-Ion Batteries in Electric Vehicle Applications: Issues and Recommendations," in IEEE Access, **6**, 19362-19378, 2018, doi: 10.1109/ACCESS.2018.2817655.

[9] E. Hossain, D. Murtaugh, J. Mody, H. M. R. Faruque, M. S. Haque Sunny and N. Mohammad, "A Comprehensive Review on Second-Life Batteries: Current State, Manufacturing Considerations, Applications, Impacts, Barriers & Potential Solutions, Business Strategies, and Policies," in IEEE Access, **7**, 73215-73252, 2019, doi: 10.1109/ACCESS.2019.2917859.

[10] J. Qi and D. Dah-Chuan Lu, "Review of battery cell balancing techniques," 2014 Australasian Universities Power Engineering Conference (AUPEC), 1-6, 2014, doi: 10.1109/AUPEC.2014.6966514.

[11] M. Daowd, N. Omar, P. Van Den Bossche and J. Van Mierlo, "Passive and active battery balancing comparison based on MATLAB simulation," 2011 IEEE Vehicle Power and Propulsion Conference, 1-7, 2011, doi: 10.1109/VPPC.2011.6043010.

[12] Das, Utpal Kumar, et al. "Advancement of lithium-ion battery cells voltage equalization techniques: A review." Renewable and Sustainable Energy Reviews **134**, 110227, 2020, doi: 10.1016/j.rser.2020.110227.

[13] Diao, Weiping, et al. "Active battery cell equalization based on residual available energy maximization." Applied energy **210,** 690-698, 2018, doi: 10.1016/j.apenergy.2017.07.137

[14] A. N. Laskovski and M. R. Yuce, "Class-E oscillators as wireless power transmitters for biomedical implants," 2010 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL 2010), 1-5, 2010, doi: 10.1109/ISABEL.2010.5702913.

[15] G. Lee, B. H. Waters, B. J. Mahoney, J. R. Smith and W. S. Park, "An investigation of cross-coupling for magnetically coupled wireless power transfer," 2013 Asia-Pacific Microwave Conference Proceedings (APMC), 80-82, 2013, doi: 10.1109/APMC.2013.6695197.

[16] T. C. Beh, M. Kato, T. Imura, S. Oh and Y. Hori, "Automated Impedance

Matching System for Robust Wireless Power Transfer via Magnetic Resonance Coupling," in IEEE Transactions on Industrial Electronics, **60(**9) 3689-3698, 2013, doi: 10.1109/TIE.2012.2206337.

[17] K. Chen and Z. Zhao, "Analysis of the Double-Layer Printed Spiral Coil for Wireless Power Transfer," in IEEE Journal of Emerging and Selected Topics in Power Electronics, **1(**2), 114-121, 2013, doi: 10.1109/JESTPE.2013.2272696.

[18] U. -M. Jow and M. Ghovanloo, "Design and Optimization of Printed Spiral Coils for Efficient Transcutaneous Inductive Power Transmission," in IEEE Transactions on Biomedical Circuits and Systems, **1(**3), 193-202, 2007, doi: 10.1109/TBCAS.2007.913130.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 4, 73-80 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

MRI Semantic Segmentation based on Optimize V-net with 2D Attention

Zobeda Hatif Naji Al-azzwi [*,1], Alexey N. Nazarov [2]

[1] *Institution Moscow of Physics and Technology, Department of School of Radio Engineering and computer Technology, 101000, Russian Federation*

[2] *Federal Research Center Computer Science and Control Russian Academy of Sciences, Moscow, Russia, 101000, Russian Federation*

ARTICLE INFO

Article history:
Received:19 April, 2023
Accepted:26 June, 2023
Online: 25 July, 2023

Keywords:
Semantic segmentation
Deep learning,
Attention-V-net

ABSTRACT

Over the past ten years, deep learning models have considerably advanced research in artificial intelligence, particularly in the segmentation of medical images. One of the key benefits of medical picture segmentation is that it allows for a more accurate analysis of anatomical data by separating only pertinent areas. Numerous studies revealed that these models could make accurate predictions and provide results that were on par with those of doctors. In this study, we investigate different methods of deep learning with medical image segmentation, like the V-net and U-net models. Improve the V-net model by adding attention in 2D with a decoder to get high performance through the training model. Using tumors of severe forms, size, and location, we downloaded the BRAST 2018 data set from Kaggle and manually segmented structural T1, T1ce, T2, and Flair MRI images. To enhance segmentation performance, we also investigated several benchmarking and preprocessing procedures. It's significant to note that our model was applied on Colab-Google for 35 epochs with a batch size of 8. In conclusion, we offer a memory-effective and effective tumor segmentation approach to aid in the precise diagnosis of oncological brain diseases. We have tested residual connections, decoder attention, and deep supervision loss in a comprehensive ablation study. Also, we looked for the U-Net encoder and decoder depth, convolutional channel count, and post-processing approach.

1. Introduction

In the past several years, convolutional neural networks (CNNs) have been used to solve issues in the disciplines of computer vision and medical image analysis. By now, deep learning-based image segmentation has gained a solid reputation as a reliable technique for picture segmentation. As the initial and crucial step in the pipeline for diagnosis and treatment, it has been routinely employed to divide homogeneous areas [1]. Today, many people experience illness, including cancer. Brain tumors can be either benign or malignant. A brain tumor is an enlargement of brain cells or cells close to the brain. Brain tissue can give rise to brain tumors. The membranes that encircle the surface of the brain, the pituitary gland, and the pineal gland are neighboring structures. An unchecked cell proliferation that threatens the central nervous system's survival is known as a brain tumor. Primary and secondary brain cancers fall into separate categories. Primary brain malignancies begin in brain cells, as opposed to secondary cancers, which spread to the brain from other organs. Primary brain tumors are gliomas. High-grade glioblastoma (HGG) and low-grade glioblastoma (LGG) are two more kinds of gliomas. The results of brain tumor segmentation can be utilized to generate quantitative measurements for glioblastoma diagnosis and treatment planning. Segmenting 3D modalities is a difficult task involving deviations and errors, in contrast to radiologists manually evaluating MRI modalities to gain quantitative information. [2]. In the medical industry, segmenting an image into its component components to extract features is a crucial task. Picture segmentation, which often comprises semantic segmentation and instance segmentation, is a significant challenge in the study of computer vision (CV).

Multimodal Brain Tumor Image Segmentation (BRATS), a yearly workshop and competition, has been held in recent years to compare several benchmark techniques created to segment the

* Corresponding Author: Zobeda Hatif Naji Al-azzwi, Iraq. Baqubqa, & zhn0hussen@gmail.com

brain tumor [3]. In clinical contexts, common standard imaging modalities include ultrasound, magnetic resonance imaging (MRI), and computed tomography (CT). Because the multiple image contrasts of these MRI protocols can be used to extract crucial supplemental information, numerous prior studies have demonstrated that multimodal MRI protocols can be used to identify brain cancers for treatment planning. T2-weighted fluid-attenuated inversion recovery (FLAIR), T1-weighted (T1), T1-weighted contrast-enhanced (T1c), and T2-weighted (T2) are among the multimodal MRI techniques. Variations in the organs' size, location, and shape make this problem more challenging. Convolutional Neural Networks (CNNs) can, however, be utilized to develop automatic segmentation techniques using MRI data [4], [5]. Deep learning algorithms produce more robust features than conventional discriminative models based on pre-defined features because they learn a hierarchy of progressively more complex task-specific features directly from the data. Deep CNNs have produced outstanding results for segmenting brain tumors [6]-[14]. In this work, we aim to segment Tumor brain MRI volumes. Due to the many appearances the tumor brain can take on in several scans because of deformations and variations in the intensity distribution, this is a difficult assignment. Moreover, due to field inhomogeneity, MRI volumes are frequently impacted by artifacts and distortions. One of the primary problems with brain lesion identification is that, because lesions only affect a small portion of the brain, unsophisticated training algorithms are more likely to choose the pointless choice of null detection. and therefore can extract the features of the three regions with greater accuracy. As a result, it has produced impressive results and made outstanding contributions to the clinical diagnosis and care of patients with brain tumors [15]. There are a lot of studies involving deep learning techniques and explaining them in details [16]-[18]. All instances of a class are found using instance segmentation, with the added ability to distinguish between distinct instances of any segment class[19]. Segmentation method entered in many medical and biological specialties [20],[21].Tumor brain segmentation interesting task to many author [22], [23]. Deep learning important technique for many applications [24].

2. Related work

The researcher In this study [5], They employed ROI masks to restrict the networks' training to relevant voxels and a cascade of two CNNs, inspired by the V-Net design, to segment brain tumors. This architecture facilitates dense training on problems with highly skewed class distributions, such as brain tumor segmentation, by limiting training to the vicinity of the tumor area. The results are based on data from the BraTS2017 Training and Validation sets. In this paper [13], For the segmentation of 3D images, they suggested a volumetric, fully convolutional neural network-based approach. After being trained from beginning to end on MRI volumes of the prostate, the network learns to predict segmentation for the full volume at once. This work introduces an objective function that optimizes during training and is based on the Dice coefficient. They deal with situations where the number of foreground and background voxels differs significantly. Histogram matching and random non-linear adjustments were utilized to augment the data and compensate for the minimal number of annotated volumes that were available for training. In this paper [14] suggested the AGSE-VNet framework for automatically segmenting MRI data from brain tumors. Using the channel relationship, the Squeeze and Excite (SE) module and Attention Guide Filter (AG) module are used to automatically enhance the useful information in the channel and suppress the useless information, and the attention mechanism is used to guide the edge detection and eliminate the influence of irrelevant information. Utilizing the online verification tool for the BraTS2020 challenge, results have been analyzed. Verification focuses on the fact that the Dice scores for the whole tumor (WT), tumor core (TC), and enhanced tumor (ET) are, respectively, 0.68, 0.85, and 0.70. Conclusion Despite the fact that MRI images have varying intensities, AGSE-VNet is unaffected by the tumor's size and therefore can extract the features of the three regions with greater accuracy. As a result, it has produced impressive results and made outstanding contributions to the clinical diagnosis and care of patients with brain tumors. In this [15] the researcher used 2D V-net K for tumor brain segmentation and enhancing predication they used data set BRATS2020.

3. Materials and Methods

3.1. Segmentation

The term "image" refers to a matrix of pixels that are ordered in rows and columns according to coordinates. There are numerous methodologies for constructing an image, including enhancement, detection, compression, segmentation, and more. As an alternative, segmentation is the act of labeling pixels.. Image segmentation is a technique used in the field of image processing that divides a digital image into several subgroups known as "image segments," which functions to simplify prospective processing or interpretation of the image by reducing the complexity of the original image. There is a distinct name for each pixel or component of a picture that belongs to the same category. Semantic segmentation, instance, and panoptic segmentation, are three prominent type of segmentation as shown in figure1.

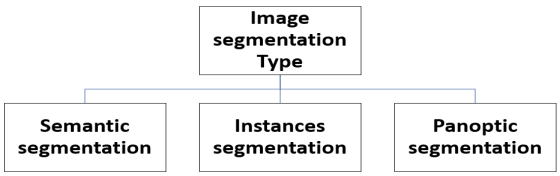


Figure 1: Show the approach of Image segmentation.

One of the most challenging problems in computer vision, particularly in medical image analysis, is semantic segmentation, the act of giving a class label to each pixel. Semantic segmentation helps with the automatic localization and detection of sick structures.[16], Semantic segmentation aims to classify different areas of a picture into categories. The segment could be viewed as a categorization issue at low resolution. Segmentation is used in computer-aided diagnosis to extract anatomical sections and has been a significant research area for information extraction[17]. When utilizing deep learning for semantic segmentation, the learning process is supervised and the training images containing actual classes that have been separated, also known as ground truth data, are connected to the labels. Semantic segmentation is used in numerous industries, including the medical sector.

When numerous instances are combined into one class, the problem description is frequently ambiguous when utilizing this approach. For example, the "people" class may be used to classify the entire crowd in a bustling scene. The semantic segmentation does not provide enough detail for this complex image.

Instance segmentation is a specific type of picture segmentation that focuses on locating instances of objects and defining their bounds. Instance segmentation has drawn interest in several computer vision disciplines, including autonomous vehicle control, drone navigation, and sports analysis. Numerous effective models have recently been created, and they can be divided into two groups: speed-focused and accuracy-focused. For this task to be used in real-time applications, accuracy and inference time are crucial.[19]. Thus, semantic segmentation differs from instance segmentation in that it treats many objects belonging to the same category as a single entity. As opposed to that, instance segmentation detects specific objects within the image[20].

Panoptic segmentation the most current segmentation task, known as panoptic segmentation, can be described as a combination of instance and semantic segmentation, where each instance of an object in the image is separated and its identification is foreseen. Popular activities like self-driving automobiles, where a significant amount of information about the immediate surroundings must be gathered with the aid of a stream of photos, find widespread application for panoptic segmentation algorithms. In figure2 [18] show the panoptic, semantic and instance segmentation.

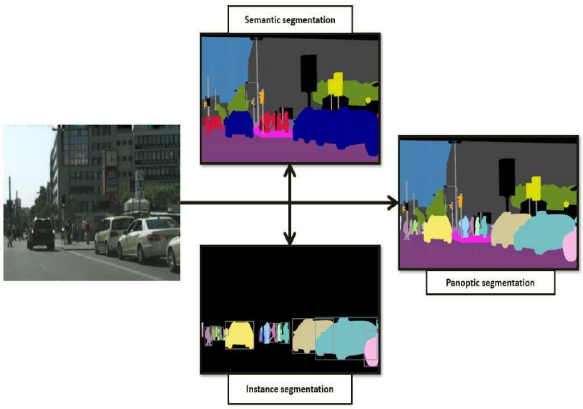


Figure 2: Demonstrate how instance segmentation, semantic segmentation, and panoptic segmentation vary from one another.

The segmentation technique belongs to the class of pixel cases. The first approach makes use of discontinuity to segment an image into lines, edges, and isolated points by identifying them based on rapid changes in local properties. The regions' boundaries are then deduced. By utilizing the homogeneity of spatially dense information, such as texture, intensity, and color, the second method's segmentation results are generated. The third strategy combines boundary-based segmentation and segmentation based on regions. [3]. Most current tumor segmentation techniques can be grouped into one of two big families based on how they segment tumors. The main drawback with image segmentation is over segmentation and high sensitivity to noise. Discriminative techniques focus on select local picture features that seem significant for the tumor segmentation job, directly learning the link between image intensities and segmentation labels without any domain knowledge. The most important picture segmentation techniques include threshold-based, edge-based, region-based, clustering-based, and "deep learning" based on artificial neural networks (3). Model training is used in machine learning-based image segmentation approaches to enhance the system's capacity to recognize key traits. For photographic segmentation tasks, deep neural network technology is highly effective.

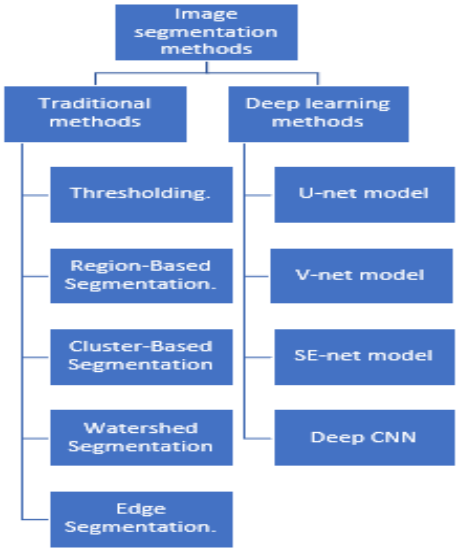


Figure 3: Show the traditional segmentation methods and deep learning methods

3.2. Deep learning methods with segmentation approach

Artificial neural networks, which are the sole foundation of the machine learning subfield of deep learning, will replicate the human brain in a manner like how neural networks do for deep learning. Because Deep Learning models can dynamically learn features from the input, they are ideal for applications like photo recognition, speech recognition, and natural language processing. The three most widely used deep learning designs are feedforward neural networks, convolutional neural networks (CNNs), and recurrent neural networks (RNN). A feedforward neural network (FNN), which has a linear information flow, is the most fundamental type of ANN. FNNs have been widely employed for tasks such as image classification, image segmentation, speech recognition, and natural language processing. Convolutional neural networks (CNNs) that were trained on manually labeled patient datasets are the foundation of most contemporary automatic segmentation methods. For image segmentation, a variety of neural network implementations and architectures are suitable. They frequently share the same essential components. An encoder is a group of layers that uses more specific, in-depth filters to extract visual characteristics. By using its earlier experience with a task that is comparable to this one (like image recognition), the encoder may be able to carry out segmentation tasks. A decoder is a collection of layers that gradually transforms the encoder's output into an interpretation according to the pixel resolution of the input image. To *enhance* model accuracy, skip

connections. Multiple long-range neural network connections that let the model recognize *significant information* at many scales are used. There are many models used to segment images; the most famous and important are U-net models[21] segment for biomedical image , V-net models [13] segment for volume medical image in follow section more description for each one.

a. V-net model:

The [13] proposed On both the left and right sides of the V-Net, there are many layers. The network's right and left halves alternately decompress and compress the signal until it is back to its original size. The network's left side is divided into several stages, each of which can operate in a different configuration. Each phase has between one and three convolutional layers. Each level involves the learning of a residual function. Each step's input is applied to its respective convolutional layers, subjected to non-linear processing, and added to the output of the final convolutional layer to enable the learning of a residual function. Comparison of this design to non-residual learning networks like U-Net ensures convergence.

The volumetric kernels used in each stage's convolutions have a 5x5x5 voxel size. (A voxel in 3D space corresponds to a value on a standard grid. Like how "voxelization" is used in point clouds, the word "voxel" is widely employed in 3D spaces.) The resolution of the compression path is decreased by stride 2 convolution using 2x2x2 voxel wide kernels. As a result, the objective of layer pooling is still accomplished while halving the size of the generated feature maps. The number of feature channels doubles at each stop along the V-compression Net's journey. Convolutional operations have a smaller memory footprint during training because they do not need switches to change the output of pooling layers back to their inputs for back-propagation. sampling backwards. They leave a lower memory footprint when training. It is possible to expand the receptive field by downsampling. A nonlinear activation function is employed with PReLU.

The network extracts feature from the lower resolution feature maps, enhances the spatial support for those features, and collects the required data to construct a two-channel volumetric segmentation. In each step, the inputs are deconvolved to increase their size, and then one to three convolutional layers are added, each of which uses half as many 5x5x5 kernels as the layer before it. Residual function is taught, just like in the left part of the network.

The output size of the final convolutional layer, which has a 1x1x1-kernel size, is the same as the input volume. The two feature mappings are generated. With the use of soft-max voxel-wise segmentation, these two output feature maps provide probabilistic foreground and background segmentations. lateral connections Similar to U-Net, compression results in the loss of location data (left). As a result, horizontal connections are used to transfer the features that were acquired from the left part of the CNN's first stages to the right part. By doing so, better precision of the final contour forecast and the correct part position data have been achieved. The model's convergence is accelerated by these linkages. Figure 4 shows the structure of V-net.

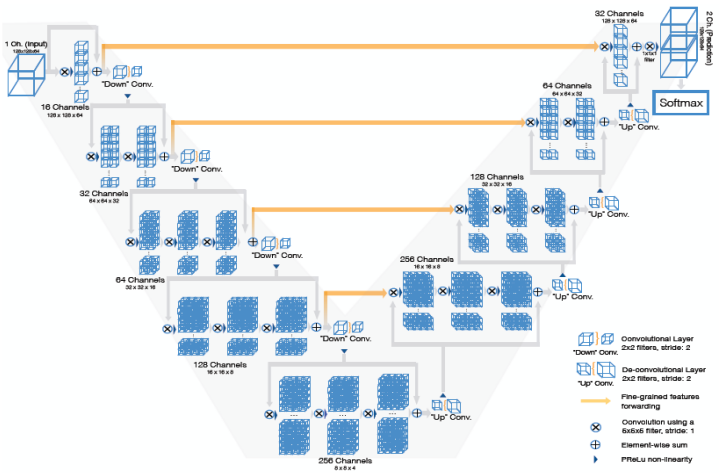


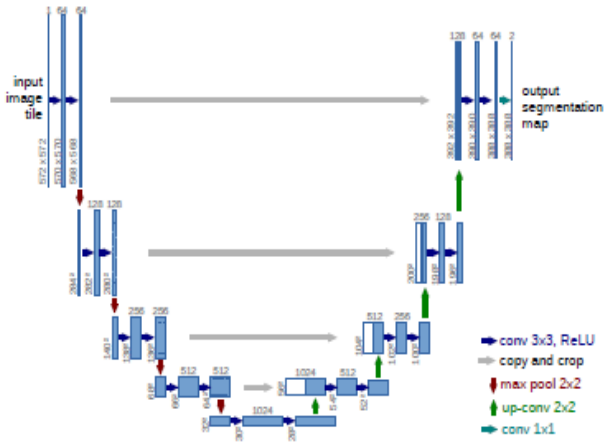
Figure 4: Show the structure of V-net [13]

b. U-net model:

The U-net model is a significant model that includes biomedical imagery. [21], take U shape as shown in figure 4. [21] it's have three part encoder, decoder and bottleneck. For classification-related issues, the tool that turns an image into a vector that can be used for classification is useful. However, in order to segment a picture, the feature map must also be turned into a vector, which must then be used to recreate the original image. This is a hard operation because it is far more difficult to convert a vector into a picture than the contrary, this is the main issue with U net-design. The three components of the U-net model are contraction, bottleneck, and expansion. In the U-Net model, the contraction is employed to expand a vector into a segmented image. This would preserve the structural integrity of the image and significantly lessen distortion.

The architecture can successfully learn the complex structures since there are twice as many kernels or feature maps after each block. A mediator between the contraction layer and the expansion layer is the bottom layer. After two 3X3 CNN layers, a 2X2 up convolution layer is applied.

The number of blocks for both expansion and contraction is the same. Yet, the extension component is where this architecture's core may be located. Several expansion building elements are included in addition to the contraction layer. A 2X2 up sampling layer is added to each block after the first two 3X3 CNN layers to handle the input. To maintain symmetry, the convolutional layer uses half as many feature mappings between each block. But each time, the input is also supplemented by feature maps from the associated contraction layer. As a result, the traits identified during the image's contraction would be considered during its reconstruction. The subsequent 3X3 CNN layer receives the output mapping, with the number of feature mappings matching the necessary number of segments. This model, U-net, has been used with the 2018 Brast data set in this paper to evaluate the optimization of V-net, as shown in Figure 6.



The figure 5: Show the U-net model

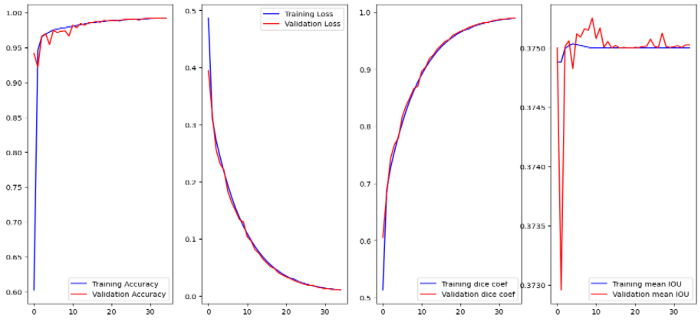


Figure 6: Show the training of U-net model with 2018Brast.

3.3. Attention 2D-layer.

Deep learning approaches such as attention models or attention mechanisms are used to add additional focus to a particular component. Focusing on something special and noting its unique value is what deep learning is all about. The attention mechanism was first made available by[24]. To address the bottleneck problem brought on by a fixed length encoding vector, where the decoder's access to the input's data would be severely limited. Because the representation's dimensionality would be limited to match that of shorter or simpler sequences, this is seen to be especially troublesome for long and/or complex sequences.

The attention layer's goal is to increase the machine translation encoder-decoder model's effectiveness. By integrating all the encoded input vectors in a weighted fashion, with the most pertinent vectors receiving the highest weights, the attention mechanism was developed to enable the decoder to use the most pertinent sections of the input sequence in a flexible manner.

As indicated in the findings section, the attention layer and V-net model were integrated in this study to increase training effectiveness. As seen in figure 7, the attention layer used in this study is composed of a convolution layer followed by Relu activation, convolution, followed by a sigmoid function.

Boost helpful functions and disable less helpful ones for the present work. The image is decompressed after each decoder receives the down sampling characteristics for the corresponding step. In the process of downsampling, the attention block is used to remove background noise and unnecessary details, while the image filter (edge information) directs the structural and feature data for the images. Remember that the model uses the skip connection idea to prevent gradient vanishing.

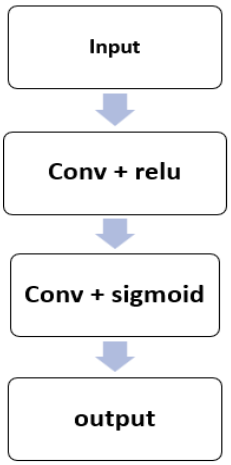


Figure 7: Show the structure of attention layer.

4. Proposed Method

Encoder and decoder components are symmetrical in the network's structure. The encoder is made up of four blocks, each of which contains several convolution layers, batch normalization, and PRelue. Between each block, there is a Max pooling layer with (2,2), which allows for down sampling and equivalent up sampling. All convolution layers have size 5x5with stride [1,1]. A full skip connection is joined with concatenation to keep the feature map and resolution of the image. Using skip connections following sampling at the decoding end, the lower-level features are fused with the first four scales of the feature map of the V-net.

Four blocks make up the decoder side, including an attention layer with a kernel size of (1,1) and a stride of 1. The layer's first three operations are convolution, relu activation, and sigmoid. Encoding equipment Batch normalization facilitates weight initialization and speeds up network training by enabling significantly higher learning rates. While building deeper networks, initializing weights might be challenging. The addition of a 2D attention layer can enhance this model. Many researcher worked on V- net model and optimized it [5,14]. The optimize in this research was by adding attention layer with encoder. There are full skip connections between encoder and decoder.

This model V-net makes use of the Brast 2018 data set to produce feature maps with modest values but concentrated, high-dimensional semantic interpretations after continuous convolution. To provide a segmentation result, the decoder is then repeatedly convolved and up sampled to the original size.

By batch normalizing, the sensitivity to the original beginning weights is reduced. To utilize a suitable stride at the conclusion of each block to reduce its resolution. The network's left side compresses the signal while the right side decompresses it until the signal's original size is reached.[13].

4.1. Evaluation criteria

To Evaluate this work, we need take in our consideration many criteria. We gauge the training of our model using the loss,

accuracy, mean IOU, dice coefficient, precision, sensitivity, and specificity.

a. Loss function: A loss function compares the target and anticipated output values to determine how well the neural network mimics the training data. During training, we make an effort to minimize this difference between the expected and desired outputs. determined by.

$$loss = \frac{1}{n}\sum_{i=1}^{n}(y - y')^2 \quad (1)$$

b. Accuracy: Accuracy is a measure of the model's performance across all classes. It is useful when all classes are given equal weight. It is calculated as the ratio of events that were successfully anticipated to all events that were predicted. One of the measures that is most frequently used when performing categorization is accuracy.. We explain each term in table 1.

$$Acc = \frac{T_p+T_N}{T_p+T_N+F_P+F_N} \quad (2)$$

Tabel.1: Describe the criteria for validity.

T_P	A true positive result is one that the model correctly identified as falling within the positive category.
T_N	When the model correctly predicts the negative class, the result is said to be True Negative.
F_P	False Negative: is a finding when the model forecasts the positive class wrongly.
F_N	False Negative: is a outcome genuine positive instances are labeled as negative

c. Mean IOU: mean intersection over union is used when calculating mean average precision.

$$mean_{IOU} = \frac{T_P}{T_P+F_P+F_N}, \quad (3)$$

d. Dice coefficient: A predicted segmentation's pixel-by-pixel agreement with the associated ground truth can be compared using the Dice coefficient. In image segmentation, the Dice score coefficient is a commonly used metric of overlap. The benefit of dice loss is that it does a good job of handling the class mismatch between the foreground and background pixel counts.

$$Dic = \frac{2|A\cap B|}{|A|+|B|} \quad (4)$$

e. Precision: One indicator of the model's performance is precision, or the quality of a successful prediction the model makes. By dividing the total number of positive predictions by the percentage of true positives, precision is determined (i.e., the number of true positives plus the number of false positives).

$$p = \frac{T_P}{T_P+F_P} \quad (5)$$

f. Sensitivity: One indicator of the model's performance is precision, or the quality of a successful prediction the model makes. By dividing the total number of positive predictions by the percentage of true positives, precision is determined.

$$sen = \frac{T_P}{T_P+F_N} \quad (6)$$

g. Specificity: The fraction of true negatives that the model correctly predicts is known as specificity. Specificity can be defined as the ability of the algorithm or model to predict a genuine negative of each possible category. It is sometimes referred to as the genuine negative rate in literature.

$$Sp = \frac{T_N}{T_N+F_P} \quad (7)$$

4.2. Descriptive for Data Set

The three sources for the BraTS 2018 challenge are indicated, respectively, by the designations "2013," "CBICA," and "TCIA." The training data set contains 20 cases of high-grade glioma (HGG) from the 2013 group, 88 cases from CBIC, and 102 cases from TCIA, as well as 10 cases from 2013 and 65 cases from TCIA participants who had low-grade gliomas (LGG), which are less active and infiltrative. [22] .One of dataset that is frequently used in the world of healthcare is BRATS 2018 which used in this paper include tow folder HHG and LLG each one has 2oo case patients we used 55 cases in this paper.

Per case, BRATS provides four multimodal MRI modalities. The four MRI modalities that the BRATS competition organizers provide, along with associated manual segmentation, include 3D brain MRIs with ground truth brain tumor segmentations annotated by physicians (T1, T1c, T2, and FLAIR): T1-weighted (T1), T1-weighted post-contrast (T1c), T2-weighted (T2), and fluid-attenuated inversion recovery (FLAIR). [6]. The link of data set; which used in this paper BRAST2018 can find on this https://www.kaggle.com/datasets/sanglequang/brats2018/download?datasetVersionNumber=10.

The volume dimension of every MRI in the BRATS 2018 dataset is 240 x240 x155. The data set which used in this study MRI 3D image (voxel) tumor brain which represented by X, Y, Z, 3D, non-informative voxels are found in medical images in general and brain MRIs. The preprocessing with this data started by converted the voxel to pixel it's mean 3D to 2D image to reduce the consuming time. The HGG have be used in this article have four modalities (T1, T1c, T2, and FLAIR). The data set in this paper divided to training and test, the test split to test and validation. The size of training is 2970 with dimension (192X192X4), and the size of test and validation is 990 with dimension (192X192X4). It is difficult to discern between a tumor and normal tissue despite the fact that tumor margins are typically fuzzy and that there is a considerable deal of variation in shape, location, and expression among patients. Despite recent developments in automated algorithms, segmenting brain tumors in multimodal MRI scans remains a difficult task in medical image processing [5]. The biggest obstacle to segmenting brain tumors is the class imbalance. Even less of the brain's tissue between 5 and 15% represents each tumor site [23].

5. Result and discussion

We improved and suggested V-Net-attention 2D based fully convolutional networks to address the brain tumor segmentation challenge. In essence, semantic segmentation is the task that deals with tumor recognition and segmentation. The data set of BRAST2018 divided to training, validation and test set with 35 epochs and the size of batch 8, because high memory consumption is still challenge in 3D-CNNs. The developed model was

implemented on colab and evaluated on brain magnetic resonance images MRI. The findings demonstrate that this proposed technique outperforms several of the already popular derivative models based on CNNs in terms of segmentation performance and tumor recognition accuracy. The experiment started by pre-processing to data set, input slice of the BRAST dataset which feed to model have dimension 192x192x4 (h x w x d, refereed to Hight, width and depth respectively) in this paper. In comparison to earlier deep learning-based studies on this subject, we train U-net on the same BRAST2018 data set as indicated in the preceding section. The results reveal that training improves V-net by adding an attention layer, which has a decent performance on HGG as show in figure8. Take optimizers Adam with learning rate (lr=1e-5).

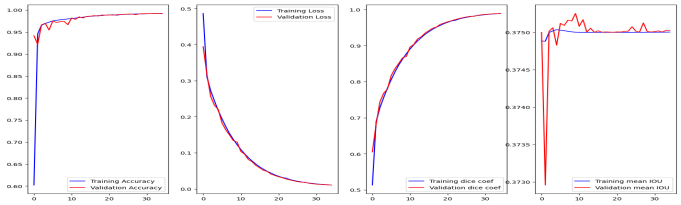


Figure 8: Show the training of V-net attention 2D on the HGG.

The measurement of classification accuracy as well as localization correctness is thought to be required in the analysis of semantic segmentation, which can be complex. The objective is to evaluate how closely the predicted (prediction) and annotated segmentation differ (ground truth). We used different benchmark to evaluate the optimized model V-net Attention 2D where take in consideration loss, Accuracy, Mean_IOU, dice_coef, precision, sensitivity, specificity demonstrated in table2, Instead, it is a helper metric that assesses the level of agreement between the ground truth and prediction.

Table 2: Show the result through training of model V-net attention 2D.

	loss	accuracy	MeanIOU	Dice_coef	precision	sensitivity	specificity
Train result	0.0105	0.99203	0.37501	0.9894	0.9926	0.99156	0.99754
Validation result	0.0108	0.99174	0.37502	0.9891	0.9923	0.99128	0.99744
Test result	0.0107	0.99175	0.37501	0.9892	0.9924	0.99124	0.99747

Here, we improve a strategy for segmenting medical pictures using fully convolutional neural networks that have been end-to-end trained to locate tumors in MRI scans. In contrast to previous recent techniques, Convolutions and BRAST 2018's slice-wise input volume processing are recommended. One of the primary issues with tumor brain is that since these lesions only impact a small portion of the brain, simple training methods are predisposed to the useless option of null detection. Only brain tissue will be used to train the modified V-NET first network model to produce raw data and predict the normal region, as shown in figure 9.

6. Conclusion

In the study developed a deep convolution network-based, fully automatic method for locating and classifying brain cancers in this study. Through tests on a well-known benchmarking dataset (BRATS 2018), which covers both HGG and LGG patients, we have demonstrated that our method can deliver both efficient and robust segmentation relative to manually determined ground truth. Our V-Net 2D attention-based deep convolution networks may also yield outcomes that are comparable for the overall tumor region and even better for the core tumor region when compared to existing state-of-the-art methods. The presented method can be used to generate a patient-specific brain tumor segmentation model without the requirement for manual assistance, which might enable objective lesion assessment for clinical tasks like diagnosis, treatment planning, and patient monitoring.

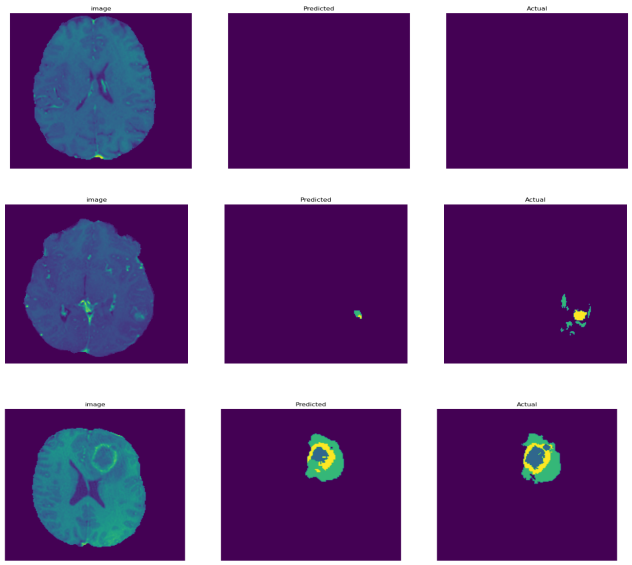


Figure 9: Illustration of the V-net-Attention 2D for normal cases HGG segmentation results in BraTS2018.

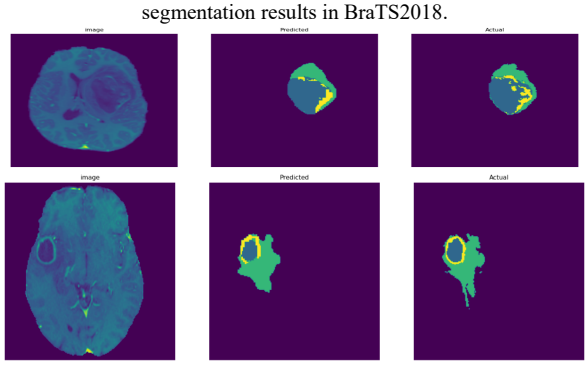


Figure 10: Demonstration of the V-net-Attention 2D segmentation findings for cases of brain tumors and brain tissue from BraTS2018.

Acknowledgment

The accomplishment of this study was supported by a grant from the institution Moscow Institute of Physics and Technology.

References

[1] M.H. Hesamian, W. Jia, X. He, P. Kennedy, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges," Journal of Digital Imaging, **32**(4), 582–596, 2019, doi:10.1007/s10278-019-00227-x.

[2] P. Ahmad, S. Qamar, L. Shen, A. Saeed, "Context Aware 3D UNet for Brain Tumor Segmentation," Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **12658 LNCS**, 207–218, 2021, doi:10.1007/978-3-030-72084-1_19.

[3] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.A. Weber, T. Arbel, B.B. Avants, N. Ayache, P. Buendia, D.L. Collins, N. Cordier, J.J. Corso, A. Criminisi, T. Das, H. Delingette, Ç. Demiralp, C.R.

Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," IEEE Transactions on Medical Imaging, **34**(10), 1993–2024, 2015, doi:10.1109/TMI.2014.2377694.

[4] H.T. Le, H.T. Thu Pham, "Brain tumour segmentation using U-Net based fully convolutional networks and extremely randomized trees," Vietnam Journal of Science, Technology and Engineering, **60**(3), 19–25, 2018, doi:10.31276/vjste.60(3).19.

[5] A. Casamitjana, M. Catà, I. Sánchez, M. Combalia, V. Vilaplana, "Cascaded V-Net using ROI masks for brain tumor segmentation," Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **10670 LNCS**, 381–391, 2018, doi:10.1007/978-3-319-75238-9_33.

[6] M. Ghaffari, A. Sowmya, R. Oliver, "Automated Brain Tumour Segmentation Using Cascaded 3D Densely-Connected U-Net," Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **12658 LNCS**, 481–491, 2021, doi:10.1007/978-3-030-72084-1_43.

[7] H. Dong, G. Yang, F. Liu, Y. Mo, Y. Guo, "Automatic brain tumor detection and segmentation using U-net based fully convolutional networks," Communications in Computer and Information Science, **723**, 506–517, 2017, doi:10.1007/978-3-319-60964-5_44.

[8] Q. Jia, H. Shu, "BiTr-Unet: A CNN-Transformer Combined Network for MRI Brain Tumor Segmentation," Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **12963 LNCS**, 3–14, 2022, doi:10.1007/978-3-031-09002-8_1.

[9] A. Atiyah, K. Ali, "Brain MRI Images Segmentation Based on U-Net Architecture," Iraqi Journal for Electrical and Electronic Engineering, **18**(1), 21–27, 2022, doi:10.37917/ijeee.18.1.3.

[10] V.P. Gladis Pushpa Rathi, S. Palani, "Brain tumor detection and classification using deep learning classifier on MRI images," Research Journal of Applied Sciences, Engineering and Technology, **10**(2), 177–187, 2015.

[11] M.A. Al Nasim, A. Al Munem, M. Islam, M.A.H. Palash, M.M.A. Haque, F.M. Shah, "Brain Tumor Segmentation using Enhanced U-Net Model with Empirical Analysis," 2022.

[12] K. Munir, F. Frezza, A. Rizzi, "Deep Learning Hybrid Techniques for Brain Tumor Segmentation," Sensors, **22**(21), 2022, doi:10.3390/s22218201.

[13] F. Milletari, N. Navab, S.A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016, 565–571, 2016, doi:10.1109/3DV.2016.79.

[14] X. Guan, G. Yang, J. Ye, W. Yang, X. Xu, W. Jiang, X. Lai, "3D AGSE-VNet: an automatic brain tumor MRI data segmentation framework," BMC Medical Imaging, **22**(1), 2022, doi:10.1186/s12880-021-00728-8.

[15] D. Rastogi, P. Johri, V. Tiwari, "Brain Tumor Segmentation and Tumor Prediction Using 2D-VNet Deep Learning Architecture," Proceedings of the 2021 10th International Conference on System Modeling and Advancement in Research Trends, SMART 2021, (August 2022), 723–732, 2021, doi:10.1109/SMART52563.2021.9676317.

[16] Y. Azzi, A. Moussaoui, M.-T. Kechadi, "Semantic Segmentation of Medical Images with Deep Learning: Overview," Medical Technologies Journal, **4**(3), 568–575, 2020, doi:10.26415/2572-004x-vol4iss3p568-575.

[17] J. Moorthy, U.D. Gandhi, "A Survey on Medical Image Segmentation Based on Deep Learning Techniques," Big Data and Cognitive Computing, **6**(4), 2022, doi:10.3390/bdcc6040117.

[18] O. Elharrouss, S. Al-Maadeed, N. Subramanian, N. Ottakath, N. Almaadeed, Y. Himeur, "Panoptic Segmentation: A Review," 2021.

[19] S. Jung, H. Heo, S. Park, S.U. Jung, K. Lee, "Benchmarking Deep Learning Models for Instance Segmentation," Applied Sciences (Switzerland), **12**(17), 1–25, 2022, doi:10.3390/app12178856.

[20] T. Scherr, A. Bartschat, M. Reischl, J. Stegmaier, R. Mikut, "Best Practices in Deep Learning-Based Segmentation of Microscopy Images," Proceedings - 28. Workshop Computational Intelligence, Dortmund, 29. - 30. November 2018. Ed.: F. Hoffmann, 175, 2018.

[21] T.B. Olaf Ronneberger, Philips Fischer, "U-Net: Convolutional Networks for Biomedical Image Segmentation," IEEE Access, **9**, 16591–16603, 2015, doi:10.1109/ACCESS.2021.3053408.

[22] L. Dai, T. Li, H. Shu, L. Zhong, H. Shen, H. Zhu, "Automatic Brain Tumor Segmentation with Domain Adaptation," 1–12, 2018.

[23] L.M. Ballestar, V. Vilaplana, "MRI Brain Tumor Segmentation and Uncertainty Estimation Using 3D-UNet Architectures," Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **12658 LNCS**(1), 376–390, 2021, doi:10.1007/978-3-030-72084-1_34.

[24] D. Bahdanau, K.H. Cho, Y. Bengio, "Neural machine translation by jointly learning to align and translate," 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 1–15, 2015.

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No.5, 77-93 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Design and Prototyping of a 3DOF Worm-drive Robot Arm

Ian Spencer Howard*

SECAM, University of Plymouth, Plymouth, PL4 8AA UK

ARTICLE INFO

Article history:
Received:03 August, 2023
Accepted:28 October, 2023
Online: 30 October, 2023

Keywords:
Key 3DOF Revolute arm
Key Worm-drive actuation
Key 3D printing

ABSTRACT

Many designs for robot arms exist. Here we present an affordable revolute arm, capable of executing simple pick-and-place tasks. The arm employs a double parallelogram structure, which ensures its endpoint angle in the plane of the upper arm remains fixed without the need for additional actuation. Its limbs are fabricated from circular tubes made from bonded carbon fiber, to ensure low moving mass while maintaining high rigidity. All custom structural elements of the arm are produced via 3D printing. We employ worm-drive DC motor actuation to ensure that stationary configurations are maintained without the necessity of continuous motor power. Our discussion encompasses an analysis of the arm's kinematics. A simulation of the arm's operation was carried out in MATLAB, revealing key operational metrics. In conclusion, we achieved extrinsic endpoint position tracking by implementing its inverse kinematics and PID control using a microcontroller. We also demonstrate the arm's functionality through simple movement tracking and object manipulation tasks.

1. Introduction

1.1. Extension of previous work

This paper builds upon work on a prototype 2D planar arm first published in the proceedings of the 2022 International Conference on System Science and Engineering (ICSSE) [1]. We explicitly highlight the extensions of previous work and novel contributions of this study at the beginning of the discussion section.

1.2. Overview

Brought about by a combination of a dwindling agricultural workforce and advancements in robotics technology, the deployment of robots for harvesting fruits and vegetables has become an increasingly important area of development [2]. As a result, there is a growing demand for low-cost robots to not only reduce the cost of harvesting but also to address the shortages in agricultural workers willing to perform the task.

As a typical application area, we consider the development of robotic arms designed to harvest berries from plants. For such an application, robotic arms are normally attached to a computer-controlled mobile robot base. These systems typically operate in structured environments, such as greenhouses, and must autonomously identify, pick, and collect ripe berries. The design of suitable robotic arms involves trade-offs, as multiple requirements need to be met [3].

1.3. Requirements

To offer a realistic alternative to employing human workers, any useful arm design needs to perform its picking tasks effectively with no greater, and preferably lower, overall operating costs. Therefore, keeping the cost of arm construction low is an important design consideration.

In terms of performance, the workspace of the arm must be sufficient for it to reach the berries chosen for picking, and it must be able to do so with reasonable accuracy (of the order of millimeters). Ideally, it should accomplish this with minimal correction during movement, avoiding the need for visual servoing [4]. The arm must also generate sufficient force to hold and operate a suitable end effector mechanism to pick the berries. Pulling off a raspberry can require up to 10N [5], although alternative removal techniques, such as cutting, are also viable [6]. Force generation requirements place demands on the joint torques needed, as well as on the stiffness and strength of the arm structure. Clearly, a balance must be struck between arm link length and stiffness, given that arm stiffness decreases rapidly as a function of link length.

The actuation and control mechanisms of the arm should prioritize power efficiency, as mobile picking systems generally

*Corresponding Author: Ian Spencer Howard, SECAM, University of Plymouth. UK, ian.howard@plymouth.ac.uk

operate from batteries. To address limitations in the use of batteries, researchers are exploring the possibilities of harnessing solar energy to recharge the batteries when daylight conditions permit [7].

The arm must ensure safe operation in the proximity of people and guidelines for safe robot operations are specified by the ISO 10218 standard [8]. To do so, it must effectively manage occasional, inevitable collisions with items in its environment, which could include greenhouse plant support infrastructure or even curious human workers standing too close to the robot. Paramount attention must be paid to this issue, to prevent any injury or damage resulting from inadvertent collisions.

Several strategies can facilitate safe robotic arm operation. One approach involves building inherent compliance into their designs [9,10]. One such design, the Gummiarm, achieves this by making use of compliant non-linear tendons driven by Dynamixels [11,12]. However, incorporating passive compliance can often compromise performance.

Another way to enhance robot safety is to limit the arm's kinetic energy due to movement. This can be achieved by restricting the maximum operational speed and ensuring the arm has a lightweight structure, thereby reducing potential damage arising from a collision.

1.4. Collaborative robots

Much work has been carried out to make industrial robots safe around people [13–15]. A collaborative robot, (also known as a cobot), is a specialized robot, engineered to work in tandem with humans within a shared workspace.

Cobots are often equipped with an array of sensors, including cameras, force sensors, and proximity sensors, which enable them to sense and interact with their surroundings. These sensors allow cobots to identify human presence, avert collisions, and react to environmental changes. Furthermore, cobots incorporate safety features, such as force-limiting and speed-limiting controls, which contribute to accident prevention and the protection of human co-workers. A popular safety strategy employed in cobots, is to monitor their force generation [16]. If force levels exceed expected task values, the robot identifies this anomaly as a probable fault condition and safely deactivates the robot in a controlled manner.

Numerous companies, researchers, and engineers have embraced the cobot concept. Universal Robots, a Danish company, was among the first to introduce a cobot, and Rethink Robotics was also an early pioneer in the cobot domain. Cobots are gaining traction across a broad spectrum of industries, including manufacturing, logistics, and healthcare.

1.5. Lightweight arms

Several lightweight arms are commercially available. The Barrett WAM, popular among researchers, features low weight and employs a cable drive. Its actuation exhibits very low backlash and low friction, while the arm itself demonstrates a low effective mass [17].

The ANYpulator is another example of a lightweight robotic arm. It employs force control and utilizes lightweight carbon fiber links in its construction to minimize moving mass [18].

The DLR lightweight robot provides another example of high-quality arms capable of safe operation [15]. Kuka has commercialized its design, which also serves as the basis for the Franka Emika arm [19].

1.6. Existing low-cost robots

Most of the commercially available robots mentioned earlier are quite expensive. However, several low-cost robotic arm designs also exist. For instance, Reachy employs off-the-shelf components and fully 3D-printed limb sections [20].

The BioRob-Arm [21] is a lightweight design that utilizes compliant actuation. It is based on an antagonistic, series-elastic actuation that employs cable transmission [22].

Another compliant, 7-DOF (Degrees of Freedom) robotic arm was developed at Stanford University. To keep costs down, the design makes use of stepper motor drive and adopts series-elastic transmission in the lower arm and base. It adopts Dynamixels for actuation in the remaining three [23].

More recently, the Berkeley Blue 7-degree-of-freedom arm has been developed with the intention of making a low-cost, force-controlled robot available to the wider AI community [24].

2. Arm design

2.1. Design and construction overview

In this section, we focus on designing a lightweight arm driven by worm-gear motors, via HTD 5M x 15mm wide timing belts, enabling it to maintain static configurations without requiring constant motor power. The arm design is based on a revolute configuration. See Figure 1 for a photograph of the arm, Figure 2 for a close-up of the pulley actuation, and Figure 3 for labelled schematics indicating important parts of the robot design.

The design of the arm was carried out using Autodesk Fusion 360. To ensure minimal weight, the upper and lower parallelogram arm sections employed carbon fiber tubing, with dimensions of 16x14mm in diameter and 250mm in length.

All parts that were custom-designed were initially exported in the STL file format, subsequently sliced employing Ultimaker Cura software, and 3D printed utilizing PLA+ material with a Creality CR-6 SE printer. To optimize stiffness and strength, arm joint components were printed with a 100% infill [25]. Additional components were printed with a 35% infill to strategically minimize their moving mass. Keeping the weight of the components low was particularly important for the upper arm mechanism support pillars, as they rotate during the operation of the arm.

Loctite Precision superglue was used to attach the carbon fiber tubes to their corresponding 3D printed PLA+ parts.

2.2. Rotary base

The first arm joint is achieved using a platform that rotates around a vertical axis. This platform supports a planar arm mechanism, which provides an additional two degrees of freedom.

The top of the rotary base assembly is fabricated from plate sections. These are both 20mm-thick and are 3D-printed. They are supported by two side pillars resting on a single, similarly thick

3D-printed base plate. These components are secured together using four long bolts. The top two base plates tightly secure a deep groove ball bearing in place. It has a large internal bore of 50mm to accommodate a 3D-printed shaft, which passes through the bearing. The shaft is affixed to the bearing assembly, using a 3D-printed clamping collar. This provides the shaft with radial and axial support. The collar incorporates an integrated 40-tooth HTD 5M x 15mm timing pulley. The shaft represents the arm's first degree of freedom. The upper end of the main shaft features a flat plate with mounting holes, so the upper arm assembly can be bolted onto it.

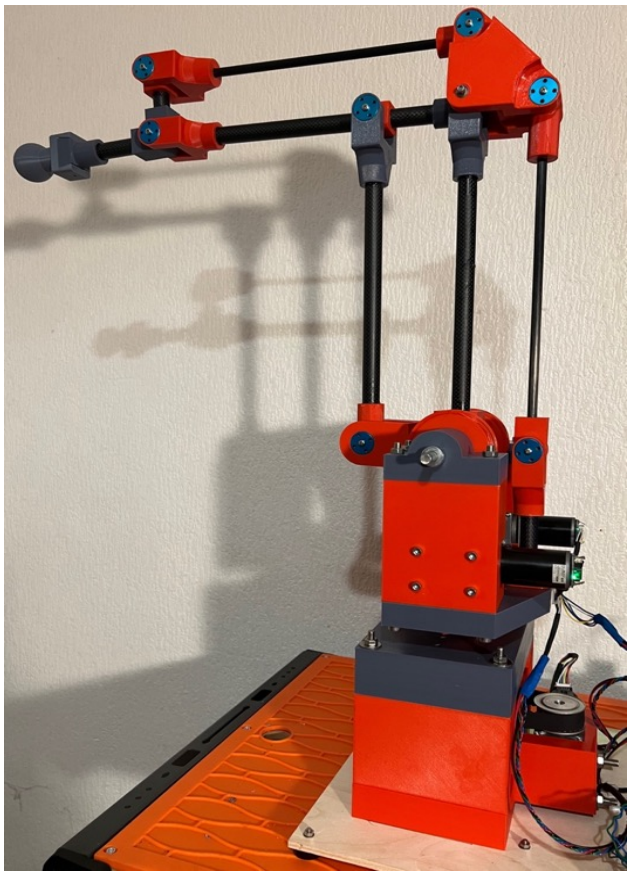


Figure 1: 3 DOF worm-gear motor robotic arm. The base mechanism serves as the initial rotational joint of the arm. Incorporating two degrees of freedom (2DOF), the upper segment adopts a parallelogram design.

The main shaft timing pulley on the collar is driven by a timing belt. A worm-gear motor, fitted with a smaller aluminum 24-tooth HTD 5M x 15mm timing pulley, drives the other end of the timing belt in order to rotate the shaft (Figure 2B). The motor itself is located on the lower base plate.

2.3. Two-degrees of freedom parallelogram design

The upper part of the arm mechanism is mounted on the main axis and is rotated by the base assembly, as depicted in Figure 3. Its design incorporates a mechanism to maintain the orientation of the endpoint in the plane of the upper arm, eliminating the need to control this linkage explicitly.

The main upper arm link L2 is supported and driven by a 3D-printed component that incorporates a 40-tooth HTD 5M x 15mm wide timing pulley. The component incorporates 2 x 8mm ball bearing races and is located on the main 8mm-diameter parallelogram mechanism shaft.

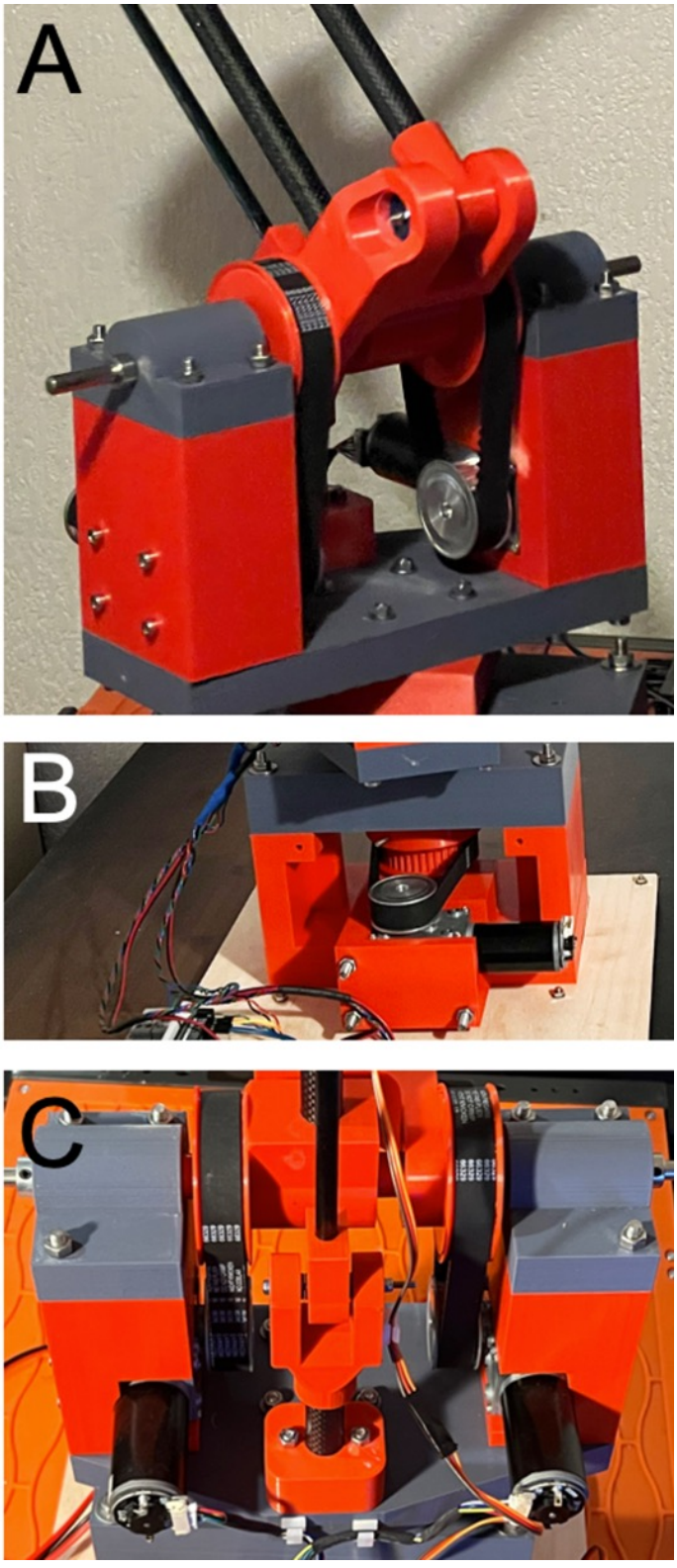


Figure 2. Close-up view of the rotating base platform mechanism. Panel A: Front view of the pulleys drive connected to the parallelogram arm mechanism. Panel B: The horizontal pulley drive for J1 for the first rotary axis of rotation. Panel C: Rear view of the parallelogram pulley drive showing worm-gear motors.

The secondary upper arm link arm is supported and driven by a 3D-printed mini-arm component, which also incorporates a 40-

tooth HTD 5M x 15mm wide timing pulley. This component also incorporates 2 x 8mm ball bearing races and is likewise located on the main 8mm-diameter parallelogram mechanism shaft.

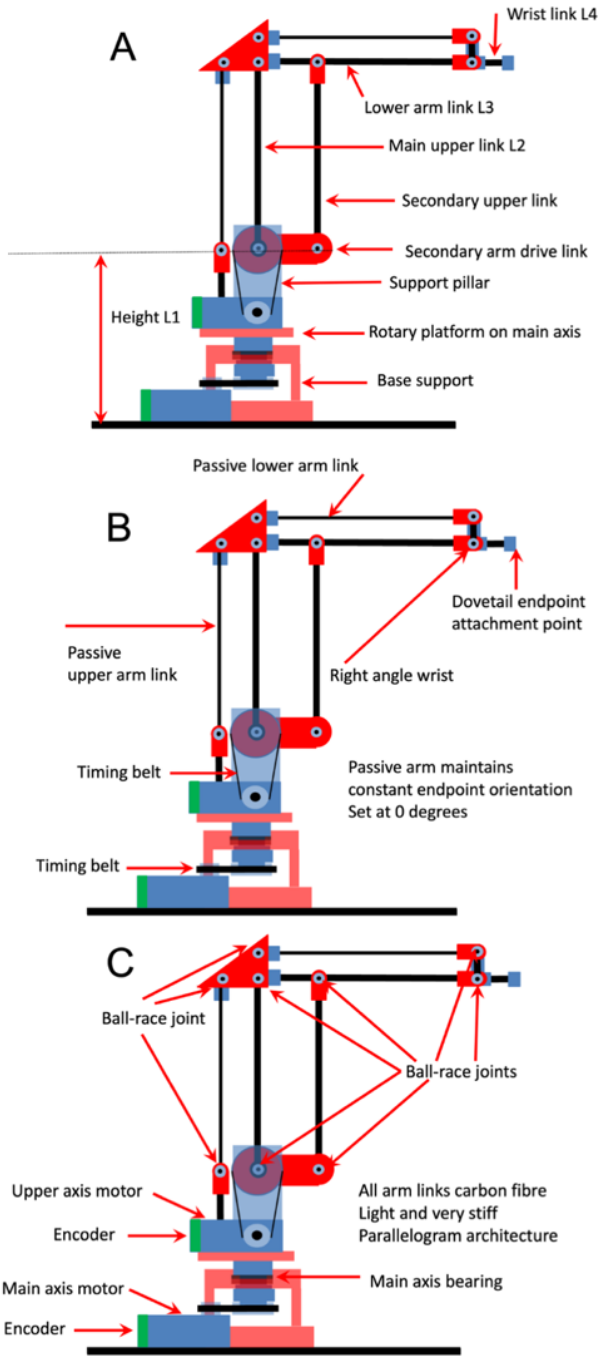


Figure 3: Robot viewed from the side. Panel A: Depicts the main components of the base and upper parallel arm links. Panel B: Shows the passive arm components and drive belts for the main arm. Panel C: Highlights the locations of the bearings, motors, and encoders.

The lower link L3 is driven by the main arm link L2 and secondary link L2 via custom made 3D-printed joint components (see Figure 3A).

To maintain the consistent orientation of the final arm link (wrist link L4), and consequently the endpoint, within the plane of the parallel arm mechanism, a passive orientation system is strategically situated behind the main arm. This ensures stable positioning and alignment throughout the operational movements of the arm mechanism. This additional parallelogram structure makes use of solid 8mm diameter carbon fiber rods in its construction. Figure 4A illustrates the endpoint alignment rocker mechanism, and the endpoint alignment mechanism and end-effector attachment point are shown in Figure 4B. This configuration enables the planar arm mechanism to preserve the endpoint orientation angle within the plane of the parallel arm without necessitating additional explicit actuation. As a result, the upper arm demands only two actuators for control, instead of three.

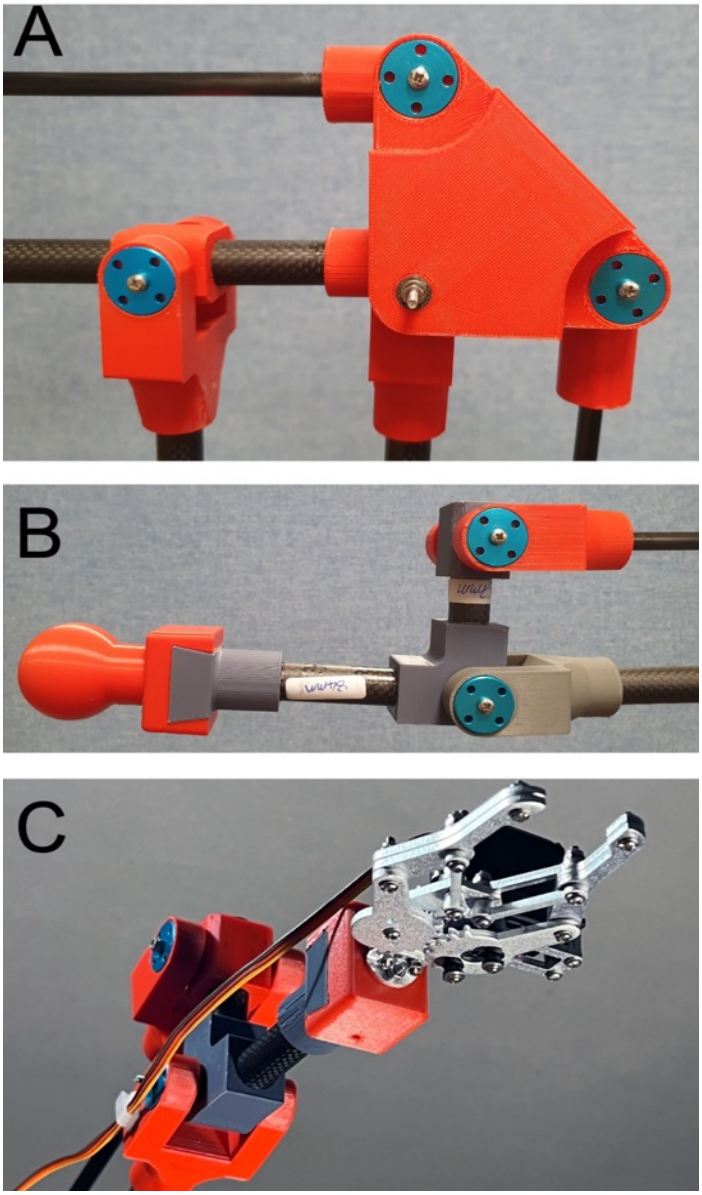


Figure 4: Panel A: The 2D upper arm features an endpoint alignment joint component located at the elbow. Panel B: The endpoint orientation mechanism. A dovetailed profile at the lower end of the last limb segment provides a connection point for end effectors. Here, a round knob is attached for safe demonstration purposes. Panel C: Simple commercially available RC-servo gripper mechanism attached to arm to demonstrate pick-and-place operations.

2.4. Arm joint components

The arm design relies on the use of joints, which utilize miniature ball-races to minimize friction. Figure 5 shows arm joint

components and their assembly is depicted in Figure 6. Each fork utilizes two lightweight aluminum flanges to support a shaft passing through the bearings, making use of the enhanced mechanical properties of aluminum compared to plastic. A long M3 screw is used as the joint shaft and is held in its support fork by means of a nylon insert hex locknut. Washers are strategically positioned on the screw shaft to achieve joint clearance between the joint bearing and the aluminum flanges.

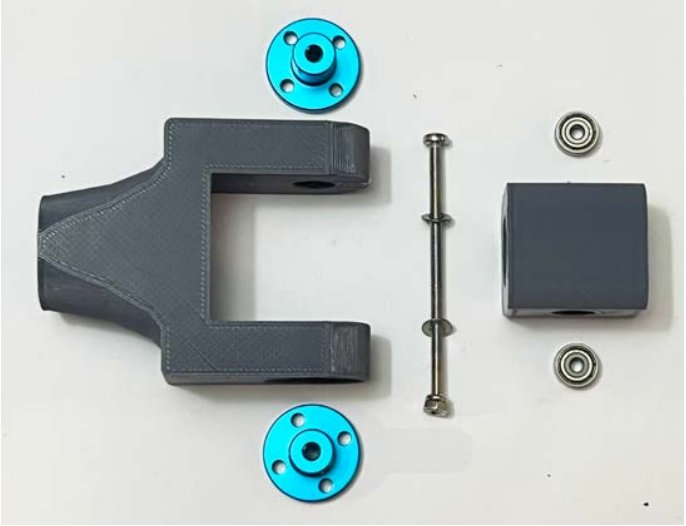


Figure 5: Components of the arm joints. The bearing component on the right-hand side hold miniature ball bearing races with a 3mm internal diameter. An M3 bolt is used as the joint shaft. When assembled, the joint bearing component locates between the two-sided support, illustrated on the left. This employs aluminum guides to hold the shaft. Washers are used to obtain suitable joint clearance.

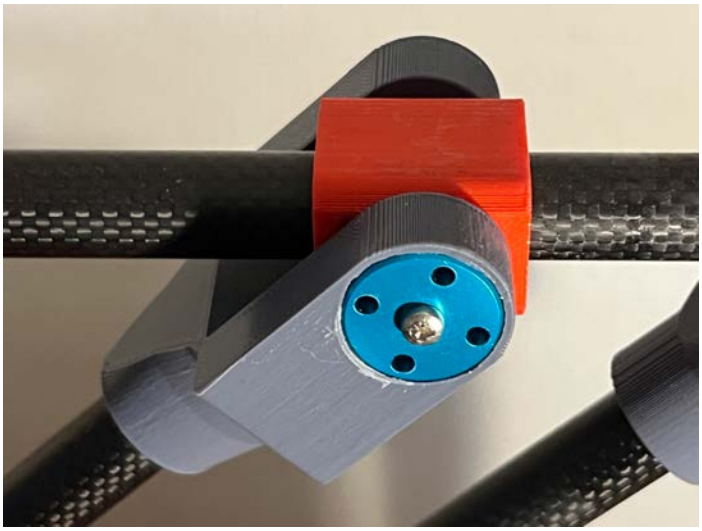


Figure 6: 2D Arm joint mechanism. Fundamental to the arm mechanism design, the arm joints utilize small ball bearings to ensure high precision and low friction operation.

2.5. Actuation using worm-drive motors

We utilize worm-gear motors for the arm's actuation. Opting for brushed worm-drive DC motors proves advantageous in constructing a robust servo drive system due to their wide availability, cost-effectiveness, and capacity to deliver high torque. Furthermore, the inherent non-back drivable characteristic of these drives guarantees the maintenance of a static joint configuration without necessitating continuous motor power [26].

The selected worm-gear motors run from a 24V supply, drawing up to 1A under full load. The worm gearing achieves a rotary speed reduction of 280, and the motor units achieve an unloaded output shaft speed of 28 RPM and an operational torque of 5Nm. These motors are equipped with Hall sensor encoders, located at the back, which deliver 11 pulses per revolution. This configuration allows the motor output shaft angle to be estimated with high accuracy and an output angular resolution of 3080 pulses per revolution was achieved. This figure can also be quadrupled further by utilizing the pulse edges.

As previously described, the motor output shafts are connected to the arm joints' drive axes utilizing 24-tooth HTD 5M x 15mm timing pulleys. The maximum angular speed of each driven joint, when using the selected worm-gear motors, is dictated by both the maximum motor rotational speed, converted from RPM to radians per second, and the mechanical advantage offered by the pulley drive. Consequently, given that all robot axis joints employ identical pulley ratios, the following value is obtained:

$$\omega_{max} = 28 * \left(\frac{24}{40}\right) * \left(\frac{2\pi}{60}\right) = 1.76\ Rads^{-1} \quad (1)$$

The maximum joint torque that can be achieved is given by the maximum continuous torque motor scaled up by the mechanical advantage of the pulley drive:

$$\tau_{max} = \left(\frac{40}{24}\right) * 5 = 8.33\ \text{Nm} \quad (2)$$

The robot arm incorporates three worm-drive motors in total. The main vertical revolute joint axis assembly, which integrates a 3D-printed timing pulley, is actuated by the first worm-gear-driven timing belt assembly. Moreover, two additional worm-drive motors mount within the pillars of the rotating vertical axis platform. The second joint motor engages a pulley to drive the main upper arm parallelogram arm link, while the third engages a timing pulley within the secondary arm drive link. Figure 2 illustrates the worm-drive motors, pulleys, and timing belts on the physical robot.

3. Gripper end effector for the arm

3.1. Gripper designs

Grippers are crucial end-effectors for robotic arms, enabling them to execute a myriad of tasks. Numerous designs have been proposed [27,28]. They are also an integral component of robotic harvesting technologies [29,30]. In addition to gripper designs that utilize hard materials, there is a significant contemporary interest in using soft materials for gripper construction [31,32].

3.2. RC servo operated gripper

Although pneumatic operation can offer a high-power density and a clean method of actuation for grippers, we chose to use RC servo actuation for easier integration into the current system. To demonstrate straightforward pick-and-place procedures with the arm, we affixed a low-cost, commercially available aluminum alloy RC-servo gripper mechanism to the arm via its dovetail connection point (refer to Figure 4C). This lightweight gripper assembly, weighing 160g, is capable of supporting simple pick-and-place tasks under direct human operator control.

We also designed and built a prototype RC servo-driven gripper with soft Ninja Flex cups at the ends of its fingers, specifically for assisting in gripping berries (refer to Figure 7). This gripper is lightweight, weighing 202g, and is attached to the arm through its dovetail connection point. We utilized this gripper to showcase the robotic arm's ability to detach berries from an artificial raspberry bush [33]. Refer to the results section for links to YouTube videos demonstrating the operation of both grippers.

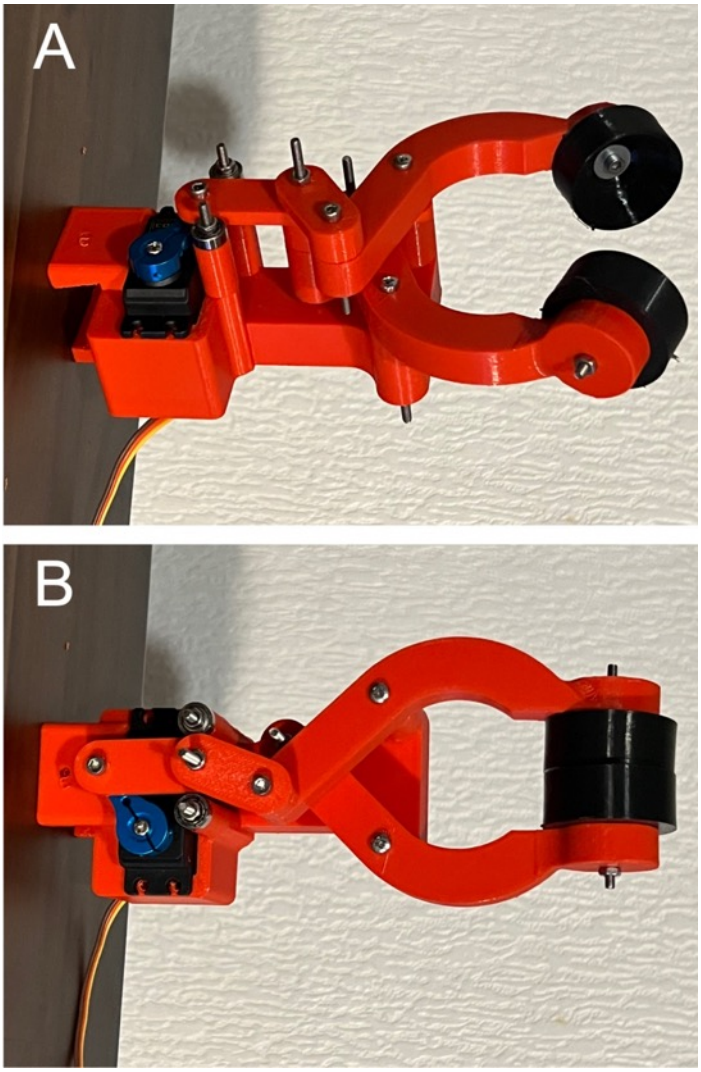


Figure 7. PLA+ 3D printed RC servo actuated gripper with soft Ninja flex endpoint cups. The gripper is shown in open (panel A) and closed (panel B) configurations.

4. Forward kinematics

4.1. Kinematic analysis of the arm

In robotic arms that operate with actuators capable of torque control, the examination of arm dynamics is an important and valuable exercise. In our design, we utilize worm-drive motors. Due to the high levels of friction introduced by this type of drive mechanism, which essentially results in a non-back-drivable system, worm drive motors are not conducive to building systems that effectively implement torque control. As a result, in our design, we do not employ torque control for the motors. Instead, we directly implement kinematic positional control of joint angles. For this reason, our analysis focuses on the forward and inverse kinematics of the mechanism as well as the Jacobian-based relationships between motor rotational velocity and endpoint velocity. Additionally, since the arm is lightweight and operates at a relatively slow speed, we also present an analysis of static force relationships using the Jacobian transpose.

4.2. Denavit-Hartenberg

Numerous textbooks on robotics offer kinematic analyses of arm structures [34–36]. For reader convenience, here we provide straightforward derivations.

Figure 8 illustrates the structure of the four links within the arm. While a straightforward derivation of the kinematics from the arm's geometry is possible, we have chosen to employ classical Denavit-Hartenberg (DH) analysis, a method often adopted in the field of robotics. Figure 9 displays the appropriate frames for DH analysis. The corresponding DH table for the arm are presented in Table 1 and the operating range limits of the robot joints are also included.

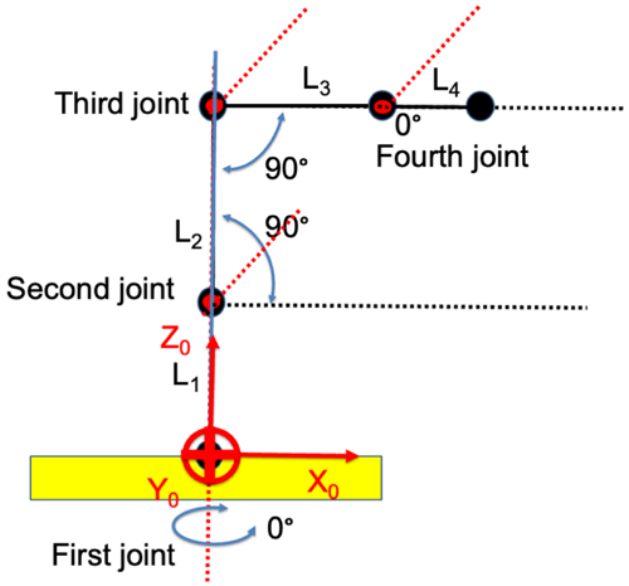


Figure 8: Default configuration.

Table 1: Denavit-Hartenberg parameters for the arm mechanism

Link	Angle θ [°]	Min θ [°]	Max θ [°]	Angle α [°]	Radius a [mm]	Offset d [mm]
1	θ_1	-45°	45°	0	0	300
2	θ_2	35°	145°	0	300	0
3	θ_3	-145°	-35°	0	300	0
4	θ_4	-110°	110°	0	175	0

4.3. Forward kinematics

The homogeneous transformation matrix for a classical DH-frame is given by the expression:

$$ {}^{i-1}_{i}H = \begin{bmatrix} c\theta_i & -s\theta_i c\propto_i & s\theta_i s\propto_i & a_i c\theta_i \\ s\theta_i & c\theta_i c\propto_i & -c\theta_i s\propto_i & a_i s\theta_i \\ 0 & s\propto_i & c\propto_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3) $$

where *'i'* is the frame, 'c' denotes cosine and 's' sine. We note that here we use the suffix notation for a homogeneous transformation H between link frames, indicating mapping *from the frame* identified by the i-suffix location *to the frame* identified in the i-1 suffix location, that is ${}^{to}_{from}H$.

In our case, the overall forward kinematic transformation for the arm is accomplished by multiplying the homogeneous matrices for the links together in the appropriate serial order, leading to the expression:

$$ {}^{0}_{4}H = {}^{0}_{1}H\ {}^{1}_{2}H\ {}^{2}_{3}H\ {}^{3}_{4}H \tag{4}$$

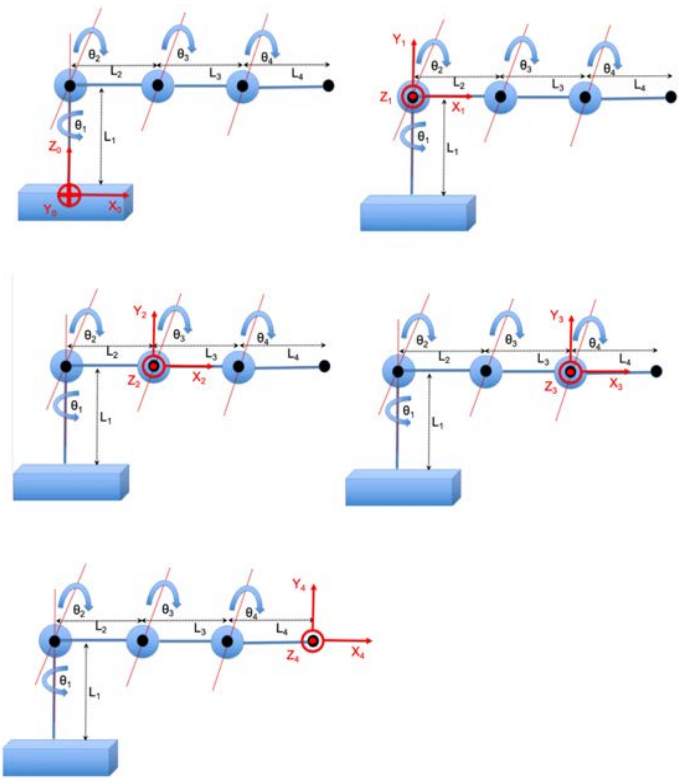


Figure 9. DH coordinate frames.

We utilize the MATLAB symbolic toolbox to evaluate the expression given in Eqn. (4), leading to the overall homogeneous transformation:

$$ {}^{0}_{4}H = \begin{bmatrix} c_{234}c_1 & -s_{234}c_1 & s_1 & c_1(L_3c_{23} + L_2c_2 + L_4c_{234}) \\ c_{234}s_1 & s_{234}s_1 & -c_1 & s_1(L_3c_{23} + L_2c_2 + L_4c_{234}) \\ s_{234} & c_{234} & 0 & L_1 + L_3s_{23} + L_2s_2 + L_4s_{234} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

In this context, L_1, L_2, L_3, and L_4 denote the four link lengths, and we introduce the following notational simplifications for clarity: $c_1 = \cos(\theta_1)$, $s_1 = \sin(\theta_1)$, $c_{23} = \cos(\theta_2 + \theta_3)$, $s_{23} = \sin(\theta_2 + \theta_3)$, $c_{234} = \cos(\theta_2 + \theta_3 + \theta_4)$, $s_{234} = \sin(\theta_2 + \theta_3 + \theta_4)$.

In our arm design, the angle θ_4 is dependent on the final link orientation angle θ_{end} as well as joint angles θ_2 and θ_3. That is:

$$ \theta_{\text{end}} = \theta_2 + \theta_3 + \theta_4 \tag{6}$$

The expression below just provides the endpoint location of the final link, together with its orientation angle θ_{end}:

$$ \begin{bmatrix} x_{\text{end}} \\ y_{\text{end}} \\ z_{\text{end}} \\ \theta_{\text{end}} \end{bmatrix} = \begin{bmatrix} c_1(L_3c_{23} + L_2c_2 + L_4c_{\text{end}}) \\ s_1(L_3c_{23} + L_2c_2 + L_4c_{\text{end}}) \\ L_1 + L_3s_{23} + L_2s_2 + L_4s_{\text{end}} \\ \theta_2 + \theta_3 + \theta_4 \end{bmatrix} \tag{7}$$

We define a neutral arm posture, as illustrated in Figure 8. Configuration of the arm can be specified by control angles that deviate from this posture. Clearly, in this state, these angles Φ_1, Φ_2, and Φ_3 have values of zero. By inspection of Figure 8, it can be seen that Φ_1 follows θ_1 directly, whereas Φ_2 and Φ_3 are offset from θ_2 and θ_3. More precisely, we calculate the control angles as follows:

$$ \Phi_1 = \theta_1 \tag{8}$$

$$ \Phi_2 = \theta_2 - 90° \tag{9}$$

$$ \Phi_3 = \theta_3 + 90° \tag{10}$$

4.4. Workspace of the arm

The robotic arm's workspace is confined not only by the lengths of its links but also by the operational range of the angular joints, which are restricted due to inherent physical limits.

The primary vertical joint of the robot arm, denoted as J1, is theoretically capable of rotating from -180° to 180°. However, due to practical constraints arising from the wiring of the upper motors and encoders we chose to limit its rotation to a range between -45° and 45°. Joint J2, associated with the main upper arm link, is intrinsically able to rotate across a range from 0° to 180°. Nevertheless, physical limitations within the end-point orientation mechanism confine its rotational scope to between 35° and 145°.

Similarly, the rotation of joint J3, linked to the secondary arm drive, is constrained by the endpoint orientation mechanism, permitting rotation only within an approximate range of -35° to -145°.

To illustrate the arm's workspace, we first generate a test dataset of random configurations of the arm. To achieve this, for each arm joint angle, we independently draw 20,000 samples from a uniform distribution with lower and upper angular limits set by the operating range of the respective joint. We then use these 20,000 random angular configurations to calculate the arm end-point positions using forward kinematics, as given by Eqn. (7). The resulting arm end-points are then plotted in 3D and displayed in three different views for clarity, as illustrated in Figure 10.

5. Differential kinematic analysis

5.1. Exploring joint and endpoint velocity relationships

The 4x4 Jacobian matrix of the arm can be used to relate its end-point velocity to joint velocity. To formulate the Jacobian requires that we find the partial derivatives of its end point position and orientation given in Eqn. (7), with respect to the arm joint angles. The Jacobian is given by the matrix expression:

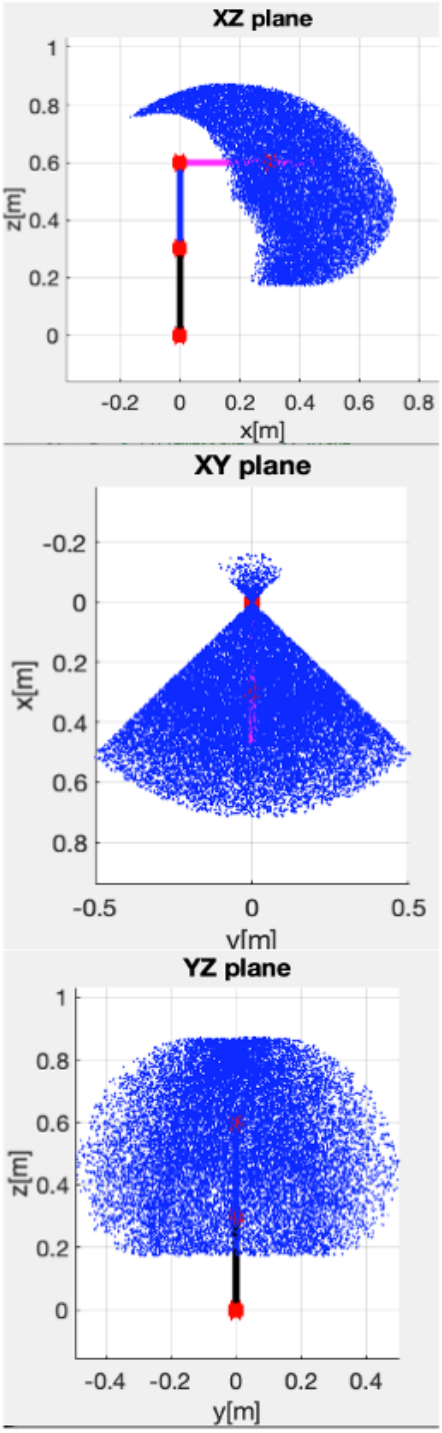


Figure 10. Workspace of the robot arm, also showing the arm displayed in its standard canonical configuration. Panel A: Side view showing the XZ axes. Panel B: Top view showing the XY axes. Panel C: Front view showing the YZ axes.

$$J = \begin{bmatrix} \frac{\partial x}{\partial \theta_1} & \frac{\partial x}{\partial \theta_2} & \frac{\partial x}{\partial \theta_3} & \frac{\partial x}{\partial \theta_4} \\ \frac{\partial y}{\partial \theta_1} & \frac{\partial y}{\partial \theta_2} & \frac{\partial y}{\partial \theta_3} & \frac{\partial y}{\partial \theta_4} \\ \frac{\partial z}{\partial \theta_1} & \frac{\partial z}{\partial \theta_2} & \frac{\partial z}{\partial \theta_3} & \frac{\partial z}{\partial \theta_4} \\ \frac{\partial \theta_{\text{end}}}{\partial \theta_1} & \frac{\partial \theta_{\text{end}}}{\partial \theta_2} & \frac{\partial \theta_{\text{end}}}{\partial \theta_3} & \frac{\partial \theta_{\text{end}}}{\partial \theta_4} \end{bmatrix} \quad (11)$$

This leads to the expression:

$$\begin{bmatrix} -s_1(L_3c_{23} + L_2c_2 + L_4c_{234}) & -c_1(L_3s_{23} + L_2s_2 + L_4s_{234}) & -c_1(L_3s_{23} + L_4s_{234}) & -c_1L_4s_{234} \\ c_1(L_3c_{23} + L_2c_2 + L_4c_{234}) & -s_1(L_3s_{23} + L_2s_2 + L_4s_{234}) & -s_1(L_3s_{23} + L_4s_{234}) & -s_1L_4s_{234} \\ 0 & L_3c_{23} + L_2c_2 + L_4c_{234} & L_3c_{23} + L_4c_{234} & L_4c_{234} \\ 0 & 1 & 1 & 1 \end{bmatrix} \quad (12)$$

The Jacobian relates the rotational velocities at the robot's joints to extrinsic velocity and rotational velocity of the endpoint through the simple expression:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \\ \dot{\theta}_{end} \end{bmatrix} = J \begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \\ \dot{\theta}_3 \\ \dot{\theta}_4 \end{bmatrix} \quad (13)$$

Examination of the Jacobian is informative regarding the singularities of the arm mechanism. Here we consider the case when the base rotation is zero. In this case the determinant of the Jacobian can be expressed as the relationship:

$$det(J) = -L_2L_3s_3(L_4 + L_3c_{23} + L_2c_2) \quad (14)$$

We observe that the determinant reaches zero when θ_3 is either 0 or π, occurring when the third link aligns with, or rotates backwards onto the second arm link. However, such configurations are not within the operating range of the arm. Consequently, on the basis of Eqn. (13), we can employ the inverse Jacobian to compute the joint velocities as follows:

$$\begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \\ \dot{\theta}_3 \\ \dot{\theta}_4 \end{bmatrix} = J^{-1} \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \\ \dot{\theta}_{end} \end{bmatrix} \quad (15)$$

5.2. Static force relationships

We can construct the 3×4 Jacobian by omitting the last row in Eqn. (12). Utilizing the transpose of the 3×4 Jacobian, we can establish a relationship between static joint torques - denoted $[\tau_1 \quad \tau_2 \quad \tau_3 \quad \tau_4]^T$ - and static endpoint forces - expressed as $[f_x \quad f_y \quad f_z]^T$ - through the following relationships:

$$\begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{bmatrix} = J^T \begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix} \quad (16)$$

6. Inverse kinematics

6.1. Main rotary axis

To determine the arm's inverse kinematics, we first define the endpoint location in the base frame coordinate system as the point (x_{end}, y_{end}, z_{end}). We note that in the base frame coordinate system, the arm mechanism's first joint constitutes a rotational axis around its z-axis, specified the angle θ_1. This rotation is the sole contributor to the end-effector's displacement along the base frame coordinate system's y-axis, as illustrated in Figure 11. Thus, determining θ_1 is relatively straightforward, given that:

$$\theta_1 = \text{atan}\left[\frac{y_{end}}{x_{end}}\right] \quad (17)$$

6.2. Planar arm mechanism

Our attention now shifts to the remaining planar arm mechanism. We define a 2D coordinate frame in the plane of the planar arm, situating its base at the origin (0,0), and defining the end effector location by the 2D point (x_{p3}, y_{p3}), as depicted in Figure 12.

Referring once more to Figure 12, we observe that the x-displacement, denoted as x_{p3}, is given by:

$$x_{p3} = \sqrt{x_{end}^2 + y_{end}^2} \quad (18)$$

We note that the direction of y-axis in the planar arm frame corresponds to the direction of z-axis in base frame. However, the 2D location of the end effector, denoted as y_3, must take the base height into account: Therefore, we have:

$$y_{p3} = z_{end} - L_1 \quad (19)$$

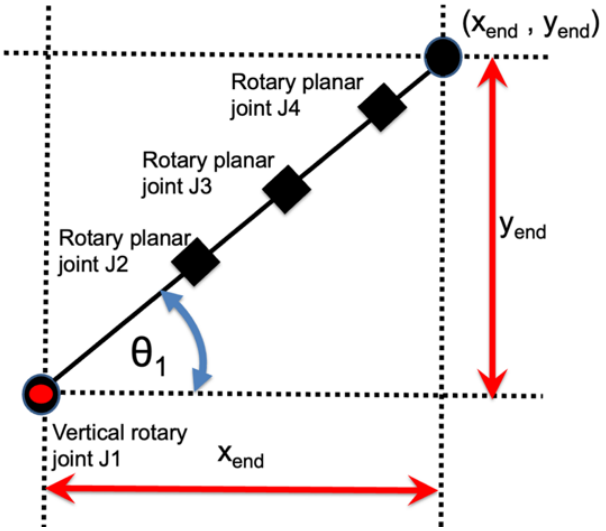


Figure 11: Top-down view of the robot arm: This perspective demonstrates the simplicity in calculating the first joint angle θ_1.

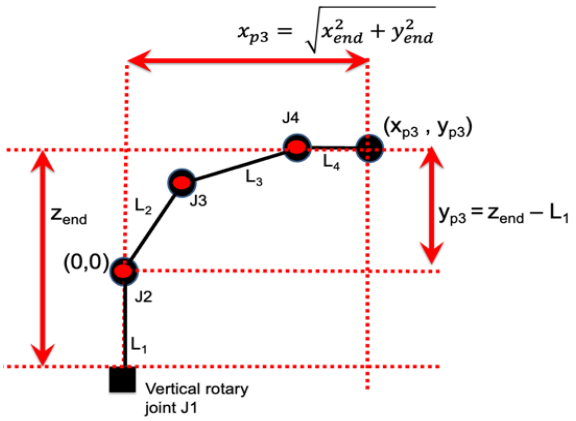


Figure 12: Side view of arm when main vertical axis joint angle is zero. The end-point of the arm in this coordinate frame is specified as the point (x_p, y_p), and its origin (0, 0) is located at J2.

We will shortly use the point (x_{p3}, y_{p3}) to calculate the 2D point (x_{p2}, y_{p2}), which we now consider obtaining the angles θ_2 and θ_3 of the remaining 2-link planar arm assembly. The cosine rule can be used to analytically derive this structure's inverse kinematics.

As depicted in Figure 13A for the 2-link planar arm mechanism, the elbow joint is situated at point (x_{p1} y_{p1}), while the upper arm joint is located at (x_{p2}, y_{p2}). Utilizing the Pythagorean theorem yields an expression for the length of the hypotenuse r:

$$r^2 = x_{p2}^2 + y_{p2}^2 \quad (20)$$

Employing the cosine rule, we see that:

$$r^2 = L_2^2 + L_3^2 - 2L_2L_3\cos(\alpha) \quad (21)$$

$$\Rightarrow \cos(\alpha) = \frac{L_2^2 + L_3^2 - x_{p2}^2 - y_{p2}^2}{2L_2 L_3} \quad (22)$$

From Figure 13A, it is noted that:

$$\theta_3 = \pi - \alpha \quad (23)$$

And from the trigonometric relationship, we have:

$$\cos(\pi - \alpha) = -\cos(\alpha) \quad (24)$$

$$\Rightarrow \cos(\theta_3) = \frac{x_{p2}^2 + y_{p2}^2 - L_2^2 - L_3^2}{2L_2 L_3} \quad (25)$$

From Figure 13B, we see:

$$\tan(\beta) = \frac{L_3\sin(\theta_3)}{L_3\cos(\theta_3) + L_2} \quad (26)$$

$$\Rightarrow \beta = \text{atan}\left[\frac{L_3\sin(\theta_3)}{L_3\cos(\theta_2) + L_2}\right] \quad (27)$$

From Figure 13C, we see that:

$$\theta_2 = \gamma - \beta \quad (28)$$

And that:

$$\gamma = \text{atan}\left[\frac{y_{p2}}{x_{p2}}\right] \quad (29)$$

Therefore, we have:

$$\theta_2 = \text{atan}\left[\frac{y_{p2}}{x_{p2}}\right] - \text{atan}\left[\frac{L_3\sin(\theta_3)}{L_3\cos(\theta_3) + L_2}\right] \quad (30)$$

To obtain a direct expression for the inverse kinematics of the arm assembly in terms of the point (x_{p3} y_{p3}), we first consider the constant orientation of the end link. In order to account for its effect, which is only a translation along the x-axis and none along the y-axis, we subtract its length from the end effector's position x_{p3} and directly use its y value:

$$\begin{bmatrix} x_{p2} \\ y_{p2} \end{bmatrix} = \begin{bmatrix} x_{p3} - L_4 \\ y_{p3} \end{bmatrix} \quad (31)$$

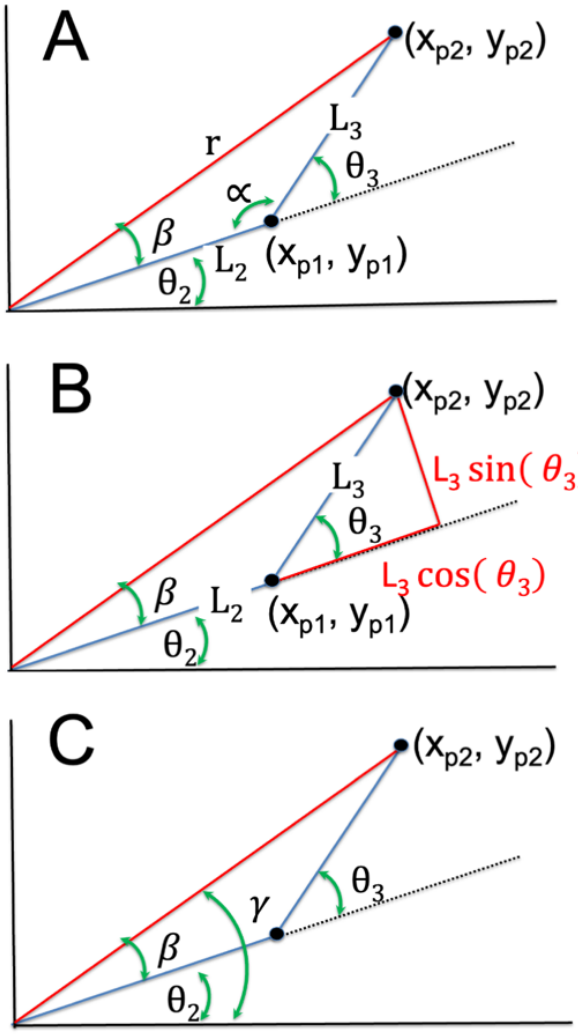


Figure 13: Schematic representation of the upper 2D parallelogram mechanism in the revolute arm: Useful for deriving inverse kinematics, the diagram illustrates the key positions, angles, and lengths.

Finally, we need to substitute the end effector location in as specified in the base coordinate frame, as defined by (x_{end}, y_{end}, z_{end}), and account for the height of the first vertical axis:

$$\begin{bmatrix} x_{p2} \\ y_{p2} \end{bmatrix} = \begin{bmatrix} \sqrt{x_{end}^2 + y_{end}^2} - L_4 \\ z_{end} - L_1 \end{bmatrix} \quad (32)$$

We substitute these values for (x_{p2}, y_{p2}) into Eqn. (30)

$$\Rightarrow \theta_2 = \operatorname{atan}\left[\frac{y_{end} - L_1}{\sqrt{x_{end}^2 + y_{end}^2} - L_4}\right] - \operatorname{atan}\left[\frac{L_3 \sin(\theta_3)}{L_3 \cos(\theta_3) + L_2}\right] \quad (33)$$

and into Eqn. (25)

$$\Rightarrow \theta_3 = \pm \operatorname{acos}\left[\frac{\left(\sqrt{x_{end}^2 + y_{end}^2} - L_4\right)^2 + (z_{end} - L_1)^2 - L_2^2 - L_3^2}{2L_2 L_3}\right] \quad (34)$$

We observe that Eqn. (34) yields two solutions and employs the value of θ_3 to compute θ_2 utilizing Eqn. (33). We opt for the negative solution for θ_3 as it aligns with the configuration illustrated in Figure 2. Lastly, we determine θ_4 using Eqn. (6):

$$\theta_4 = \theta_{end} - \theta_2 - \theta_3 \quad (35)$$

We note that this relationship was also used to calculate the minimum and maximum values of θ_4 shown in Table 1 on the basis of the minimum and maximum values of θ_2 and θ_3.

7. Control electronics

7.1. Microcontroller operation

A microcontroller system, built upon the Arduino Mega platform, was designed to operate the arm and power the worm-drive motors. This microcontroller offers an ample number of digital I/O ports with interrupt capability, facilitating operation with three incremental encoders located on the back of the motors through interrupt pins. This setup allowed for effective determination of the motor positions before the gear reduction via their worm drive mechanisms. To control the robot, a USB serial connection with the Arduino Mega was used.

7.2. Motor speed regulation using H-Bridges

To operate the DC motors used in the arm, we utilized two L298N Dual H-Bridges. These components allowed precise control over both the speed and direction of all three motors. Motor voltage is set using pulse-width modulation (PWM) activation of the on-time of the H-bridges. Drive power is provided from a 24V regulated power supply.

7.3. DIN rail controller implementation

All controller components were mounted on DIN rails, providing a simple and robust construction method. Please refer to Figure 14 to see its physical implementation. The 5V supply required to power the H-Bridge was provided by the Arduino Mega. In addition, an LM2596S DC-DC Buck converter module was attached to the DIN rail, directly driven from the 24V regulated power supply, to operate accessories such as the RC servo-operated gripper mechanism.

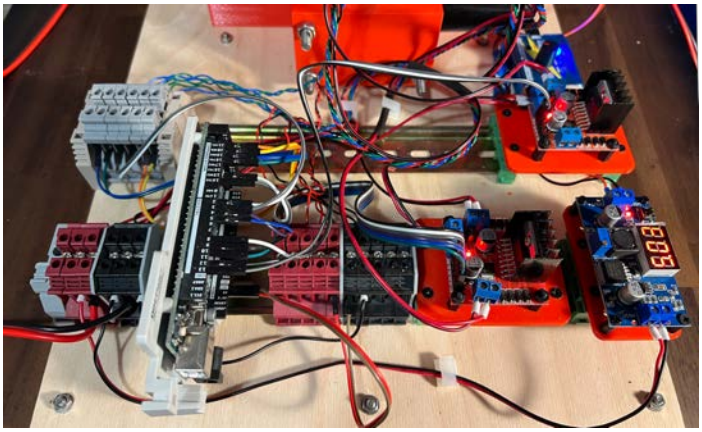


Figure 14: Controller electronics organized using DIN rail construction. The Arduino-based controller, terminal blocks, and H-bridges for the three worm drive motors can be seen mounted on the two DIN rails. A Buck controller is also shown and provides a 5V drive for the RC servo operated gripper mechanism.

7.4. Microcontroller program structure

Code was developed on the Arduino Mega to implement the kinematic control of the arm. The operation of the code consists of three distinct stages:

1. Initialization and Definitions: In this stage, variables and functions utilized by the algorithm are defined and initialized. Class objects are also constructed.
2. Setup: In this phase, operating variables are initialized and communications with the host PC are established.
3. Poll Loop: The poll loop is an endless processing loop that is used to service the program's needs. As well as implemented arm control, it must also service incoming messages from the host PC via a USB connection.

7.5. Microcontroller program menu

Within the poll loop, incoming commands from the serial interface are read and decoded. The available commands include:

- Manually set positive and negative rotation of individual drive motors.
- Reset encoders. This action needs to be carried out after the arm is positioned into its standard configuration, leading to calibration of the joint angles.
- Activate movement tracking along a predetermined trajectory.
- Activate point mode so can manually increment or decrement the robot's end-point along either the x, y or z axis in extrinsic space.
- Stop all movement.
- Display help menu and view parameters.

7.6. PID controller to track target trajectories

The arm is preliminary intended to carry out point-to-point movements for pick-and-place operations. For such tasks, the arm is typically given positional targets, or trajectories, to which it is required to reach. Inverse kinematics are then used to calculate the corresponding target motor angles. To accomplish this trajectory tracking task, a position control mode was implemented. When this mode is active, motor encoders are read to estimate the arm's joint angles, taking into considering the mechanical advantage of the worm gearing and also the drive pulley system. PID control is applied to ensure they match up with the requested joint control angles. In total there are three separate PID controllers, one for each motor. Output from the PID controllers constitutes velocity commands which make use of the H-Bridges to drive the motors.

Although the parallelogram joints J_1 and J_2 can be directly driven by the arm control angles Φ_1, Φ_2 , the third parallelogram joint J_3 must be driven by the sum of Φ_2 and Φ_3 to achieve the desired configuration of the parallelogram.

The Proportional-Integral-Derivative (PID) gains were determined empirically through experimental trials conducted on the actual robot. Notably, the integral gain was set to zero. In the next section we present the pseudo-code for the primary controller loop.

7.7. Main controller loop pseudocode

Pseudocode for the main poll loop running trajectory tracking is shown below. This makes use of function calls to calculate inverse kinematics and implement PID control and these functions are also described with pseudocode in subsequent sections.

Main Poll loop

Result: Moves arm along target trajectory
Perform Initialization of variables and trajectory
while *want tracking* **do**

Call the menu object for input commands and act accordingly

Calculate time interval 'h' since the last update

Read the desired target endpoint location (x, y, z)

Call Inverse Kinematics function to get target joint angles θ_1, θ_2 and θ_3

Compute corresponding arm control angles Φ_1, Φ_2 and Φ_3

Read the three motor actuator angles M_1, M_2 and M_3

Call PID function with control target angle Φ_1 and motor angle M_1 as input motor which returns motor velocity control U_1

Call PID function with control target angle Φ_2 and motor angle M_2 as input motor which returns motor velocity control U_2

Call PID function with control target angle $(\Phi_2 + \Phi_3)$ and motor angle M_3 as input motor which returns motor velocity control U_3

Generate PWM control velocities using Arduino for U_1, U_2 & U_3 and using H-bridges apply drive to the respective worm-drive motors

end

7.8. Inverse kinematics pseudocode

An essential aspect of the tracking process involves mapping the desired extrinsic position targets to the corresponding arm joint angles using inverse kinematics. We employ an implementation of inverse kinematics based on Eqns. (17, 33, 34), which is presented here as pseudocode in C++ style. The actual inverse kinematics code is written in C++ and makes use of a constructor to set up the link lengths L_1, L_2, L_3. L_4 and the endpoint angle θ_{end} in member variables. The inverse kinematics pseudocode is as follows:

Inverse Kinematics

Input: Desired target end point location (x_{end}, y_{end}, z_{end})
Link lengths: L_1, L_2, L_3, L_4
Endpoint angle: θ_{end}
Result: Target joint angles θ_1, θ_2 , θ_3, θ_4
Return: IK success or fail flag

Function call

Calculate $\theta_1 = \text{atan2}(y_{end}, x_{end})$

Calculate x-axis of planar arm: $x_{p3} = \text{sqrt}\,(x_{end}^2 + y_{end}^2)$

Account for end link: $x_{p2} = x_{p3} - L_4$

Account for base height of planar arm: $y_{p2} = z_{end} - L_1$

Calculate $\cos(\theta_3)$: $c_3 = (x_{p2}^2 + y_{p2}^2 - L_2^2 - L_3^2) / (2 * L_2 * L_3)$

Calculate $\sin(\theta_3)$: $s_3 = \text{sqrt}(1 - c_3^2))$

Check if can reach target: if $((1 - c_3^2) < 0)$ return **false**

Calculate $\theta_3 = -\text{atan2}(s_3, c_3)$

Calculate variable $k_1 = L_2 + L_3 * c_3$

Calculate variable $k_2 = -L_3 * s_3$

Calculate $\theta_2 = \text{atan2}(y_{p2}, x_{p2}) - \text{atan2}(k_2, k_1)$

Calculate $\theta_4 = \theta_{end} - \theta_2 - \theta_3$

Succeeded reaching target so: return **true**

end

7.9. PID block pseudocode

The PID function control code is written in C++. This code includes a constructor for setting up the gain parameters as member variables, a reset function which is needed to initialize the last time and last error member variables, and to reset the error integral. It also includes and a PID error block that calculates the control signal based on the target and measured joint angles. Pseudocode is for the computation of the control U is shown below.

PID Control

Input: Desired control target angle Φ
Measured motor angle θ_M
PID gains: k_p, k_I, k_d
Last error: e_{last}
Integral of error: e_i
Current and last times $T_{current}$, T_{last}
Result: Motor velocity control **U**

Function call

Calculate proportional error: $e_p = \theta_M - \Phi$

Calculate time step: $h = T_{current} - T_{last}$

Update last time: $T_{last} = T_{current}$

Calculate error derivative: $\frac{de}{dt} = (e_p - e_p)/h$

Calculate error integral: $e_i = e_p + e_i$

Calculate PID control: $U = e_p * k_p + e_i * k_i + \frac{de}{dt} * k_d$

end

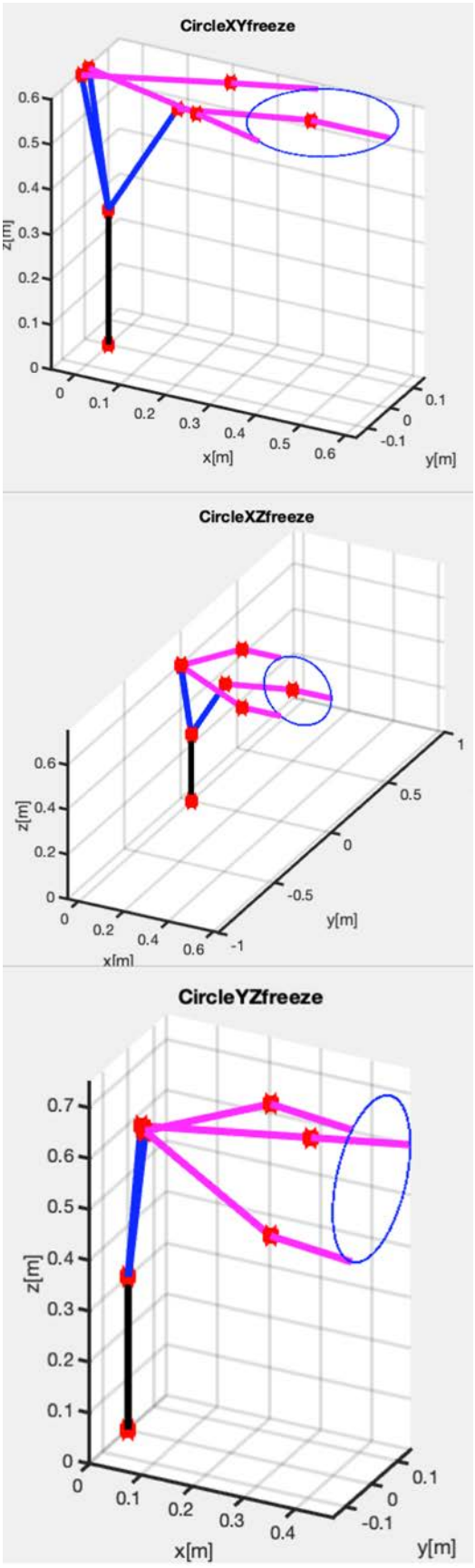


Figure 15. Arm configurations to selected targets. This visualization demonstrates how the arm accesses three distinct points on target circles within the XY, XZ, and YZ planes. The arm joint angles are computed via inverse kinematics, while the arm joint positions are computed through the application of forward kinematics

8. Arm operation

8.1. MATLAB Simulation

To test and showcase the inverse kinematics, we implemented a set of simple tracking tasks in MATLAB. We generated three distinct trajectories. Each of these trajectories is composed of 500 points, uniformly distributed around the perimeters of circles in the XY, XZ, and YZ planes. These circles, with their centers positioned at the (x, y, z) coordinates of (0.475, 0.0, 0.60) meters and a radius of 0.15 meters, comfortably fit within the robotic arm's operational workspace. Utilizing inverse kinematics, we computed the necessary arm angles to target specific points on these circles. Figure 15 visually presents both the desired circles and the requisite arm postures to access three chosen points from each circle.

We made use of Eqn. (15), to calculate motor shafts angular velocities corresponding to the endpoint of the arm moving around the circle every 2 seconds. Figure 16 illustrates this, revealing it corresponds to an endpoint speed of 0.47 m/s. The peak joint speed was identified to be 1.64 Rad/s, which is notably beneath the maximum speed the motors are capable of achieving, 1.77 Rad/s, as denoted by Eqn. (1). It's noteworthy that pursuing the target circle with the arm's endpoint through inverse kinematics and calculating velocities via finite differences deliver consistent results.

Using Eqn. (16) and discounting the impact of the arm's inherent dynamics (a simplification that holds true primarily at low movement velocities), we computed the torques at the motor shafts. These torques are required to sustain a realistic payload of 1kg while applying a force of 1Kgf in the negative x-direction. This is for endpoint positions encircling the target circle in the YZ plane. Figure 17 reveals a peak joint torque of approximately 8.56Nm. Notably, this value marginally surpasses the designated joint torque the motors can produce, which is 8.33Nm as referenced in Eqn. (2). However, this is attainable during intermittent operation.

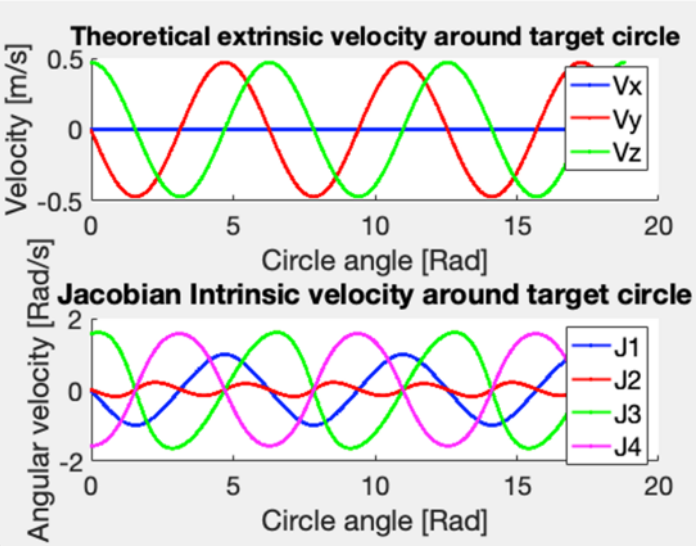


Figure 16. Arm velocity relationships: The arm's endpoint navigates around a target circle with a 0.15 m radius, executing three full rotations within a 10-second time frame. The center of the circle coincides with the arm's endpoint in its canonical position. Panel A: Extrinsic endpoint velocities of the arm. Panel B: Angular velocities of the arm joints

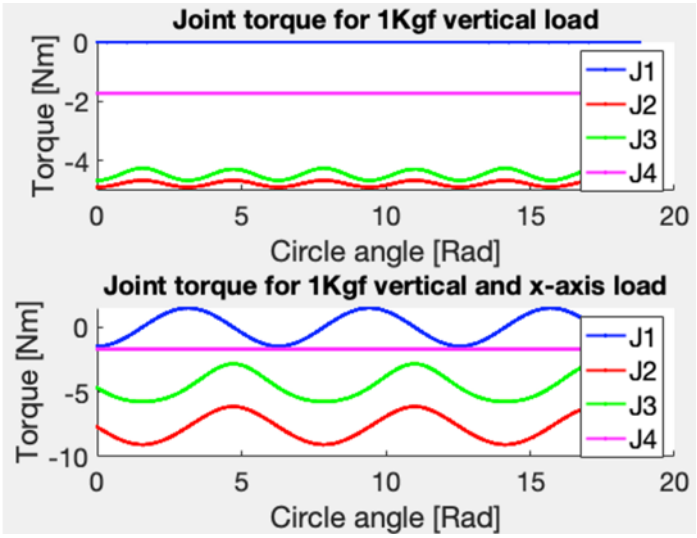


Figure 17. Simulated joint torques under combined loading conditions: Effect of simultaneous 1 Kgf horizontal and 1 Kgf vertical loads. The scenario involves the endpoint moving around a target circle with a 0.15 m radius, the center of which aligns with the endpoint of the arm in its canonical position.

8.2. Trajectory tracking using the physical robotic arm

Demonstrations of tracking around XY, XZ and YZ circles were performed using the actual physical robotic arm.

To implement tracking around square and circular trajectories, firstly target 3D trajectories were generated offline using the microcontroller and held in an array. During the trajectory tracking operation, target points were read out from the array at the appropriate time-index, and their values converted to target joint angles using the microcontroller's implementation of the arm's inverse kinematics.

A Proportional-Derivative (PD, i.e., PID with a zero integral gain component) controller implemented on the microcontroller was then used to ensure the motors were appropriately driven so that the target joint angles were tracked.

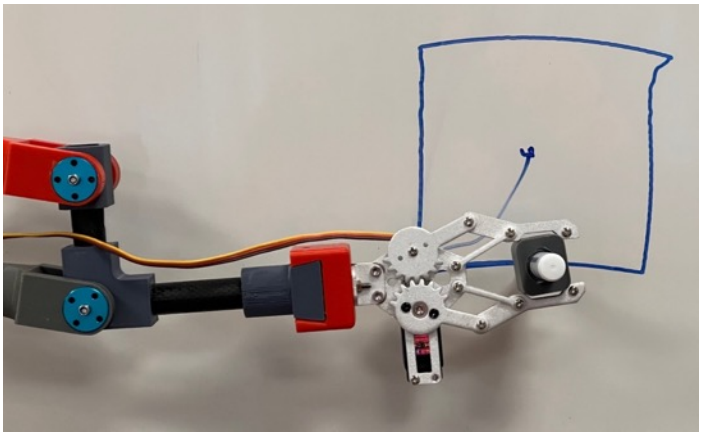


Figure 18: Arm using a low-cost commercially available gripper to hold a felt-tip marker pen and draw on a whiteboard.

8.3. Tracking using the physical robotic arm

For demonstration purposes, a simple commercial gripper was affixed to the arm's endpoint, enabling it to grasp and subsequently release objects. Tracking around a square while holding a pen and drawing on a whiteboard is depicted in the photograph shown in

Figure 18 (refer also to section 8.5 for online YouTube videos). It's worth noting there is some overshoot in the upper right-hand corner due to friction caused by the arm pressing against the whiteboard before suddenly slipping. Future improvements to the controller should address this issue.

8.4. Pick and place operation using the physical robotic arm

We also demonstrate the pick-and-place teleoperation of the arm through a berry-picking task. This task entailed the remote control of the arm to harvest artificial berries from an artificial bush. Both the berries and the bush were designed for a study that evaluated gripper mechanisms under direct human operator control [33].

The teleoperation is implemented as a variant of the previous trajectory tracking task, utilizing inverse kinematics and PID control to appropriately drive the actuating worm-gear motors. However, in this instance, rather than tracking a pre-defined trajectory, keyboard commands were used in real time to increment or decrement a stationary target position in extrinsic space. In this manner, an operator could direct the endpoint of the arm to the desired location in space. Opening and closing the gripper was implemented using an RC servo tester unit.

8.5. YouTube demonstrations of arm simulations and operation

Video demonstrations of these real robot tasks are available on our YouTube channel, "Robotics, Control, and Machine Learning," within the playlist “Design and prototyping of a 3DOF worm-drive robot arm.” The playlist is accessible directly via the following link:

https://youtube.com/playlist?list=PLjKvJX8cBCKWBZDUbaPWgSSMs_msS1MA0.

9. Discussion

9.1. Summary

In this project, we engineered a lightweight robotic arm mechanism, capable of operating in a three-dimensional workspace, primarily designed with the agricultural industry in mind. However, it could also find applications in various other industries, as well as providing a platform suitable for education in robotics.

The integration of a two-dimensional planar mechanism, affixed to a vertical rotary axis, achieves comprehensive 3D spatial coverage. The parallelogram design takes inspiration from the vBOT robotic manipulandum, a research instrument utilized to explore human sensorimotor control in arm movements [37].

We opted for worm-drive actuation to ensure that static postures could be sustained without necessitating an active motor drive, thereby mitigating power consumption. By leveraging off-the-shelf and light-weight structural components, the design not only becomes cost-effective but due to its minimal moving mass, also remains inherently safe particularly when operated in close proximity to individuals. Overall, this culminates in a design that is easy to fabricate, low-cost, lightweight, and is power-efficient.

9.2. Extension of previous work

The current work extended a previous prototype design in several important ways. An additional vertical rotary axis is added, enabling the original planar arm mechanism to rotate and thus operate in a 3D workspace. The kinematic analysis is expanded, and forward, inverse, and differential kinematics for the new 3D arm structure are derived. Furthermore, the former planar arm support structure has been fully redesigned, and we now incorporate 3D-printed pulleys on the arm itself, replacing the previous cumbersome aluminum pulleys. Finally, in the results section, we present evidence of the arm's point-to-point operation, including pick-and-place tasks, demonstrating its ability to perform practical real-world activities.

9.3. Novel contribution of the work

The novel aspects of the design arise from combining the use of 3D printing, carbon fiber sections for arm fabrication, a parallelogram planar upper arm structure that passively maintains endpoint orientation, and the use of worm gear motors.

The key features of the arm, as well as its similarities and differences with the previous prototype arm presented in [1], are summarized in Table 2.

9.4. Bill of materials and cost of the arm

The arm design incorporates numerous low-cost components, resulting in a construction that can be assembled relatively inexpensively. Table 3 presents a bill of the main materials for the project, indicating the costs of individual components as listed by online suppliers for small quantities.

The total cost of constructing a single unit (in 2023) is approximately £480, although discounts would likely apply for larger quantities. While assembly costs have not been included in this calculation, the arm, can easily be put together within a day by a single person. This assumes that the 3D printed custom parts have already been printed and are available for assembly.

Clearly, a construction cost of less than £500 makes the design competitive against many commercial arms, as well as those targeted more towards the hobby market. The former typically start at several thousand pounds per unit, while the latter are still on the order of £1000.

Table 2: Features of the new arm design highlighting differences with the previous 2D prototype

ARM FEATURE	NEW ARM	PROTOTYPE ARM
SIMILARITIES IN DESIGN		
Worm-drive motor actuation	Yes	Yes
Low weight carbon fiber upper limb construction	Yes	Yes
Passive end-link horizontal orientation mechanism	Yes	Yes
Construction makes use of low-cost off-the-shelf components wherever possible	Yes	Yes
Construction technique employs 3D printing for custom parts	Yes	Yes

DIFFERENCES IN DESIGN		
Linear endpoint degrees of freedom (DOFs) of the arm in extrinsic space	3	2
Arm configuration	Revolute arm mechanism	Planar 2D mechanism
Dovetail end-effector attachment point	Yes	No
End-effector accessories	The new design integrates the use of both a commercial and a new bespoke gripper mechanism	No accessories
Demonstration of pick-and-place operation	Yes	No
Timing belt drive mechanism	Driven from custom 3D printed pillars and baseplate	Driven from baseplate
Forward kinematic analysis and implementation	FK derived and implemented for 4-link 3DOF arm	FK derived and implemented for 3-link 2DOF arm
Inverse kinematic analysis and microcontroller implementation	IK for 4-link 3DOF arm	IK for 3-link 2DOF arm

Table 3: Main arm mechanism bill of materials indicating component costs in (as of 2023)

DESCRIPTION	UNIT COST	NUMBER	TOTAL
DRIVE MECHANISM COMPONENTS			
Worm-drive motor DC 24V 28RPM with Hall encoder	£35	3	£105
50mm I/D 90mm O/D 20mm Thick deep-groove bearings	£10	1	£10
HTD 5M 24 tooth pulley	£12	3	£36
HTD 345 5M belt	£16	3	£48
ARM STRUCTURAL COMPONENTS			
CF tube 16x14 x250mm	£12	3	£36
CF tube 16x14 x100mm	£6	1	£6
CF rod 8mmx250mm	£12	2	£24
8mm 200mm stainless shaft	£5	1	£5
Aluminum flange 3mm I/D x H13 x D10	£2	14	£28
M3 x 95mm bolt	£1	8	£8
M6 x 150mm bolt	£2.50	6	£15
M6 x 170mm bolt	£2.50	4	£10
8mm I/D x 22mm O/D x 7mm thick bearing	£2	2	£2
Miniature bearings I/D 3mm x 10mm O/D x 4mm	£0.8	16	£13
Shaft collars 8mm	£1.5	2	£3
ARM CONTROLLER COMPONENTS			
Arduino Mega clone	£16	1	£16
Dual H-Bridge Motor Drive Controller Board L298N	£2.5	2	£5
Buck converter to power RC servo driven accessories	£2.5	1	£2.5
DIN rails 300mm	£3	2	£6
DIN clips	£2	6	£12
DIN rail terminals	£1	22	£22
Arduino DIN rail Mount	£8	1	£8
3D PRINTING MATERIAL			
1Kg roll PLA+	£20	3	£60
TOTAL SUM			£ 478

9.5. Simple upgrades to the current design

In our current design, the fourth end-link consistently maintains a 0° orientation angle, resulting in a translation of the overall endpoint along the horizontal x-axis. However, alternative fixed endpoint orientations could be achieved, by adapting the design of the wrist component.

We utilized parts fabricated from PLA+ through 3D printing. While this material is a suitable choice for prototyping due to its printing ease, more sturdy and environmentally resilient materials could be seamlessly substituted for actual deployment, such as in fruit-picking applications. Potential alternatives might encompass the use of PETG, ABS, nylon, or even composite materials [33, 34].

In line with the approach adopted by an earlier prototype [1], incorporating an efficient controller that estimates motor torque by monitoring motor current would be a simple task. This feature would enable the incorporation of a safety mechanism into the controller, devised to identify unforeseen collisions. Motor current could be compared with anticipated values for a specific task, enabling automatic deactivation of the control if these limits were surpassed.

9.6. Future improvements

The current arm system employs a simple controller based on the Arduino Mega and L298N Dual H-bridges. While this cost-effective combination displays reasonable performance and is suitable for the current proof-of-concept arm design, it does exhibit significant shortcomings. The Arduino Mega's default capability to generate a PWM operating frequency for controlling three

outputs is restricted, achieving only a relatively low operating frequency of 490 Hz. Additionally, the microcontroller's clock speed is quite slow (16 MHz), limiting the control loop update rate to around 100 Hz. Future developments of the arm will involve creating a higher-performance controller, leveraging faster, cost-effective microcontrollers such as the ESP32.

Current measurement of endpoint position using motor encoders cannot account for potential deformations or flexibilities inherent in the arm's structure, or the slight backlash in the worm gear motors. Integrating additional angular position sensors on the arm's joint axes, e.g., using low-cost magnetic encoders, would improve the estimation of endpoint position which could be used to improve arm operational accuracy.

Direct force sensing using force transducers would provide an effective method for the estimation of endpoint force. Although commercial industrial force transducers are prohibitively expensive, research suggests that 3D-printed sensors may provide a suitable low-cost solution to this issue [38].

Finally, it is noteworthy that in the design described here, direct-drive brushless DC motors could serve as a substitute for the current method of actuation and support force and torque control performance [39]. However, this would also incur higher motor costs and escalate power consumption, particularly during the maintenance of static arm configuration.

Conflict of Interest

The author confirms there are no conflicts of interest to disclose.

Acknowledgements

We thank Enterprise Solutions at the University of Plymouth for their financial support of this work on a project entitled, "Low-cost Compliant Robotic Arm for Fruit and Vegetable Harvesting". We also thank our industrial partner, Fieldwork Robotics Ltd., and in particular, Martin Stoelen of Western Norway University of Applied Sciences, for their invaluable support and insightful discussions.

References

[1] I.S. Howard, "Design and prototyping of a low-cost light weight fixed-endpoint orientation planar Cobot," in 2022 International Conference on System Science and Engineering (ICSSE), IEEE: 47–54, 2022, doi:10.1109/ICSSE55923.2022.9947353.

[2] T. Duckett, S. Pearson, S. Blackmore, B. Grieve, W.-H. Chen, G. Cielniak, J. Cleaversmith, J. Dai, S. Davis, C. Fox, "Agricultural Robotics: The Future of Robotic Agriculture," UK-RAS Network, **UK-RAS White Papers**, 2018, doi:https://www.ukras.org.uk/wp-content/uploads/2021/01/UKRASWP_AgriculturalRobotics2018_online.pdf.

[3] A. Bechar, C. Vigneault, "Agricultural robots for field operations: Concepts and components," Biosystems Engineering, **149**, 94–111, 2016, doi:https://doi.org/10.1016/j.biosystemseng.2016.06.014.

[4] F. Chaumette, S. Hutchinson, P. Corke, "Visual servoing," Springer Handbook of Robotics, 841–866, 2016, doi:https://doi.org/10.1007/978-3-319-32552-1_34.

[5] D.F. Kutz, A. Wölfel, T. Meindl, D. Timmann, F.P. Kolb, "Spatio-temporal human grip force analysis via sensor arrays," Sensors, **9**(8), 6330–6345, 2009, doi:10.3390/s90806330.

[6] B. Jia, A. Zhu, S.X. Yang, G.S. Mittal, "Integrated gripper and cutter in a mobile robotic system for harvesting greenhouse products," in 2009 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE: 1778–1783, 2009, doi:10.1109/ROBIO.2009.5420430.

[7] S. Gorjian, H. Ebadi, M. Trommsdorff, H. Sharon, M. Demant, S. Schindele, "The advent of modern solar-powered electric agricultural machinery: A solution for sustainable farm operations," Journal of Cleaner Production, **292**, 126030, 2021, doi:https://doi.org/10.1016/j.jclepro.2021.126030.

[8] I.O. for Standardization, "Robots for industrial environments–safety requirements–Part 1: robot, ISO10218-1: 2006," 2006.

[9] B. Vanderborght, A. Albu-Schäffer, A. Bicchi, E. Burdet, D.G. Caldwell, R. Carloni, M. Catalano, O. Eiberger, W. Friedl, G. Ganesh, "Variable impedance actuators: A review," Robotics and Autonomous Systems, **61**(12), 1601–1614, 2013, doi:https://doi.org/10.1016/j.robot.2013.06.009.

[10] G. Grioli, S. Wolf, M. Garabini, M. Catalano, E. Burdet, D. Caldwell, R. Carloni, W. Friedl, M. Grebenstein, M. Laffranchi, "Variable stiffness actuators: The user's point of view," The International Journal of Robotics Research, **34**(6), 727–743, 2015, doi:https://doi.org/10.1177/027836491456651.

[11] M.F. Stoelen, R. de Azambuja, B. López Rodríguez, F. Bonsignorio, A. Cangelosi, "The GummiArm Project: A Replicable and Variable-Stiffness Robot Arm for Experiments on Embodied AI," Frontiers in Neurorobotics, **16**, 24, 2022, doi:https://doi.org/10.3389/fnbot.2022.836772.

[12] M.F. Stoelen, F. Bonsignorio, A. Cangelosi, "Co-exploring actuator antagonism and bio-inspired control in a printable robot arm," in From Animals to Animats 14: 14th International Conference on Simulation of Adaptive Behavior, SAB 2016, Aberystwyth, UK, August 23-26, 2016, Proceedings 14, Springer: 244–255, 2016, doi:https://doi.org/10.1007/978-3-319-43488-9_22.

[13] A. De Luca, A. Albu-Schaffer, S. Haddadin, G. Hirzinger, "Collision detection and safe reaction with the DLR-III lightweight manipulator arm," in 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE: 1623–1630, 2006, doi:10.1109/IROS.2006.282053.

[14] S. Haddadin, A. Albu-Schäffer, G. Hirzinger, "Safety evaluation of physical human-robot interaction via crash-testing," in Robotics: Science and systems, Citeseer: 217–224, 2007, doi:10.15607/RSS.2007.III.028.

[15] A. Albu-Schäffer, S. Haddadin, C. Ott, A. Stemmer, T. Wimböck, G. Hirzinger, "The DLR lightweight robot: design and control concepts for robots in human environments," Industrial Robot: An International Journal, **34**(5), 376–385, 2007, doi:https://doi.org/10.1108/01439910710774386.

[16] J.E. Colgate, W. Wannasuphoprasit, M.A. Peshkin, "Cobots: Robots for collaboration with human operators," in Proceedings of the 1996 ASME international mechanical engineering congress and exposition, 1996, doi:https://doi.org/10.1115/IMECE1996-0367.

[17] W.T. Townsend, J.K. Salisbury, "Mechanical design for whole-arm manipulation," in Robots and Biological Systems: Towards a New Bionics?, Springer: 153–164, 1993, doi:https://doi.org/10.1007/978-3-642-58069-7_9.

[18] K. Bodie, C.D. Bellicoso, M. Hutter, "ANYpulator: Design and control of a safe robotic arm," in 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE: 1119–1125, 2016, doi:10.1109/IROS.2016.7759189.

[19] S. Haddadin, S. Parusel, L. Johannsmeier, S. Golz, S. Gabl, F. Walch, M. Sabaghian, C. Jähne, L. Hausperger, S. Haddadin, "The franka emika robot: A reference platform for robotics research and education," IEEE Robotics & Automation Magazine, **29**(2), 46–64, 2022, doi:10.1109/MRA.2021.3138382.

[20] S. Mick, M. Lapeyre, P. Rouanet, C. Halgand, J. Benois-Pineau, F. Paclet, D. Cattaert, P.-Y. Oudeyer, A. de Rugy, "Reachy, a 3D-printed human-like robotic arm as a testbed for human-robot control strategies," Frontiers in Neurorobotics, **13**, 65, 2019, doi:https://doi.org/10.3389/fnbot.2019.00065.

[21] T. Lens, J. Kunz, O. Von Stryk, C. Trommer, A. Karguth, "Biorob-arm: A quickly deployable and intrinsically safe, light-weight robot arm for service robotics applications," in ISR 2010 (41st International Symposium on Robotics) and ROBOTIK 2010 (6th German Conference on Robotics), VDE: 1–6, 2010.

[22] G.A. Pratt, M.M. Williamson, "Series elastic actuators," in Proceedings 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots, IEEE: 399–406, 1995, doi:10.1109/IROS.1995.525827.

[23] M. Quigley, A. Asbeck, A. Ng, "A low-cost compliant 7-DOF robotic manipulator," in 2011 IEEE International Conference on Robotics and Automation, IEEE: 6051–6058, 2011, doi:10.1109/ICRA.2011.5980332.

[24] D. V Gealy, S. McKinley, B. Yi, P. Wu, P.R. Downey, G. Balke, A. Zhao, M. Guo, R. Thomasson, A. Sinclair, "Quasi-direct drive for low-cost compliant robotic manipulation," in 2019 International Conference on Robotics and Automation (ICRA), IEEE: 437–443, 2019, doi:10.1109/ICRA.2019.8794236.

[25] M. Fernandez-Vicente, W. Calle, S. Ferrandiz, A. Conejero, "Effect of infill parameters on tensile mechanical behavior in desktop 3D printing," 3D Printing and Additive Manufacturing, **3**(3), 183–192, 2016, doi:https://doi.org/10.1089/3dp.2015.0036.

[26] V.H. Pinto, J. Gonçalves, P. Costa, "Model of a DC motor with worm gearbox," in CONTROLO 2020: Proceedings of the 14th APCA International Conference on Automatic Control and Soft Computing, July 1-3, 2020, Bragança, Portugal, Springer: 638–647, 2021, doi:https://doi.org/10.1007/978-3-030-58653-9_61.

[27] B. Zhang, Y. Xie, J. Zhou, K. Wang, Z. Zhang, "State-of-the-art robotic grippers, grasping and control strategies, as well as their applications in agricultural robots: A review," Computers and Electronics in Agriculture, **177**, 105694, 2020.

[28] J. Hernandez, M.S.H. Sunny, J. Sanjuan, I. Rulik, M.I.I. Zarif, S.I. Ahamed, H.U. Ahmed, M.H. Rahman, "Current designs of robotic arm grippers: a comprehensive systematic review," Robotics, **12**(1), 5, 2023.

[29] F. Rodríguez, J.C. Moreno, J.A. Sánchez, M. Berenguel, "Grasping in agriculture: State-of-the-art and main characteristics," Grasping in Robotics, 385–409, 2013.

[30] J.F. Elfferich, D. Dodou, C. Della Santina, "Soft robotic grippers for crop handling or harvesting: A review," IEEE Access, **10**, 75428–75443, 2022.

[31] E. Navas, R. Fernandez, D. Sepúlveda, M. Armada, P. Gonzalez-de-Santos, "Soft grippers for automatic crop harvesting: A review," Sensors, **21**(8), 2689, 2021.

[32] J. Hughes, U. Culha, F. Giardina, F. Guenther, A. Rosendo, F. Iida, "Soft manipulators and grippers: a review," Frontiers in Robotics and AI, **3**, 69, 2016.

[33] L. Alvarez-Hidalgo, I.S. Howard, Human evaluation of robotic grippers for berry picking, Springer Nature Switzerland, 2023, doi:https://doi.org/10.1007/978-3-031-43360-3.

[34] B. Siciliano, L. Sciavicco, L. Villani, G. Oriolo, "Motion control," Robotics: Modelling, Planning and Control, 303–361, 2009, doi:https://doi.org/10.1007/978-1-84628-642-1.

[35] K.M. Lynch, F.C. Park, Modern robotics, Cambridge University Press, 2017, doi:10.1017/9781316661239.

[36] M.W. Spong, S. Hutchinson, M. Vidyasagar, Robot modeling and control, Wiley New York, 2006.

[37] I.S. Howard, J.N. Ingram, D.M. Wolpert, "A modular planar robotic manipulandum with end-point torque control," J Neurosci Methods, **181**(2), 199–211, 2009, doi:10.1016/j.jneumeth.2009.05.005.

[38] N. Hendrich, F. Wasserfall, J. Zhang, "3D printed low-cost force-torque sensors," IEEE Access, **8**, 140569–140585, 2020, doi:10.1109/ACCESS.2020.3007565.

[39] H. Asada, T. Kanade, "Design of direct-drive mechanical arms," 1983, doi:https://doi.org/10.1115/1.3269106.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 6, 78-86 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Implementation of a GAS Injection Type Prefabricated Lifting Device for Underwater Rescue Based on Location Tracking

Jong-Hwa Yoon[1], Dal-Hwan Yoon[2,*]

[1]*TheQuest Co. Ltd., Seoul, 05853, Korea*

[2]*Semyung University, Department of Electronic Engineering, Jecheon, 27136, Korea*

ARTICLE INFO

Article history:
Received:14 August, 2023
Accepted:15 October, 2023
Online: 30 November, 2023

Keywords:
Lifting fixture
Air cartridge
Marine disaster
Monitoring system
Underwater communication

ABSTRACT

In this paper, we have developed a gas injection-type prefabricated lifting device based on location tracking to efficiently lift the human body in the event of an accident that occurs underwater on the sea or land. The efficiency of the lifting system is very important to ensure the golden time of the rescue and the safety of divers in the event of casualties underwater. Divers performing underwater safety rescue operations must endure up to 30 minutes with two air vents, and always consider the safety accident environment due to difficulty in securing visibility or high flow rates due to underwater turbidity. Particularly, there are many cases where life is threatened by hypothermia in the water. Therefore, both divers and the deceased need location tracking connected to the lifting device, and a fast and efficient lifting system was studied in underwater activities. The monitoring device uses a communication speed of 115.2 kbps from the sensor to the monitoring, and a communication speed of 2.4 kbps from the controller to the receiving unit. The gas injection-type prefabricated lifting device with a high elastic structure is lightweight and portable, and which consists of a baggy bag with minimal components to increase usage and work efficiency based on the instinctive behavior of divers. Accordingly, the entrance element design combining a bow and hinge that maintains a moment of force with TPU-based materials, a balanced design using weight balancing technology of a network structure, an SMB linkage design that induces water surface rise through gas injection, and an underwater experiment.

1. Introduction

In the event of an accident in the water of the sea and land, it is very important to secure a golden time for saving lives. In addition, divers who search the human body underwater should also be accompanied by securing safety. The mobile navigation control system has been studied that tracks the position of the human body while lifting the human body and monitors the movement of the position [1]. These underwater communication devices are attached to rescue workers and the human body and serve to transmit location data. Figure 1 shows a location monitoring that transmits location data from underwater. J. H. Yoon, S. I. Kang, and D. H. Yoon developed a location monitoring device using underwater ultrasonic sensor signals and conducted research to transmit location tracking signals to mother ships or land control centers in LTE or Lora [2, 3].

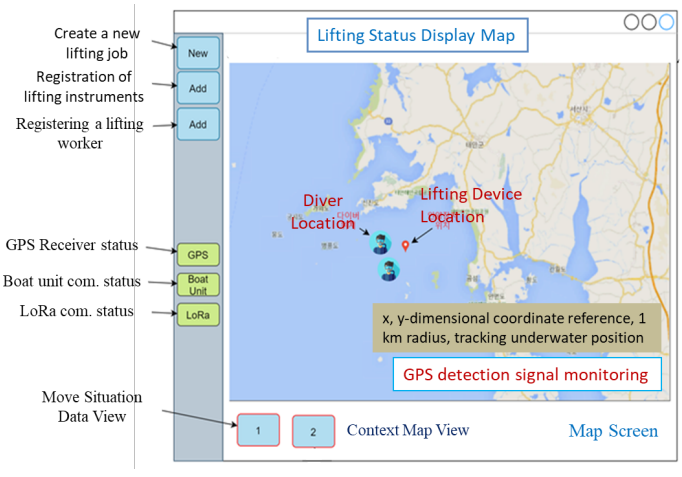


Figure 1: The surface location of ICT-based multi-communication system

Figure 2 shows a block diagram of the IoT injury motion monitoring system. The monitoring device may relay communication on a moving position and relay it to the land by

*Corresponding Author: Dal-Hwan Yoon, Semyung University, 65 Semyungro, Jecheon City, Chung-cheong Buk Do, 27136, Korea. E-mail: yoondh@semyung.ac.kr

LTE communication. The GPS data can then be used to track the location of the accident.

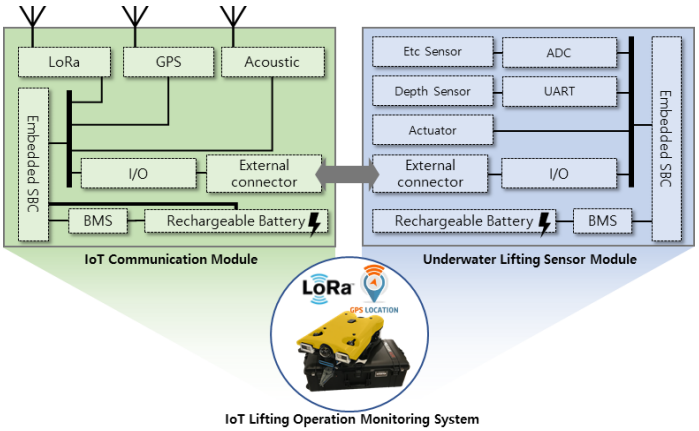


Figure 2: Block diagram of IoT behavior monitoring system

H. K. Min and D. H. Yoon studied a light diffusion device in which the information transmission system can float on the water surface and light is displayed [4]. It consists of a detection unit mounted on a lifting device underwater to detect a lifting situation and a control module that converts related information into an optical signal. This study includes an optical transmission line connected to transmit the optical signal of the control module to the optical diffusion mechanism.

D. H. Yoon and M. H. Park studied to provide a device and method for tracking the location of a target by continuously identifying the current location of a recovery object or rescue object such as a ship or drowning in the event of a water accident or emergency such as sinking, collision, subversion, fire, or drowning [5]. It consists of a hydraulic sensor that detects the water pressure applied to the object, a floating object separation function that separates the floating object from the object if the sensing water pressure exceeds the set value, a communication unit that communicates with the floating object, and a GPS receiver that calculates the current position.

Meanwhile, in [6] the author developed process fusion and production automation technology based on the infusion method to produce a boat with enhanced stability. Process fusion-based boat manufacturing devices to strengthen safety analyzed the balance and center of gravity of the boat in 3D (dimension), and studied the size of the bonding area, curvature of the bonding surface, and strength of the bonding surface. Particularly, the boat manufacturing device has the function of adjusting the mixing ratio of the resin and the curing agent and adjusting the discharge amount per unit time based on the results of the joint area analysis. This time, it includes controlling the manufacturing process, such as the width joint of the boat, makeup taping, accessory adhesion, and sealing by spraying adhesive at a set temperature according to 3D flow analysis.

Existing human lifting technology consists of zipping the instrument that holds the human body on the seabed, and most of the technologies are applied to raise the instrument by injecting air into a separate lift bag when lifting it to the surface [7, 8]. However, the integrated package set that combines roll-fold-based gas expansion technology does not show reliable performance for consumers [9]. Particularly, when diving into the water, it is subject to a lot of buoyancy resistance due to the weight and volume of the diver's equipment, and which complains of difficulties due to the time it takes to work underwater and the complexity of the work manual.

Technology for lifting the human body underwater should not miss Golden Time in the event of a marine accident, and the safety of divers should be secured first. It is very important not only to complete the work within 30 minutes of oxygen consumption time, but also to secure underwater turbidity, underwater flow rate, and visibility. To this end, various studies are needed, including durable materials, weight reduction, structure considering flow rates, and easy-to-handle manuals. In addition, since the marine area is wider than the land area, the cost of on-site helicopters, police patrol boats, diving personnel, and on-site command centers in the event of an accident is enormous [10]. Particularly, as seen in the Ferry Sewol incident in Korea, the underwater turbidity and flow rate were very severe, making it very difficult for divers to access the sunken ship, and it was also confirmed in cases where the safety of field workers could not be guaranteed.

In this study, we develop a package-type human lifting device system in which the lifting device rises to the surface through the quantity and flow rate flowing through the lifting device, renewable materials and durable designs, and underwater shooting of CO2 cartridges. In addition, the equipment is designed by applying an air cartridge method to minimize the time spent underwater through the convenience, weight reduction, and rapid operation of rescue workers. Through this, research is conducted on securing the reliability of the rapid lifting and lifting process.

2. Implementation of GAS Dispensing Human Lifting Device

2.1. Underwater Environment

The design of the lifting device is optimized based on various underwater environment data through prototype development. Korea's underwater environment has severe waves on the east coast, and it takes longer to darken than the bright time underwater between sunrise and sunset. In the case of the West Coast, the turbidity is so serious that it is difficult to check the front 1m, and there is a risk that everyone will be exposed to safety accidents [2]. Moreover, the difference between the tides, where the sea level changes vertically twice a day every 12 hours and 25 minutes, represents 9m from the coast in Incheon and 3m from Mokpo [9]. Compared to the tidal phenomenon in which the height of seawater changes vertically, algae move horizontally, and Korea's west coast is known to be the three most severe tidal regions in the world, along with Canada's Pun Dee Bay and France's Saint-Michel Bay. As such, if the vertical tidal phenomenon of seawater is severe, the horizontal tidal flow becomes fast, and marine accidents such as ship stranded occur frequently. Particularly, seawater cooled in winter begins to rise in temperature from the surface of the sea from April to May, and when seawater is mixed due to the temperature difference between the surface of the coast and the deep sea, the flow of algae increases. Frequent fog occurs due to the formation of electric wires, so caution is needed in ship navigation.

Another aspect of the underwater environment is that the bottom of the sea floor is piled up with light-type soil of tens of centimeters to tens of meters, and soil is pouring out of the Yellow River in China, making the sea foggy. Particularly, it is very

important to secure visibility in the event of a marine accident for the soil environment caused by severe turbidity and floating matter. Figure 3 shows diving activities in various underwater environments.

Figure 3: Diving activities in various underwater environments

In this study, depending on the worst underwater environment, the lifting device is implemented in consideration of the mesh structure, portability, weight reduction, and operability to reduce buoyancy in water. It is characterized by designing a lifting device in the form of an integrated lifting device and a surface marker Buoy (SMB), and firing a CO2 cartridge to quickly discharge it to the surface of the water. The 33g CO2 cartridge attaches two manipulators to attach two to the lifting mechanism. Until now, it is understood that no lifting net-integrated airbag has been developed yet.

The lifting mechanism utilizes the effect of rising to the surface of the water through the underwater bombardment of the CO2 cartridge. The air cartridge method was applied while requiring the portability and rapid operation of the lifting device. Therefore, research on securing the reliability of the rapid lifting and lifting process is continuing. Figure 4 compares the mesh-type lifting device designed as an early prototype with the lifting device used by the National Police Agency's scientific investigation unit.

2.2. Design of Lifting Instruments

The lifting device design is designed in consideration of mass production through portability, convenience and simplicity of use, production cost reduction, and mold improvement, and components are designed in consideration of the efficiency and convenience of underwater work. At this time, the flow rate and quantity are calculated by applying Boyle's law to secure the flexibility of the lifting mechanism in water. Figure 4 explains the Boyle law for the quantity and flow rate experiment flowing through the lifting mechanism. Here, d represents the length, A represents the cross-sectional area, Q represents the flow rate, t represents the time taken, and $\bar{v}$ is the average speed. We can explain the Figure 4 as follows an equation (1)

$$\mathrm{Q} = \frac{V}{t} = \frac{Ad}{t} = A\,\bar{v},\ \bar{v} = \frac{d}{t} \tag{1}$$

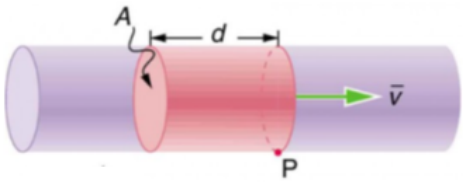


Figure 4: Designing the flow rate and quantity flowing through the lifting device

Lightweight is needed to minimize weight underwater and ease portability. The flow rate is calculated to secure flexibility as follows the equation (2).

$$\mathrm{Q} = \mathrm{Av},\ \ \mathrm{A} = \frac{Q}{v} = \frac{\pi}{4}\,d^2, d = \sqrt{\frac{4Q}{\pi v}} \tag{2}$$

Figure 5 uses TPU, a light and renewable durable material that taking account of the quantity and flow rate flowing through the lifting mechanism. The inlet of the lifting instrument plays an important role in underwater work. In addition, it is designed with a mesh structure so that the human body and water entering the entrance exit at a high speed.

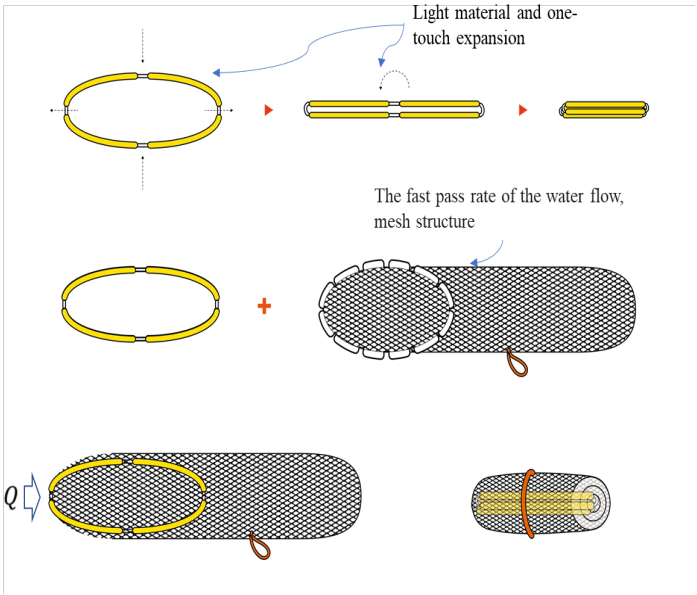


Figure 5: Detailed Design of Lifting Appliances

When the diver enters the water, the amount of oxygen in the oxygen tank is compressed, increasing the fatigue of the underwater worker. The deeper you get into the water, the more oxygen (within 30 minutes) shrinks, and the faster the work should be done in consideration of the falling time, working time, and rising time.

Figure 6 shows the operation of underwater using the designed lifting mechanism rule. Use the hook law for tensile stress and load during underwater operation. In equation (3), $\sigma(\mathrm{kg}/cm^2)$ is tensile stress, P is load (tension), E is cross-sectional area, and is elastic coefficient.

$$\sigma = \frac{P}{A} = E\epsilon = \frac{M}{Z} \tag{3}$$

here, ϵ is the coefficient, $M(cm.kg)$ is the silver bending moment, and $Z(cm^3)$ represents the cross-sectional coefficient. Use durable material TPU 80 and 95 for the bow.

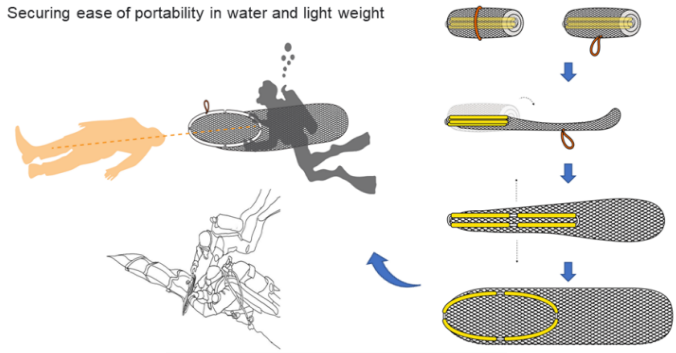


Figure 6: Underwater working behavior

In order to maintain extreme bending or the shape of the round part when combining the hinge and the bow, the occurrence of play due to elastic loss over time was minimized. The hinge connection was designed with a cross-shaped double support structure, but it was very difficult to maintain its shape due to tension loss. Therefore, it was designed with a support structure of thickness and width that gradually became thinner around the bow, and which developed into a two-piece one-stage structure. Figure 7 shows the double support structure and design for maintaining the roundness value and shape.

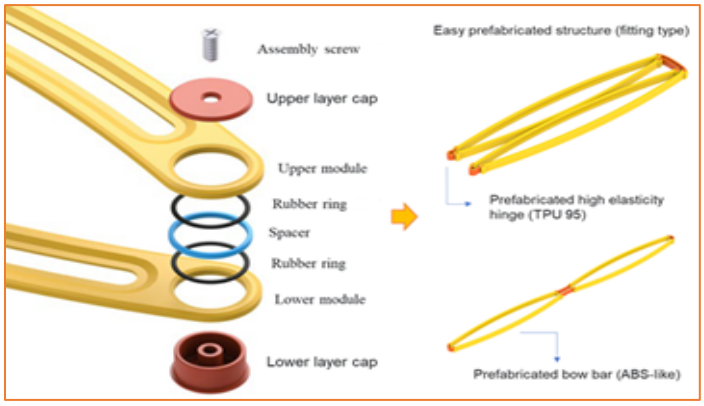


Figure 7: High elasticity hinge design and detailed schematic

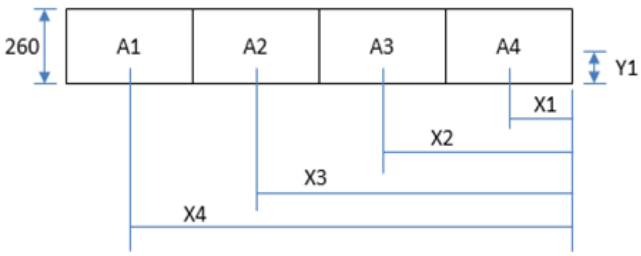


Figure 8: The lifting mechanism of network structure

When the human body is inserted into the lifting device and raised to the sea level, the weight balance technology of the network structure is applied to configure the head to face the sea level. The lifting mechanism uses a formula $h1 = h2 = h3 = h4 = h$ and $Y1 = Y2 = Y3 = Y4$ to design a weight-balanced algorithm for 2000x260mm SMB and network structures, which can calculate X1, X2, X3, and X4 differently. Figure 8 shows the weight balance algorithm of the lifting mechanism network structure. Figure 8 shows the lifting mechanism of network structure.

Figure 9 shows the weight balance algorithm of the lifting mechanism network structure based on figure 8.

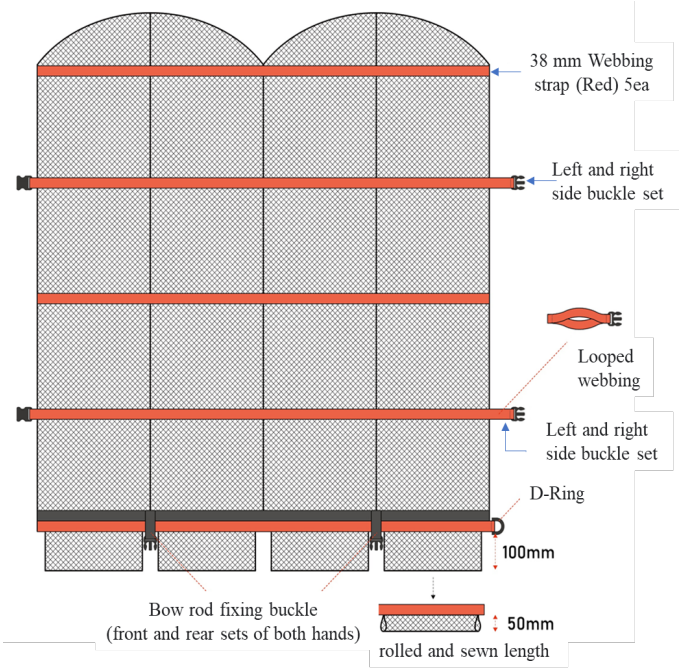


Figure 9: Weight balance algorithm of the lifting mechanism network structure

On the other hand, in water, SMB (Surface Marker Buoy) is designed in a form with a certain slope. The equation for volume and area S of SMB is as follows;

$$S = \frac{(a+b)}{2} h, \quad F_i = \rho V_i g \tag{4}$$

here, F_i is the buoyancy, ρ is the density of water and g is gravity acceleration. The scientific design of the lifting mechanism take into account for the weight held when underwater and the ease of work due to underwater internal force. Particularly, after inserting a dead body into a lifting device, weight balancing technology is applied to configure the head to face the sea level. Figure 10 shows the design cases according to the length of the SMB and mesh net with a certain slope.

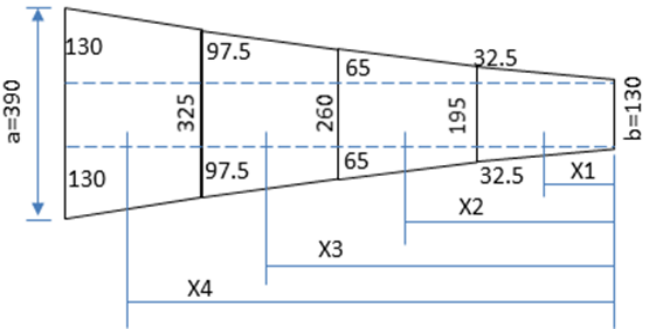


Figure 10: SMB and mesh gradient angle with constant gradient

Figure 11 shows the lifting mechanism packaged with the designed SMB and mesh network. Tightening devices were added to both sides in consideration of the expansion of the lifting mechanism. In the SMB header part, it has the pump valve, and the oral inflator are valves that take account of CO2 loss, and consist of an air inlet using the last CO2 cartridge.

Figure 11: Lifting device packaged with a mesh net

When working at a depth of 20 → 30 → 40m each, the worker's physical strength decreases severely due to securing time to the surface of the water and the difference in buoyancy of the human body [5]. When working in pairs, it is very important not to lose a team member. Considering that the maintenance time of the oxygen tank held by the underwater worker is 30 minutes, it takes more than 2 hours of sufficient rest time on the water surface. Figure 12 shows the folding method and size of the lifting device packaged with SMB and mesh net.

Figure 12: Folding method and size of lifting appliance packaged with SMB and mesh net

Figure 13 shows the monitoring tracking system and surrounding accessories. The monitoring system plays a very important role in securing the location of the lifting instrument and the location of underwater workers [2, 6]. The underwater ultrasonic sensor signal is supplied using a USB port, and the O/S is configured with Linux. In the commercialization stage of the lifting system, low-end type development is easy and compatibility is wide.

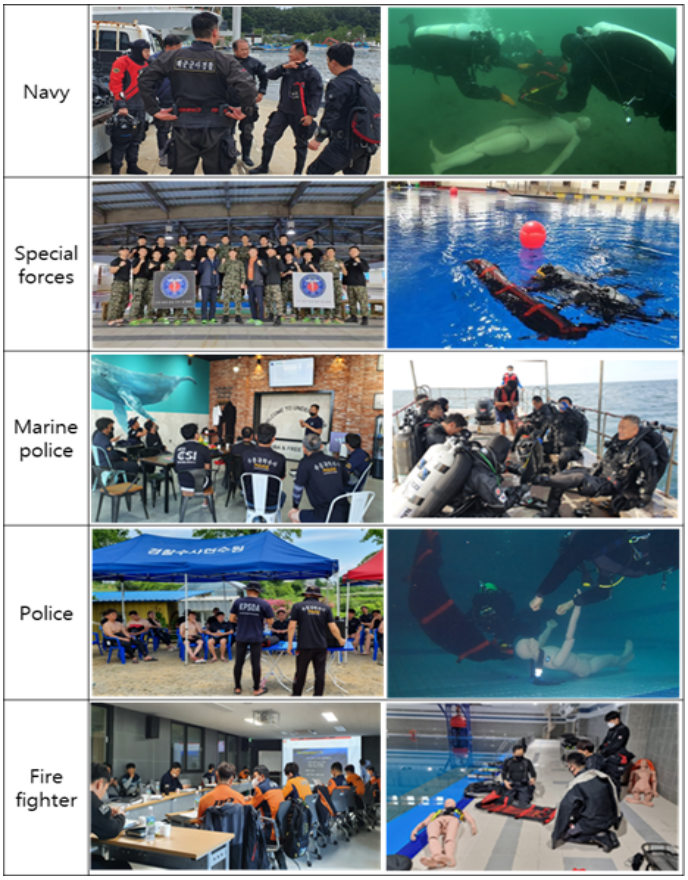


Figure 13. Monitoring position tracking system and peripheral accessories

3. Experimental Results

3.1. Experimental Scenarios

The design of the lifting device is optimized based on various underwater environment data through prototype development. Korea's underwater environment is characterized by severe waves on the east coast, and the water quickly darkens except between sunrise and sunset. In the case of the west coast, the turbidity is so severe that it is difficult to check 1m ahead, and there is a risk that everyone will be exposed to safety accidents. Particularly, despite the difficulty of the test environment due to the tangled waste and mud, this can be explained as the worst lifting condition in the event of an accident.

The quantity and flow rate flowing through the lifting mechanism, renewable materials, and durable designs were selected. Safety and efficiency are the top priorities of the package-type human lifting system in which the lifting device rises to the surface through the underwater shooting of the CO2 cartridge. Particularly, the GAS injection method SMB was packaged with a mesh network to minimize the time spent underwater through the convenience, weight reduction, and rapid

operation of rescue workers. Through this, an experiment was conducted on securing the reliability of the rapid lifting and lifting process.

When the depth of the water deepens, the diver's decompression limit time is 8 minutes for work above 40m, and the effective air capacity of the air tank is reduced by 1/4 by water pressure. After 20→30→40m of underwater entry work, the worker's physical strength is severely deteriorated due to the difference in buoyancy of the human body and securing time to come to the surface of the water. When working in pairs, it is very important not to lose the group due to underwater forgetting. Considering that the maintenance time of the oxygen tank held by the underwater worker is 30 minutes, it takes more than 2 hours of the sufficient rest time, Particularly, musculoskeletal diseases, mental diseases, and rheumatism that appear in divers are the reasons why sufficient rest and management are needed after underwater work. Figure 14 shows awareness of fatigue and stress that may appear after work.

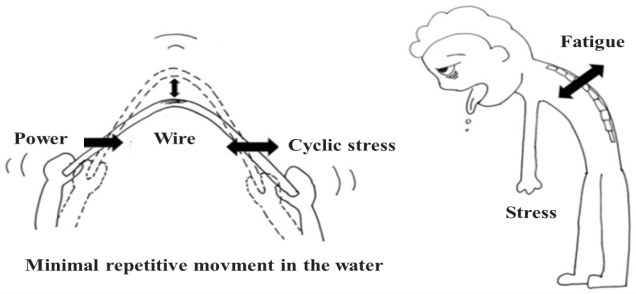


Figure 14: Awareness of fatigue and stress after work

The underwater communication test was conducted in real time from 6m to 40m in depth. The ultrasonic sound sensor is converted to 2,400bps and verifies the transmission error according to the binary. The communication speed from sensor to monitor is 115,200bps, and the communication speed from controller to receiver is 2,400bps.

Figure 15: Performing underwater work based on various environment

In order to gain the reliability of the experiment, public-private forces (Navy, Marine Police, Fire Department, and Public Safety Submersible Association) jointly conduct underwater human lifting training and experiments. Particularly, it was based on the experience of participating in Ferry Sewol events in Korea's West Sea (Korea Public Safety Submersible Association), and it was found that securing the safety of divers was also very important by conducting experiments based on various marine environment data. Figure 15 shows each participating organization that performs underwater work based on various environmental data, including the author. The organizations that participated in the study included naval special teams, maritime police, land police, firefighters and divers.

3.2. Lifting Appliances Experiment

Figure 16 compares the performance of each prototype with the proposal of a lifting mechanism of the rotary ring method, zipper and yelcro method, and ring cage method.

Figure 16. Comparison of the rotating ring method, zipper and Velcro method, and Linkage method

The rotating ring method of the initial prototype takes time to fasten in the rotating ring, and the weight is increased underwater using a mesh net made of plastic. In particular, it was difficult to dive more than 10m due to buoyancy when diving into the water. The zipper and Velcro method is a form of combining a straight SMB and fastening the inlet to a mesh net made of plastic with a zipper. When gas is injected into a straight SMB, it takes time because the body is not centered when it rises to the surface, and in the case of an old body, damage was feared. Lastly, the SMB integrated package bag type of the GAS injection method is a lightweight TPU 80 material and the entrance is designed for durability and flexibility.

The product used in the diving investigation department of the Korea Coast Guard requires rapid discharge due to the hops of the human body and the water accommodated through the entrance, and swells underwater, causing inconvenience in working. In order to protect the lifting device SMB, the relief valve provides a function of automatically maintaining a constant pressure by discharging air when the pressure inside the airbag increases. The water pressure corresponding to 5 atmospheric pressure is formed on the seabed surface of 40m deep, so the air capacity is reduced to 1/5 when the air cartridge is operated. When the water pressure decreases as the lifting mechanism rises, the air inside the lifting mechanism relatively increases. When the pressure or more is

formed in the lifting mechanism, air is discharged through the relief valve, and the pressure is controlled. Figure 17 shows the disassembly and assembly process of the relief valve.

Figure 17: Disassembly and assembly of relief valves

Figure 18 shows the water surface rise of the lifting mechanism according to buoyancy and weight after injecting CO2 gas into SMB. The lifting mechanism begins to tilt according to buoyancy and weight, tilting in the vertical direction such as 16.68° for S1, 22.78° for S2, 28.40° for S3, and 33.39, and rising to the surface of the water.

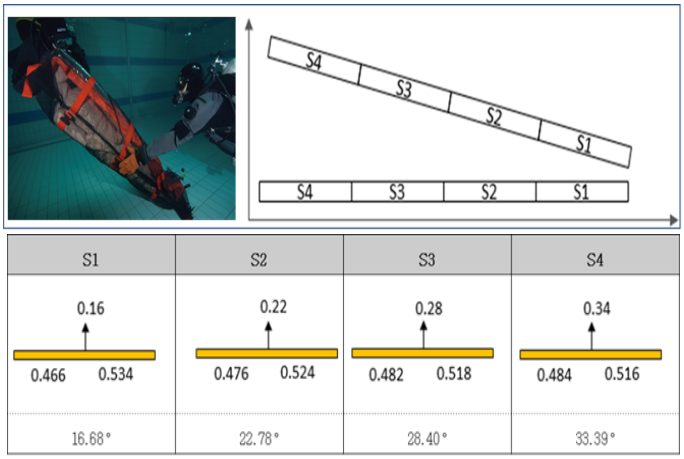


Figure 18: Water level rise of lifting equipment according to buoyancy and weight after CO2 gas injection into SMB

Table 1 shows the underwater characteristics according to diameter, length, weight, and burst pressure according to the CO2 cartridge specification.

Table 1: Underwater data according to CO2 cartridge specification

Vol.(cc)	Dia.(mm)	Length(mm)	weight (g)	Burst pressure (Bar)	Mouth Con.
18	22	88	70	450	No thread
25	25	95	90	450	3/8-24 UNF
32	25	107	105	450	1/2-20 UNF
42	22	160	117	450	No thread
45	22	248	225	450	No thread
60	30	120	155	450	1/2-20 UNF
61	22	248	225	450	No thread
95	40	138	275	450	5/8-18 UNF
110	40	150	450	245	M16 x 1.5

The buoyancy of the lifting instrument was 7-5, 9-5, rotary, and when folded, buoyancy and rising experiments were conducted. At this time, the measured weight = actual weight - was calculated as the weight of the substituted fluid, and the density / fluid density = weight / weight of the substituted fluid.

Table 2: Buoyancy of lifting instruments

Type	Trans. (m)	Length (m)	Height (m)	Volume (㎥)	Buoyancy (N)	Decision
7-5 type	2.00	0.70	0.05	0.070	686.70	rising
9-5 type	2.00	0.90	0.05	0.090	882.90	rising
Cir. Guide	6.00	3.14	0.03	0.00565	55.45	rising
Folded type	1.00	3.14	0.18	0.00565	55.45	rising

The lifting device uses a lifting device with a 4-link bridge structure as a group of two, but it was developed to enable rescue activities for one person. Particularly, the 4 -links inlet of the lifting device is designed to be easily handled by folding once after capturing the human body. In addition, a fastening device was installed to facilitate the expansion of the package that the human structures were simultaneously performed. Figure 19 shows cases such as SMB-integrated package type, SMB separation type, human lifting buoyancy experiment scene, and three-person expansion.

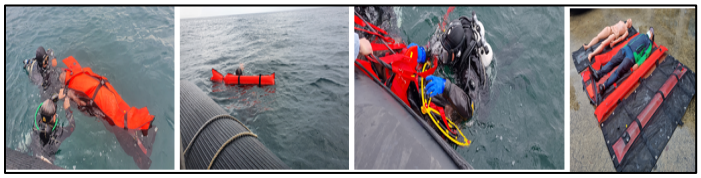


Figure 19: Experiments of SMB package, SMB detachable type, human lift buoyancy check.

Figure 20 shows that marine police and naval investigators are equipped with lifting instruments and conduct underwater experiments. Underwater gravity, underwater weight, weight of the lifting instrument, and total weight of the lifting instrument were measured for the total weight of 120 kg equipped with the equipment equipped with the diver and the lifting instrument package.

Figure 20: Underwater experiments of lifting instruments by maritime police and naval personnel

Gloves are worn to maintain the body temperature of underwater divers, but the working time may be extended due to a decrease in body temperature, and if the lifting device is complicated, it may take excessive working time.

Table 3: Total weight of underwater gravity, underwater and weight

Weight	Gravity of the lifting object	Weight of under water(kg)	Air lifting appliance	Appliance weight (kg)	Total weight (kg)	Gravity (N)
120.00	1.05	6.00	3.00	1.00	7.00	68.67
	1.10	12.00		1.00	13.00	127.53
	1.20	24.00		1.00	25.00	245.25

An example of experimenting is a lifting mechanism packaged with a GAS injection-type SMB and a mesh net at a depth of 40m.

At this time, the season was around May, and the experiment was conducted from 10 a.m. to 12 hours. To secure an underwater view, an LED lantern was equipped, and an experiment was conducted in an environment with a good wave of 2 to 3m. The experiment was completed in a short time as the oxygen supply was reduced to 20 to 25 minutes.

Figure 21 shows the buoyancy formation, acceptable weight, underwater pressure, air injection time and buoyancy, results of the lifting mechanism in accordance with the CO2 cartridge explosion in the experiments of Table 3.

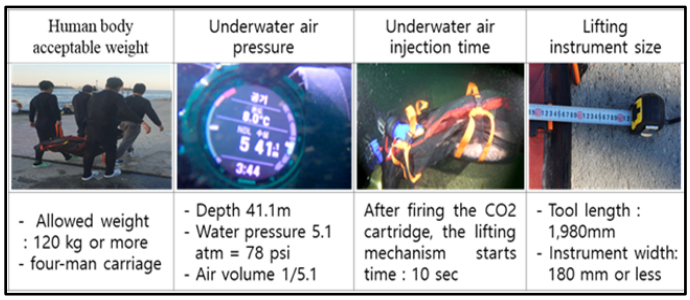


Figure 21: Flotation experiment according to CO2 GAS cartridge explosion at a depth of 40m

Underwater human lifting tests were conducted from 10:00 a.m. to 12:00 p.m. and 14:00 p.m. to 16:00 p.m. from November to December 2021, spring for a month between April and May 2022, and 30 to 40 experiments were conducted over four seasons from October to November 2022.

3.3. Experiment of a location tracking monitoring device

Figure 22 is the main screen of the monitoring device and shows the monitor layout in consideration of the environment of use, such as the lifting instrument state map, location information, and additional functions. The control system is a portable system that is convenient to move and receives sensor signals based on Linux O/S using a USB port. At this time, the temperature measurement data according to gyro, acceleration, and water depth are shown.

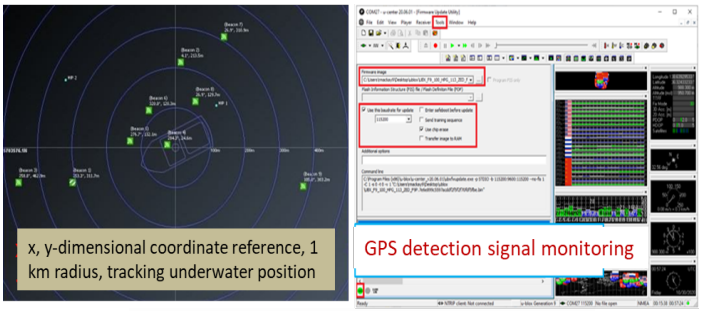


Figure 22: Lifting Status Map and Monitoring Display

Table 4: Measurement data according to gyro, acceleration, depth, and temperature

	Gyro			Acceleration			Depth of the water	Temperature
	yaw	pitch	roll	x-acc	y-acc	z-acc		
0m	130.31	-64.5	77.31	-8.85	-4.29	0.61	-0.11	23.29
-1m	38.13	-81.12	179.25	-9.66	-0.13	-1.4	0.85	23.38
-2m	191.44	-82.56	-172.13	-9.69	0.07	-1.36	1.94	23.49
-3m	97.56	-83.19	179.19	-9.71	-0.19	-1.15	2.96	23.54
-4m	10.81	-82.94	166.56	-9.75	-0.28	-1.03	3.93	23.58
-5m	87.69	-84.25	177.5	-9.78	-0.45	-0.9	4.93	23.63
-6m	252.63	0.37	-119.31	0.05	8.54	-4.61	5.63	23.63

Figure 23 shows the results of experiments on the communication distance and communication speed between the monitoring device and the underwater transmitter located on land and sea.

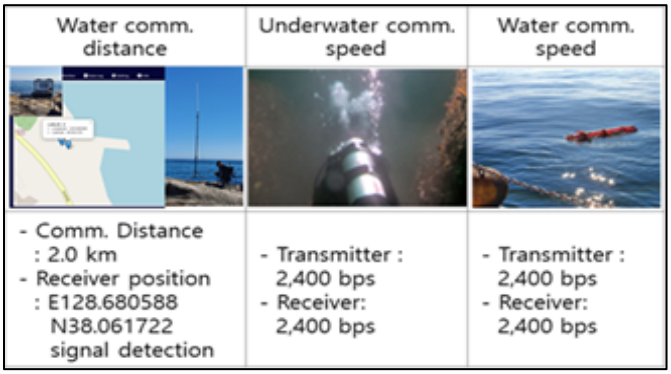


Figure 23: Experiments on the communication distance and speed of the monitoring device

The control monitoring device is portable and receives a sensor signal using a USB port. It can be configured with Linux O/S to provide a wide range of compatibility.

4. Conclusion

In the event of an underwater accident in the ocean or land, this study efficiently lifted the human body and implemented a gas injection-type prefabricated lifting device based on location tracking. To secure golden time and the safety of divers, the efficiency and safety of the lifting system were considered as the top priority for saving lives underwater.

At 40 m underwater, divers must endure up to 30 minutes with two air cylinders, and the worst underwater experimental environment conditions such as securing visibility and dangerous environments with high flow rates were selected due to underwater turbidity. Particularly, there are many cases where life is threatened by hypothermia in the water. Therefore, both divers and the deceased need location tracking connected to the lifting device, and a fast and efficient lifting system was studied in underwater activities.

The gas injection-type prefabricated lifting device with a high elastic structure is lightweight and portable, and a baggy bag is composed of minimal components to increase the use method and work efficiency based on the instinctive behavior of the diver. Accordingly, the entrance element design combining a bow and hinge that maintains a moment of force with TPU-based materials, the balance design using weight balancing technology of the network structure, the SMB linkage design that induces water

surface rise through gas injection, and the entire assembly were tested underwater.

Coast Guard, special investigators, and fire officials participated to improve reliability by using a situation where turbidity was very high and the flow rate of underwater sites was high as the worst field experiment conditions. Due to turbidity, the West Coast's two-person rescue experiment was also exposed to safety accidents, and the study was conducted in a way that auxiliary diving personnel performed guard duties in preparation for safety accidents.

As the water pressure increased by 1 atmosphere every 10m deep, the amount of air in the lifting appliance was compressed to 1/2 compared to the land, and it was found that using one CO_2 cartridge for 60kg underwater mannequin sufficiently reached the surface of the water.

When the water depth decreases in the process of the lifting instrument rising from the seabed to the surface of the water, the water pressure decreases at the same time. At this time, the air inside the lifting device expands and the pressure rises, as it reaches the water surface, the pressure control valve attached to the lifting device operates to discharge air and maintain constant pressure. When 38g of the CO_2 cartridge was injected at a depth of 40m, it took about 10 seconds for the water lift to start, and at the same time as lifting the human body, a manual for securing the safety of the diving source could be established.

At 40m underwater, water depth, water temperature, and direction angle were transmitted to rescue personnel on the water surface as a location detection signal through monitoring and the start of water rise of the lifting device. At a depth of 40m, the lifting test calculated the diving agent's acquisition time, working time, and time taken to get out, considering the oxygen content of about 30 minutes. At this time, the time available for work was as short as eight minutes.

The location tracking monitoring device of the lifting device was implemented in consideration of underwater lifting situation information storage, display, and waterproof functions. LTE technology for short-range and long-distance communication was applied to the acoustic sensors obtained underwater, and interface functions were expanded through USB ports. In the future, we will focus on portability, weight reduction, low battery power, and extended life time.

References

[1] J. H. Yoon, D. H. Yoon, “A Human Lifting and Navigation System Based on ICT for Underwater Rescue,” 2022 The 22nd International Conference on Control, Automation and Systems (ICCAS 2022) BEXCO, Busan, Korea, Nov. 27~Dec. 2022

[2] J. H. Yoon, S. I. Kang, D. H. Yoon, “Implementation of ICT-based Underwater Communication Monitoring Device for Underwater Lifting,” Institute of Korean Electrical and Electronics Engineers (IKEEE), **26**(3), 396~400, 2022

[3] D. H. Yoon, “Location Tracking Apparatus and Method in Water Accident or Emergency Situation,” Patents No.10-2020-0142370, Sept. 2020

[4] H. K. Min, D. H. Yoon et al, "Information Transmission System of Lifting Apparatus" Patents No.10-2020-0127360, 2020

[5] D. H. Yoon, M. H. Park, "Water Accident Emergency Response System and Method Thereof”, Patents 10-2020-0126397, 2020

[6] D. H. Yoon, “Boat Aerostat Production Apparatus and Method for Safety Strengthen Based on Manufacture Convergence,” Patents No.10-2020-0137878, 2020

[7] Science and Technology, “System Assessment and Validation for Emergency Responders (SAVER),” U.S. Department of Homeland Security, 2021

[8] Central Maritime Safety Tribuna, “Announcement of Maritime Accident Statistics in 2021”

[9] H. K. Min, N. H. Seong, et al, “Development of ICT-based human body lifting system for underwater structures”, Techinical Report, Ministry of SMEs and Startups, 2021

[10] J. G. Kim, “The Design and Implementation of Seabed Auto-interpretation System Using Ultrasonic Signal Processing”, 2001

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 6, 87-95 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Towards Real-Time Multi-Class Object Detection and Tracking for the FLS Pattern Cutting Task

Koloud N. Alkhamaiseh[*,1], Janos L. Grantner[2], Ikhlas Abdel-Qader[2], Saad Shebrain[3]

[1]*Department of Computer Science, Michigan Technological University, Houghton, 49931, MI, USA*

[2]*Department of Electrical and Computer Engineering, Western Michigan University, Kalamazoo, 49008, MI, USA*

[3]*Western Michigan University Homer Stryker MD School of Medicine, Kalamazoo, 49008, MI, USA*

ARTICLE INFO

Article history:
Received:26 April, 2023
Accepted:08 October, 2023
Online: 30 November, 2023

Keywords:
Laparoscopic surgery
Object detection
Bag-of-freebies
FLS pattern cut

ABSTRACT

The advent of laparoscopic surgery has increased the need to incorporate simulator-based training into traditional training programs to improve resident training and feedback. However, current training methods rely on expert surgeons to evaluate the dexterity of trainees, a time-consuming and subjective process. Through this research, we aim to extend the use of object detection in laparoscopic training by detecting and tracking surgical tools and objects. In this project, we trained YOLOv7 object detection neural networks on Fundamentals of Laparoscopic Surgery pattern-cutting exercise videos using a trainable bag of freebies. Experiments show that YOLOv7 has a mAP score of 95.2, 95.3 precision, 94.1 Recall, and 78 accuracy for bounding boxes on a limited-size training dataset. This research clearly demonstrates the potential of using YOLOv7 as a single-stage real-time object detector in automated tool motion analysis for the assessment of the resident's performance during training.

1. Introduction

This paper is an extension of work originally presented in DICTA 2021 [1]. In this project, more data is collected and prepared to train YOLOv7 [2] as a real-time object detector of laparoscopic tools and objects in the Fundamentals of Laparoscopic Surgery (FLS) pattern-cutting exercise using a box trainer [3]. Laparoscopic procedures have become increasingly popular in operating rooms worldwide due to their numerous benefits, leading medical schools to incorporate this technique into their surgery curricula [4]. However, the one-on-one apprenticeship model is subjective and time-consuming. To address this issue, laparoscopic trainers and simulators have become well-accepted alternatives that allow for safe and harm-free training [5]. Although simulation systems offer objective measurements and remote training, expert surgeons are still required to assess surgical skills proficiency. Virtual Reality (VR) training provides a completely virtual environment with haptic feedback and complex software, but it is expensive and requires highly sophisticated mechanical design [6]. By improving the real-time object detection of laparoscopic tools and objects, this project aims to enhance the effectiveness and accessibility of laparoscopic training.

Box trainers and physical trainers provide a practical environment for using real laparoscopic instruments to improve basic skills such as knotting, handling objects, and cutting tissues. While time is currently the primary metric for evaluating a surgeon's performance using statistical tools, studies have shown that box trainers enhance trainee confidence and dexterity [7], [8]. However, objective assessments of laparoscopic skills still require experienced surgeon evaluations. To address this issue, hybrid trainers combine the benefits of simulators and physical trainers to recreate real-world conditions and provide objective assessments through integrated software. Hybrid trainers provide a comprehensive approach to evaluating real-world situations by merging both simulators and physical trainers.

The emergence of deep learning [9] has proven to be a highly effective machine learning approach for detecting and classifying objects from raw data by learning representations from the data. Its superior feature extraction and expression capability has surpassed other machine learning methods in many areas, especially when dealing with large data sets. Therefore, deep learning appears to be a very promising method for detecting tool presence [9].

[*]Corresponding Author: Koloud N Alkhamaiseh, +12695449311 & kalkhama@mtu.edu

The main contributions of this research are as follows:

- Contribution to the creation of the first laparoscopic box trainer custom dataset, i.e., the WMU's Laparoscopic Box-Trainer Dataset [10]. This custom dataset was developed through a research collaboration between Western Michigan University's Department of Electrical and Computer Engineering and the Department of General Surgery at Homer Stryker M.D. School of Medicine. Researchers are free to download the dataset at their convenience. This dataset was created specifically to aid research in the field of Laparoscopic Surgery Skill Assessment. It consists of videos showcasing four different tasks on the Laparoscopic Box-Trainer - two precision cutting tests, intracorporeal suturing, and peg transfer. These videos were recorded by surgeons, surgical residents, and OB/GYN residents at the Intelligent Fuzzy Controllers Laboratory at WMU. You can access the dataset at https://drive.google.com/drive/folders/1F97CvN3GnLj-rqg1tk2rHu8x0J740DpC. The dataset also contains labeled images with related labels for all tasks, and more files will be added as the research progresses.
- Proposing a robust real-time multi-class object detection and tracking module based on YOLOv7 as a single-stage real-time object detection neural network for surgical tools and objects in FLS pattern cutting test.

The adoption of YOLOv7 is a wise choice because of its superior network architecture, precise object detection, efficient label assignment, and resilient loss function and model training. Moreover, YOLOv7 is more cost-effective than other deep learning models [2] and is highly proficient in detecting and tracking surgical tools and objects within spatial boundaries.

In order to fully explain our proposed system, this paper is organized as follows: Section 2 gives a brief introduction to the methods used for evaluating the performance of laparoscopic surgery training. In Section 3, we present our methodology. Section 4 contains a summary of our experimental findings, and in Section 5, we outline our plans for future work.

2. Background

FLS tool is widely used for psychomotor skill training in surgery. The American College of Surgeons (ACS) has created didactic instructions and manual skills to improve the basic laparoscopic surgery skills of surgical residents and practicing surgeons using the FLS box trainer [3].

The FLS box trainer, along with didactic instructions and manual skills, can help surgical residents and practicing surgeons improve their basic surgical skills. Current assessment methods focus on detecting surgical tools and analyzing motion, but it's also crucial to track surgical instruments during operations or training to analyze operations and assess training.

In a previous study [5], computer vision algorithms were used to assess performance during surgical tool detection, categorization, and tracking in real-time FLS surgical videos. An artificial neural network learned from expert and non-expert behaviors and a web-based tool was created for uploading MIS training videos securely and receiving evaluation scores with trainee performance analysis over time. The assessment used a multi-dimensional vector consisting of smoothness of motion, proficiency of surgical gestures, and number of errors.

Another study [6] presented a trainer for assessing laparoscopic surgical skills using computer vision, augmented reality, and AI algorithms on a Raspberry Pi programmed in Python. The assessment method employs an artificial neural network based on a predetermined threshold for the peg transfer task. A simulation of pattern-cutting was used to track laparoscopic instruments, while computer vision libraries counted the number of transferred points during the transfer task.

Recent advancements in deep learning, specifically CNN networks, have shown remarkable progress in computer vision tasks [11]. Various studies have implemented deep learning architectures to detect surgical tool presence in laparoscopic videos [12], [13], [14], and phase recognition [15].

Some projects have created systems that can detect laparoscopic instruments in real-time during robotic surgery, using the real-time detection algorithm of the CNN network. These systems are based on the object detection systems YOLO [16] and YOLO9000 [17], with a mean average precision of 84.7 for all tools. They also have a speed of 38 frames per second (FPS).

The available skill assessment frameworks have some limitations when it comes to evaluating fundamental laparoscopic skills based on globally accepted standards and criteria. These frameworks only focus on tool motion and do not consider surgical objects and their manipulation during training. For instance, the Objective Structured Assessment of Technical Skills (OSATS) [16] and the Global Operational Assessment of Laparoscopic Skills (GOALS) [17] are examples of such frameworks. To address this limitation, a new system was developed using a deep learning algorithm called YOLACT [18]. This system tracks surgical tool motion and detects surgical objects, including their deformability, shapes, and geometries in the surgical field of view. The system was tested on a modified FLS peg transfer exercise and provided a more comprehensive evaluation of laparoscopic skills beyond tool motion alone.

To evaluate the abilities of those performing intricate intracorporeal suturing, automated systems have been suggested. One such system, developed by the authors in [19], uses the latest versions of One-Stage-Object-Detectors like YOLOv4, Scaled-YOLOv4, YOLOR, and YOLOX. A dataset of suturing tasks was used to train this system, which strikes a balance between cutting-edge architectures. In [20], the authors proposed a skill evaluation system that employs Scaled-YOLOv4 and a centroid tracking algorithm.

The authors presented a fuzzy logic supervisor system for assessing surgical skills in [21]. This system used multi-class detection and tracking of laparoscopic instruments during standard FLS pattern-cutting tests. However, the system had limitations when the instruments or objects were not within view of the camera. To address this issue, a new autonomous evaluation system was proposed by authors in [22], which utilized two cameras and multi-thread video processing to detect laparoscopic instruments.

In addition, two fuzzy logic systems were implemented in parallel to evaluate left and right-hand movement. The authors in [23] have improved the YOLOV7x algorithm significantly for detecting surgical instruments. These enhancements effectively address concerns about dense arrangements, mutual occlusion, difficulty in distinguishing similar instruments, and varying lighting conditions.

To provide a more comprehensive assessment of surgical quality, our approach includes examining circle shape deformability, as well as laparoscopic tool tracking and detection in box trainer pattern cut test recorded videos.

3. Methodology

The FLS box trainer boasts multiple exercises, but we've focused on perfecting a pattern-cutting exercise. Our system centers around two circles printed on artificial tissue, with a radius of 2.5 and 3.0 centimeters for the inner and outer circles respectively Figure 1. Keeping the scissors within these circles is essential for this exercise, as crossing either circumference will result in an incorrect cut.

- In this paper, we propose a system that works seamlessly with our intelligent FLS box trainer. This system utilizes two cameras placed inside the box to record videos, which are then used to train a deep-learning object detector and tracker. We have employed powerful deep learning algorithms YOLOv7, which have set a new standard in real-time object detection. Our model has been trained, validated, and tested using a custom data set to track three objects - the scissors, the clipper, and the circle - within the box trainer's surgical view. The circle is the object that needs to be cut out, and this is achieved using laparoscopic scissors and clippers.
- Figure 2 illustrates the complete workflow of our method. First, we divide the recorded videos into frames and preprocess them before training the YOLOv7 network. After successfully training the model, it can accurately detect and track intended objects in both tested videos and real-time videos. Lastly, we calculate the performance evaluation parameters and generate an output video that displays the labeled surgical objects.

3.1. Network Architecture

The YOLO family of models has a long-standing association with the Darknet framework [24], tracing back to its inception in 2015 [25].

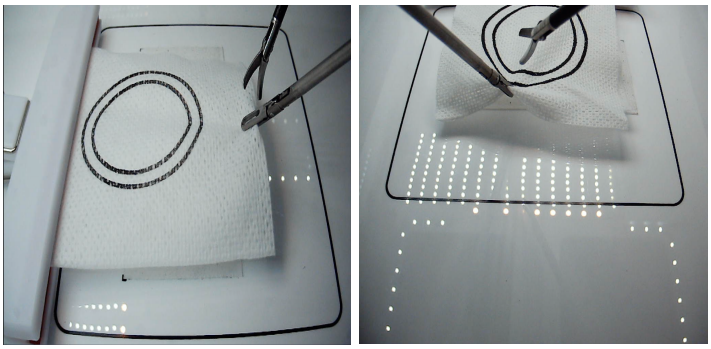


Figure 1: FLS Pattern cutting test setup views as captured by the two cameras.

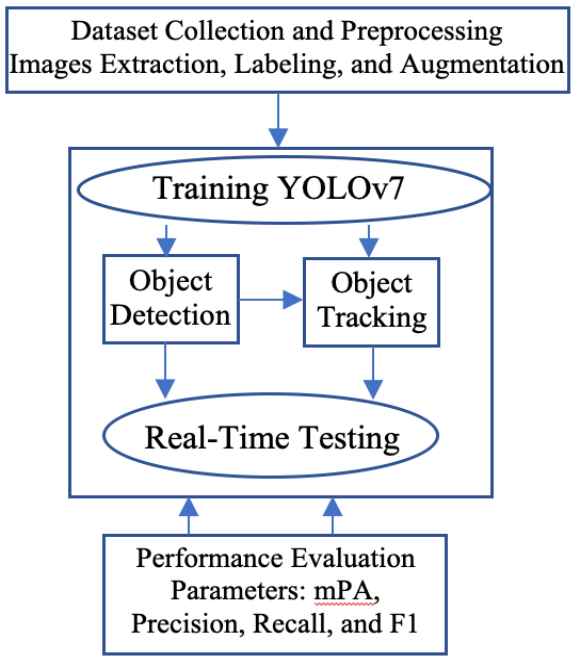


Figure 2: The proposed workflow chart.

The YOLO (You Only Look Once) detection layers rely on regression and classification optimizers to determine the necessary number of anchors. The image is divided into cells using a 19x19 grid, where each cell can predict up to five bounding boxes. However, some of these cells and boxes may not contain an object, so a probability of object presence (PC) is utilized to remove low-probability bounding boxes. Non-max suppression is subsequently used to select the bounding boxes with the highest shared area. YOLO has many versions and variants that enhance performance and efficiency.

The most recent official version, YOLOv7, was created by the original authors of the architecture. It is a single-stage real-time object detector and, according to the YOLOv7 paper [2], it is currently the fastest and most accurate real-time object detector available.

- E-ELAN (Extended Efficient Layer Aggregation Network): It is the computational block in the YOLOv7 backbone, in which the network learns faster by expanding, shuffling, and merging cards to continuously improve its ability to learn without destroying its gradient path.

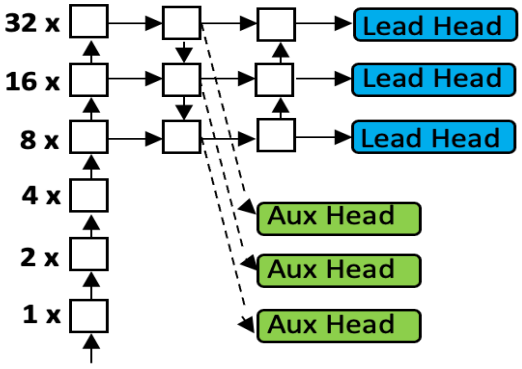
- Concatenation-based model scaling allows the model to maintain the properties that it had at the initial design and thus maintain the optimal structure.
- To replace the convolutional layer or residual with re-parameterized convolution, planned re-parameterized convolution without an identity connection uses RepConv.
- Coarse for auxiliary and Fine for lead loss: As YOLOV7 includes multiple heads, the Lead Head is responsible for the final output, and the Auxiliary Head is used to train in the middle layers as illustrated in Figure 3 (a). Loss assists with updating the weights of these heads, allowing for Deep Supervision and better model learning. A Label Assigner mechanism was introduced to enhance deep network training, which considers the network prediction results and ground truth before assigning soft labels as shown in Figure 3 (b). Unlike traditional label assignment methods that generate hard labels based on given rules by directly referring to the ground truth, reliable soft labels use calculation and optimization methods that also consider the quality and distribution of prediction output together with the ground truth.

YOLOv7 introduces important reforms that significantly improve real-time object detection accuracy while keeping inference costs low. Compared to state-of-the-art real-time object detections, YOLOv7 reduces parameter and computation costs by about 40% and 50%, respectively, resulting in faster inference speeds and higher detection accuracy [2]. YOLOv7 has a fast and robust network architecture that integrates features, provides better object detection performance, and employs an efficient model training process with a robust loss function and label assignment. Overall, YOLOv7 represents the best option for optimizing real-time object detection.

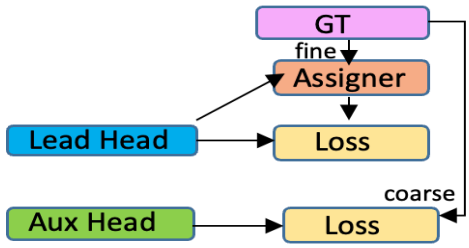
3.2. Dataset

Our laparoscopic detection and tracking project requires a dataset annotated with spatial bounds for objects and tools. To achieve this, we extracted 1572 labeled images from 13 videos ofthe FLS pattern-cutting test, recorded by an expert surgeon and residents from the School of Medicine at Western Michigan University [26]. These videos have a resolution of 640x480 pixels and a frame rate of 30 frames per second and were carefully selected to accurately depict various instrument scenarios, lighting conditions, and angles. We resized the images to 416x416 pixels with auto orientation as a preprocessing step.

Figure 4 displays a ground-truth example utilizing a free preprocessing tool from Roboflow [27]. This tool manually labels the three intended objects- circle, clipper, and scissors in each frame. To augment our dataset's sample size, we employed various techniques, such as affine transformations, rotations, cropping, shearing, hue saturation, and blurs, as depicted in Figure 5. To ensure accuracy, new pixels were filled with the average RGB value of the corresponding image, which has proven to be highly reliable [28]. Our augmentation step yielded 3,458 images, with 87% designated for training, 8% for validation, and 5% for testing. The dataset's distribution of instruments and objects is outlined in Table 1.



(a) The addition of an auxiliary head is included in the model.



(b) Coarse-to-fine lead guide assigner.

Figure 3: YOLOV7 multiple heads and Label Assigner [2].

3.3. Evaluation Criteria

To evaluate our work, we used box loss, objectness loss, classification loss, precision, recall, and mean Average Precision(mAP) as performance metrics. Figure 6 depicts these metrics. By measuring box loss, we were able to determine the algorithm's ability to accurately locate the center of an object and ensure that its bounding box adequately covers it. Objectness measures the probability that an object exists within a proposed region, serving as a confidence metric. High objectness indicates a greater likelihood that an object will be visible in the image window. Classification loss evaluates the algorithm's ability to predict the correct object class. Ground-truth intersection over union (GIoU) refers to the overlap between the ground-truth region and the detection result region. In the context of GIoU judgment, precision is the ratio of true positives to total detections. Meanwhile, recall is the ratio of successful detections to the total number of classes. The mAP, on the other hand, displays our bounding box predictions based on various GIoU thresholds set at mAP@0.5:0.95 and mAP@0.5 on average. To obtain the final estimate, the AP value is computed for each class across all GIoU thresholds, and the mAP is averaged for all classes.

Table 1: Details of the dataset: The number of instances for each instrument type is shown, distributed over 1572 images, and the resulting augmented sample size is included.

Object type	Instant samples	Augmented sample size
Scissors	1253	3759
Clipper	1289	3867
Circle	1432	4296
Total instances	3974	11922
Images	1572	3458

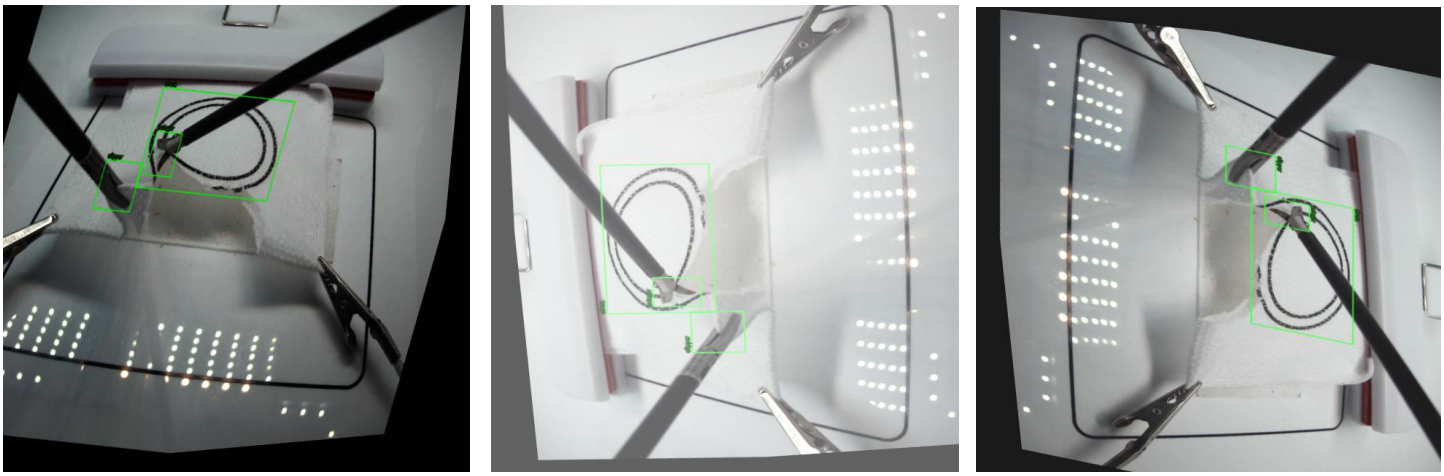


Figure 5: Some sample images for preprocessing using augmentation techniques.

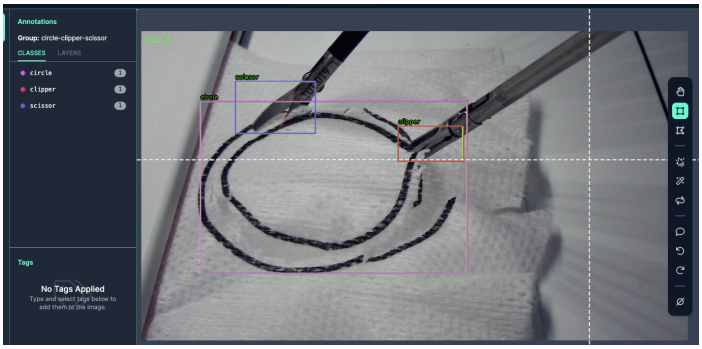


Figure 4: Image labeling process ground-truth example.

4. Experimental Results

This section outlines the setup used to train and evaluate CNN-based models for FLS trainer pattern-cutting test instrument recognition and tracking. We utilized a notebook developed by Roboflow.ai [29] to train YOLOv7 with our custom dataset and yolov7_training.pt pre-trained weights.

4.1 Training

To train YOLOv7, the proposed system used the Tesla P100-PCIE-16 GB GPUs with 56 processors and 16,280 MB of memory obtained from Google Colab [30]. To implement the process and validate the performance of the model scripts, several Python libraries were used (e.g., in [31] and [32]). The YOLOv7 model was trained with 416 x 416-pixel images, 16 batches, and 200 epochs. The training process took 3.357 hours with 10.6G of GPU memory.

4.1 Quantitative and Qualitative Results

In this subsection, we quantitatively and qualitatively evaluate the performance of our approach to detecting and tracking the FLS laparoscopic instruments and objects in the pattern-cutting test using YOLOv7.

The performance details during the training and validation phases are presented in Figure 6 and Table 2. Along with the losses, precision, recall, and mean average precision were calculated using GIoU thresholds of 50% and 50%:95% for up to 200 iterations. The model's precision, recall, and mean average precision exhibited rapid improvement, stabilizing after roughly 100 epochs with minor fluctuations at the start. Moreover, the validation data's classification loss decreased significantly until approximately epoch 50. To select weights, early stopping was employed.

Further, Figure 7 presents a precision-recall curve that provides a granular performance indicator for each class. The PR curve analysis indicated that the circle provided the highest performance (98.7%), followed by the clipper (95.9%), and the scissors (91.0%). It was expected since the scissors were not always clearly visible. They may be obscured by the gauze while cutting and exhibit different orientations as they move. Table 3 shows the precision, recall, and mean average precision values for each class.

Moreover, the F1 score measures the model accuracy by calculating the harmonic mean of precision and recall for the minority positive class. The harmonic mean emphasizes similar precision and recall values; the more precision and recall scores differ, the worse the harmonic mean. This score provides both recall and precision, which means that it captures both positive and negative cases. For all classes, the F1 score for the proposed model is 0.95 at a confidence level of 0.78 as shown in Figure 8.

Figure 6 showcases qualitative results for the suggested module. The top row displays the actual boxes, and the second row reveals the detection and classification outcomes obtained from YOLOv7. Despite the laparoscopic instruments and circles having deformities, varied orientations, locations, and some covering, the detection accuracy is quite high.

Moreover, this system has demonstrated an exceptional ability to detect and track targeted objects with a high degree of accuracy in pattern-cutting test videos, taking only 12.3 milliseconds per frame for processing. An example of a pattern-cutting test with active tracking and detection can be seen in Figure 9, which shows some selected frames from two videos recorded by two different cameras. The trained model can detect and track the scissors, graspers, and circles despite their varying orientations, locations, and coverage.

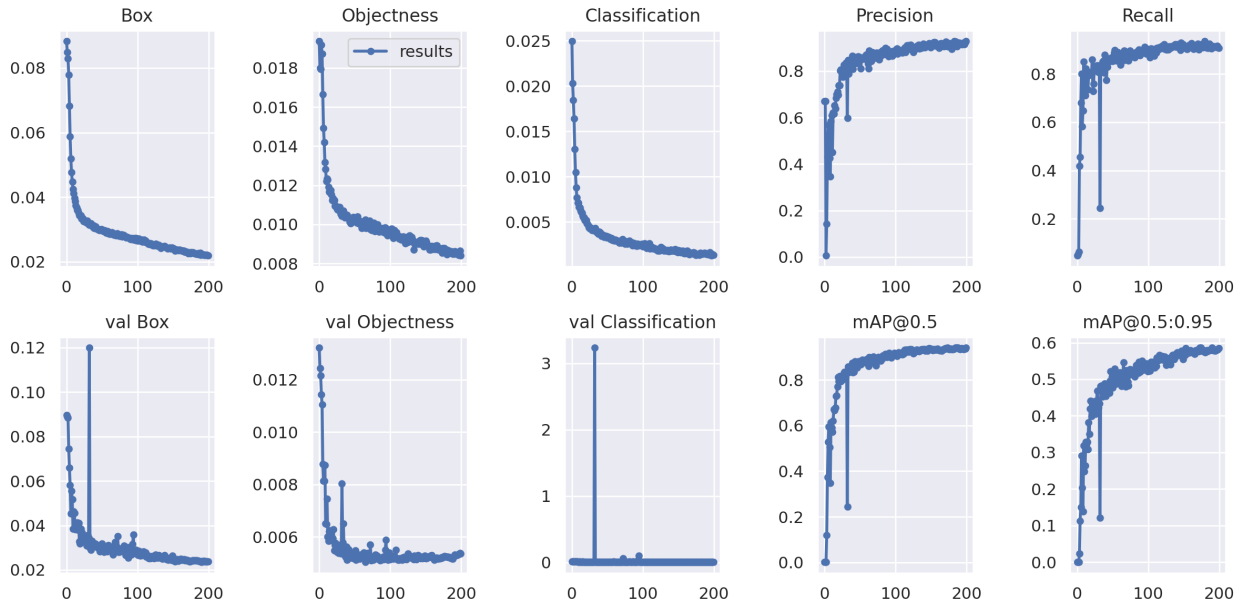


Figure 6: Plots of the box loss, objectness loss, classification loss, precision, recall, and mean average precision (mAP) for both the training and validation sets over the training epochs.

Table 2: Losses, mean average precision, precision, and recall final values.

Evaluation Criteria	Final Value
Box Loss	0.01391
Objectness	0.004788
Class. Loss	0.0003568
mAP@0.5	0.951
mAP@0.5:0.95	0.641
Precision	0.95
Recall	0.941

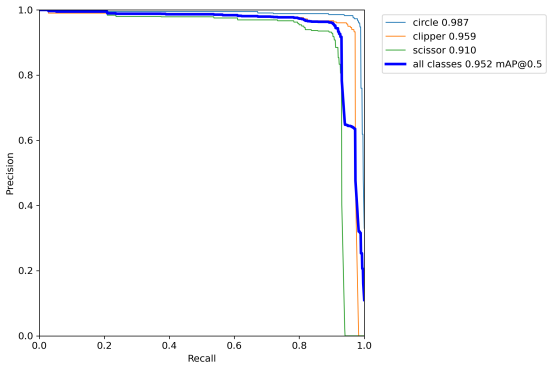


Figure 7: Precision-recall (PR) curves for all classes.

Table 3: Precision, recall, and mean average precision values for each class.

Evaluation Criteria	Circle	Clipper	Scissors
Precision	0.979	0.949	0.932
Recall	0.966	0.957	0.9
mAP@0.5	0.987	0.959	0.91
mAP@0.5:0.95	0.889	0.537	0.495

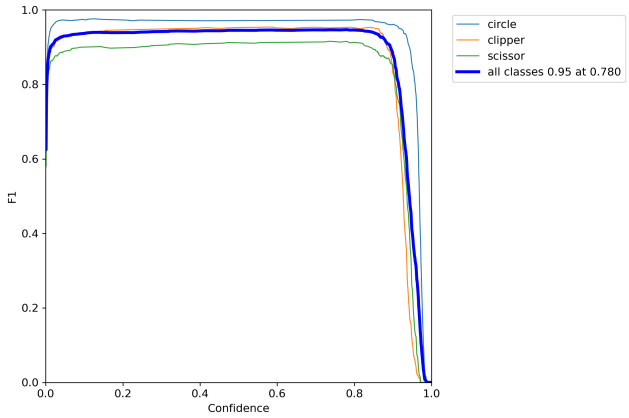


Figure 8: F1 score curves for all classes.

Afterward, testing was conducted on the lab's main workstation. This workstation is equipped with a 2.5GHz Intel(R) Xeon(R) CPU E5-1650 v4 and 32.0GB of RAM. With a delay of 1.67 seconds, the model can detect and track the circle, the grasper, and the scissors. In this case, the delay was caused by the time it took to capture and process frames. To conduct real-time assessments, more powerful hardware is required.

Our study proposes a new approach for detecting and tracking laparoscopic instruments, using deep-learning neural networks.

We have compared our approach with other methods reported in the literature, summarizing the number of extracted images, labeling tools, model built, and results in Table 4. Our proposed approach outperforms previously reported models, achieving an F1 score of 0.95 at a confidence level of 0.78, and a mAP score of 95.2, 95.3 precision, 94.1 Recall, and real-time processing speed of 83.3 FPS, despite the limited number of videos. The models in previous studies may differ in construction, which could explain the differences in results. Overall, our study presents a reliable and efficient method for assessing the performance of trainers in laparoscopic instrument use.

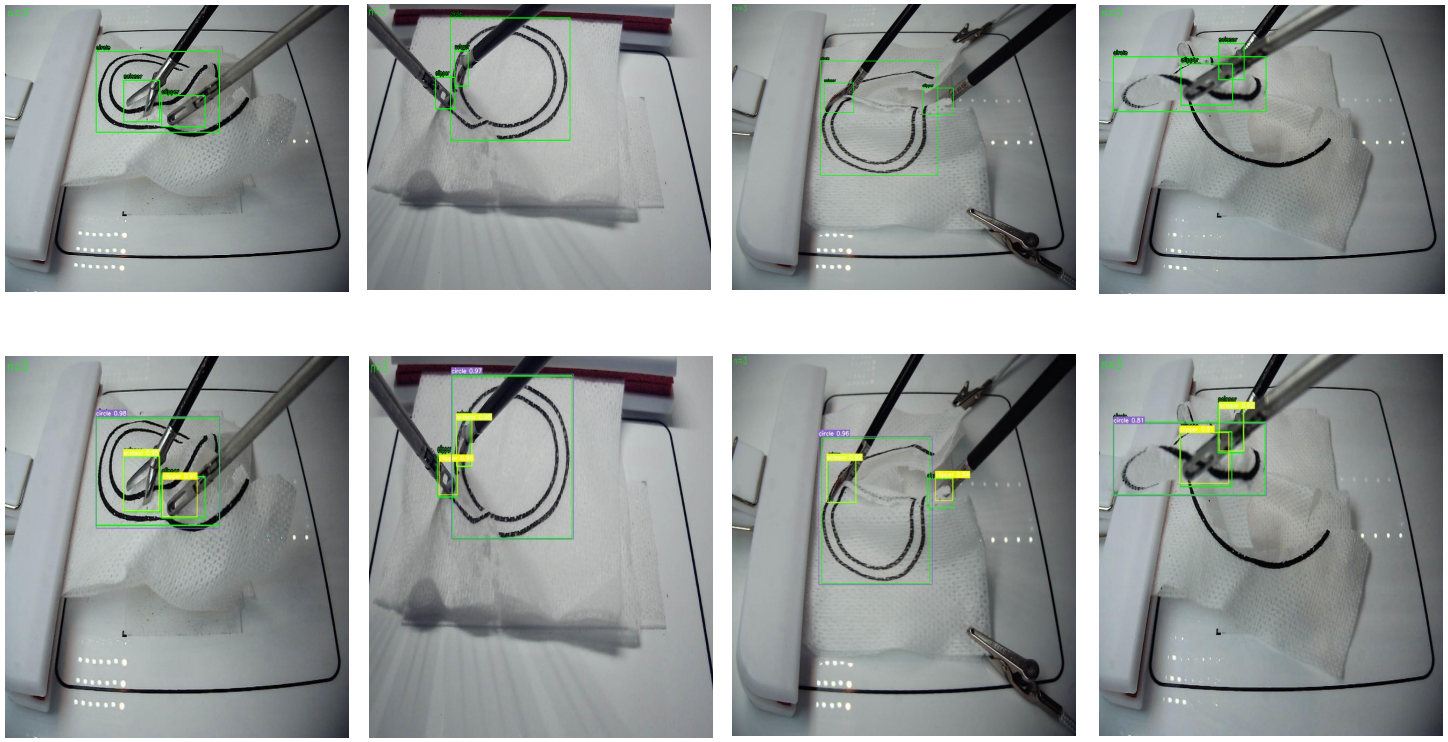
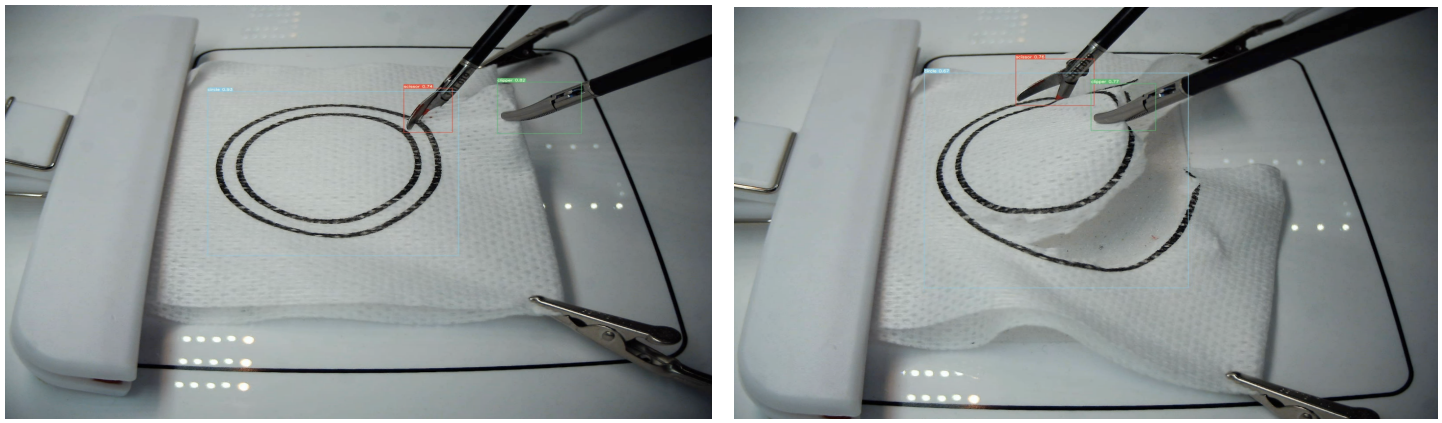
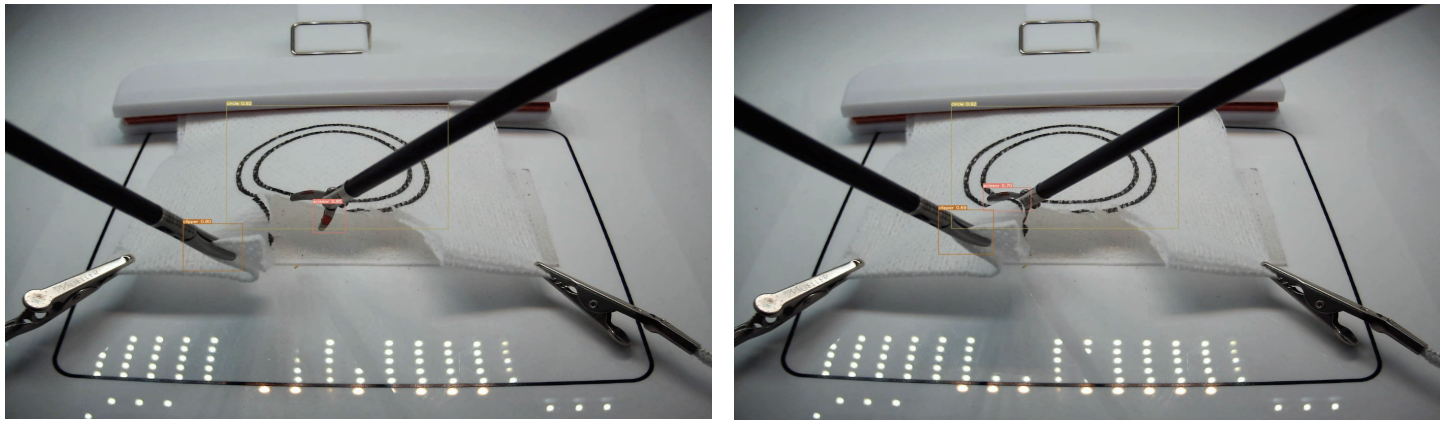


Figure 9: Qualitative results showing the detection of four test images.



(a) First camera tested video.



(b) Second camera tested video.

Figure 10: Qualitative results showing detection and tracking of laparoscopic instruments in two tested videos

One significant use of this model lies in the field of surgical education and performance evaluation. The output of the model, which includes images and videos, can serve as feedback for surgical performance or as a means of deliberate practice for cognitive behaviors during tests. Moreover, these results are accurate and can be utilized as input for further analysis.

5. Conclusions and Future Work

The aim of this study was to enhance laparoscopic surgical training and assessment by developing an extended dataset of instruments and objects for a box trainer pattern-cutting test and implementing a real-time object detection approach based on YOLOv7. Our findings demonstrate that our method effectively detects and tracks spatial tool and object movements and could be used to create a reliable real-time assessment system. Moving forward, we plan to integrate these results into a fuzzy logic decision support system to develop an automated GOALS assessment system.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors acknowledge the Department of General Surgery at Western Michigan University Homer Stryker M.D. School of Medicine for their assistance in producing the dataset used in this work.

Table 4: A comparison summary of the proposed model with the related reported approaches in the literature.

Approach	Dataset	Labeling Tool	Model	Accuracy %	Precision %	Speed
The proposed model	1572 images extracted from our lab dataset [10]	Roboflow [27]	Yolov7	78	mAP 95.2	In real-time at 83.3 FPS
Surgical tools detection based on modulated anchoring network in laparoscopic videos [12]	5696 extracted from m2cai16-tool-locations and AJU-Set datasets	The data already labeled	Faster R-CNN	69.6% and 76.5% for each dataset	mAP 69.6% and 76.5% for each dataset	not reported
Real-Time Surgical Tool Detection in Minimally Invasive Surgery Based on Attention-Guided Convolutional Neural Network [13]	4011 extracted from EndoVis Challenge, ATLAS Dione, and Cholec80-locations datasets	not reported	ResNet50 with multi-scale pyramid pooling.	not reported	100, 94.05, and 91.65 for each dataset	In real-time at 55.5 FPS
Deep learning based multi-label classification for surgical tool presence detection in laparoscopic videos [14]	29478	Pixel Annotation Tool36	EndoNet	not reported	mAP 63.36	not reported
Identifying surgical instruments in laparoscopy using deep learning instance segmentation [15]	333	not reported	Mask R-CNN	not reported	AP 81	not reported
Surgical-tools detection based on convolutional neural network in laparoscopic robot-assisted surgery [17]	M2CAI 2016 Challenge videos	not reported	YOLO	not reported	mAP 72.26	48.9 FPS
Robust real-time detection of laparoscopic instruments in robot surgery using convolutional neural networks with motion vector prediction [18]	7492 extracted from m2cai16-tool-locations dataset	not reported	YOLO9000	not reported	mAP 84.7,	38 FPS
Instrument Detection for the Intracorporeal Suturing Task in the Laparoscopic Box Trainer Using Single-stage object detectors [19]	900 images extracted from our lab dataset [10]	Roboflow [27]	YOLOv4, Scaled-YOLOv4, YOLOR, and YOLOX	not reported	mAP50 0.708 0.969 0.976 0.922 for each model	not reported
Surgical Skill Assessment System Using Fuzzy Logic in a Multi-Class Detection of Laparoscopic Box-Trainer Instruments [21]	950 images extracted from our lab dataset [10]	not reported	SSD ResNet50 V1 FPN and SSD Mobilenet V2 FPN	not reported	65% and 80% reliability, 70 and 90 of fidelity for each architecture	not reported
Surgical Instrument Detection Algorithm Based on Improved YOLOv7x. [23]	452	LabelImg	Improved YOLOV7x algorithm based on RepLK Block and ODConv	94.7	not reported	not reported

References

[1] K. Alkhamaiseh, J. Grantner, S. Shebrain and I. Abdel–Oader, "Towards Automated Performance Assessment for Laparoscopic Box Trainer using Cross-Stage Partial Network," 2021 Digital Image Computing: Techniques and Applications (DICTA), 2021, 01-07, doi: 10.1109/DICTA52665.2021.9647393.

[2] Wang, Chien-Yao, A. Bochkovskiy, and H. Liao., " YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," (2022), arXiv. https://doi.org/10.48550/arXiv.2207.02696

[3] "Fls homepage." [Online]. Available: https://www.flsprogram.org. ˜ Last accessed 2 Dec 2022.

[4] R. Aggarwal, T. Grantcharov, K. Moorthy, J. Hance, and A. Darzi, "A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill," American journal of surgery, **191**(1), 128—133, January 2006. [Online]. Available: https://doi.org/10.1016/j.amjsurg.2005.10.014

[5] G. Islam, K. Kahol, B. Li, M. Smith, and V. L. Patel, "Affordable, web-based surgical skill training and evaluation tool," Journal of Biomedical Informatics, **59**, 102–114, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046415002397

[6] G. A. Alonso-Silverio, F. P´erez-Escamirosa, R. Bruno-Sanchez, J. L. Ortiz-Simon, R. Mu˜noz-Guerrero, A. Minor-Martinez, and A. Alarc´on-Paredes, "Development of a laparoscopic box trainer based on open source hardware and artificial intelligence for objective assessment of surgical psychomotor skills," Surgical Innovation, **25**(4), 380–388, 2018, pMID: 29809097. [Online]. Available: https://doi.org/10.1177/1553350618777045

[7] N. J. Hogle, W. D. Widmann, A. O. Ude, M. A. Hardy, and D. L. Fowler, "Does training novices to criteria and does rapid acquisition of skills on laparoscopic simulators have predictive validity or are we just playing video games?" Journal of surgical education, **65**(6), 431–435, 2008.

[8] M. A. Zapf and M. B. Ujiki, "Surgical resident evaluations of portable laparoscopic box trainers incorporated into a simulation-based minimally invasive surgery curriculum," Surgical innovation, **22**(1), 83– 87, 2015.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning. nature 521 (7553), 436444," Google Scholar Google Scholar Cross Ref Cross Ref, 2015.

[10] https://drive.google.com/drive/folders/1F97CvN3GnLj-rqg1tk2rHu8x0J740DpC

[11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., "Imagenet large scale visual recognition challenge," International journal of computer vision, **115**(3), 211–252, 2015.

[12] B. Zhang, S. Wang, L. Dong, and P. Chen, "Surgical tools detection based on modulated anchoring network in laparoscopic videos," IEEE Access, 8, 23 748–23 758, 2020.

[13] Shi, P., Zhao, Z., Hu, S. and Chang, F., "Real-Time Surgical Tool Detection in Minimally Invasive Surgery Based on Attention-Guided Convolutional Neural Network," IEEE Access **8**, 228853-228862, 2020.

[14] S. Wang, A. Raju, and J. Huang, "Deep learning based multi-label classification for surgical tool presence detection in laparoscopic videos," in 2017 IEEE 14th International Symposium on Biomedical Imaging. IEEE, 620–623, 2017.

[15] S. Kletz, K. Schoeffmann, J. Benois-Pineau, and H. Husslein, "Identifying surgical instruments in laparoscopy using deep learning instance segmentation," in 2019 International Conference on Content-Based Multimedia Indexing (CBMI). IEEE, 1–6, 2019.

[16] E. Kurian, J. J. Kizhakethottam, and J. Mathew, "Deep learning based surgical workflow recognition from laparoscopic videos," in 2020 5th International Conference on Communication and Electronics Systems (ICCES). IEEE, 928–931, 2020.

[17] B. Choi, K. Jo, S. Choi, and J. Choi, "Surgical-tools detection based on convolutional neural network in laparoscopic robot-assisted surgery," in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 1756–1759, 2017.

[18] K. Jo, Y. Choi, J. Choi, and J. W. Chung, "Robust real-time detection of laparoscopic instruments in robot surgery using convolutional neural networks with motion vector prediction," Applied Sciences, **9**(14), 2865, 2019.

[19] M. Mohaidat, J.L. Grantner, S.A. Shebrain, and I. Abdel-Qader, " Mohaidat M, Grantner JL, Shebrain SA, Abdel-Qader I. Instrument detection for the intracorporeal suturing task in the laparoscopic box trainer using single-stage object detectors," In2022 IEEE International Conference on Electro Information Technology (eIT) May 19, 455-460, 2022.

[20] M. Mohaidat, J.L. Grantner, S.A. Shebrain, and I. Abdel-Qader, "Multi-Class Detection and Tracking of Intracorporeal SuturingInstruments in an FLS Laparoscopic Box Trainer Using Scaled-YOLOv4," In Proceedings of the Advances in Visual Computing:17th International Symposium, ISVC 2022, San Diego, CA, USA, 3–5 October 2022

[21] F.R. Fathabadi, J.L. Grantner, S.A. Shebrain, and I. Abdel-Qader, "Surgical skill assessment system using fuzzy logic in a multi-class detection of laparoscopic box-trainer instruments," In 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 1248-1253, 2021.

[22] F.R. Fathabadi, J.L. Grantner, S.A. Shebrain, and I. Abdel-Qader, "3D Autonomous Surgeon's Hand Movement Assessment Using a Cascaded Fuzzy Supervisor in Multi-Thread Video Processing," Sensors, 23, 2623, 2023. https://doi.org/10.3390/s23052623

[23] R. Boping, B. Huang, Sh. Liang, and Y. Hou, "Surgical Instrument Detection Algorithm Based on Improved YOLOv7x," Sensors (Basel, Switzerland) 23,11 5037. 24 May. 2023, doi:10.3390/s23115037

[24] https://pjreddie.com/darknet/

[25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 779–788, 2016.

[26] "School of medicine." [Online]. Available: https://med.wmich.edu/. ˜ Last accessed 2 Dec 2022.

[27] [Online]. Available: https://aroboflow.com/. Last accessed 2 Dec 2022.

[28] A. M. Obeso, J. Benois-Pineau, M. G. Vázquez, and A. R. Acosta, "Saliency-based selection of visual content for deep convolutional neural networks," Multimedia Tools and Applications, 78(8), 9553–9576, 2019.

[29] https://colab.research.google.com/drive/1X9A8odmK4k6l26NDviiT6dd6TgR-piOa#scrollTo=GD9gUQpaBxNa

[30] https://colab.research.google.com/signup\

[31] Keras image preprocessing. Available at: march 2015; accessed 3 november 2021.

[32] Googleresearch, "tensorflow: large-scale machine learning on heterogneous systems," google res., 2015.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 8, No. 6, 96-106 (2023)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

A Secure Medical History Card Powered by Blockchain Technology

Samiha Fairooz[1], Shakila Yeasmin Miti[2], Zihadul Islam[2], Meem Tasfia Zaman[*,2]

[1]*University of Adelaide, Faculty of Sciences, Engineering and Technology, Adelaide, South Australia*

[2]*North South University, Department of Electrical and Computer Engineering, Dhaka, Bangladesh*

ARTICLE INFO

Article history:
Received: 29 September, 2023
Accepted: 26 November, 2023
Online: 30 December, 2023

Keywords:
Medical history card
Web application
Blockchain
Decentralized database
Electronic medical records
Healthcare

ABSTRACT

A reliable healthcare system ensures that the population has access to top-notch medical services, ultimately enhancing their overall health most efficiently. At times, data are not secured or handled appropriately. Addressing these concerns, blockchain technology is projected to bring about a substantial revolution in the medical industry by assuring the confidentiality of electronic health information. This research not only seeks to rectify the shortcomings in Bangladesh's existing health system but also explores the potential of blockchain technology's decentralized database to fortify the entire healthcare framework. More importantly, it showcases a web-based application, particularly a medical history card that displays a patient's details, diagnoses,vaccines, medication records, investigation background, familial information, blood donation history, and many additional information starting from birth. Alongside, the paper emphasizes the transformative impact of implementing blockchain technology in the healthcare sector, paving the way for a more secure and efficient healthcare ecosystem. All in all, the array of medical information captured within the pack face of a single card could hasten medical decisions and ensure the effectiveness of any treatment.

1 Introduction

The health sector holds an exceedingly central position in safeguarding an overall sustainable socio-economic advancement. The health world is immensely diversified with the presence of several technical and biomedical analyses. The previous concepts of medical management are now found substituted over time. Currently, people are more focused on constituting a community of healthiness in which disease prevention outweighs the importance of the cure, unlike in the past. Notwithstanding that, the healthcare management system confronts the most critical challenges and barriers. Although technological and medicinal advancements are now quite prominent, no steps have yet been taken to incorporate the patient's medical history in a decentralized manner. The patients of private and public Healthcare Centers (HCs) are still found carrying the baggage of reports, diagnoses, and prescriptions while visiting a physician. Reports or prescriptions are either lost or misplaced, resulting in harassment and stress during emergencies. Medical health records are irreplaceable not only to patients but also to the hospital authorities themselves. A detailed history can put additional value to the current treatment by the medical consultants and physicians. The continuing development of the Internet of Medical Things (IoMT) could accelerate the maintenance of digital documents while offering greater convenient services to clients. Despite these precautions, healthcare frequently encounters virtual hacking incidents that end up resulting in the deletion and disclosure of personally identifiable information. Software vulnerabilities, security collapse, and human error are ordinarily the reasons behind this catastrophe. Amid such a situation, the data breach discloses many loopholes in the healthcare database system that can be encapsulated under three keynotes: 1) centralized aggregation of essential data, 2) unauthorized access to the public, and 3) disclosing confidential data purposely due to inadequate information security. Notable, health records need to be breach-proof as any data tampering can cause critical medical emergencies, including the death of patients. With the significance of medical history in mind, we are introducing a medical history card that would contain all a person's medical records under a platform. To explain, the patients will be emancipated from the burden of bearing all their medical history before visiting a physician. Furthermore, the likelihood of doctors encountering incorrect and deceptive information will decrease. The utilization of blockchain technology in the card ensures the decentralized accuracy of medical information, eliminating the need for third-party electronic storage. Blockchains are often referred to as distributed ledger technology (DLT) as blockchain records are immutable and cannot be modified, changed, or deleted.

[*]Corresponding Author: Meem Tasfia Zaman, North South University, Email: zaman.tasfia@northsouth.edu

2 Literature Review

Conventionally, the idea of digitally archiving medical information has traditionally been in development for a significant period of time. The current paper is dedicated to extending the paper entitled "A medical history card utilizing the Blockchain technology" which was published in the 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) [1]. In the past in 1973, The Regenstrief Medical Record System (RMR), a type of paper-based electronic record system has been utilized in outpatient healthcare facilities. The RMR obtained information about patients from a wide range of sources including physician chambers, forensic lab data, pharmacies, and others. It generated a summary report, surveillance report, and the Patient Encounter Form (PEF) to deliver protocol-based feedback to doctors about critical clinical conditions. It was extensively believed that the widespread use of EMRs had decreased surgical errors, boosted health, and decreased expenditures on healthcare. Likewise, the organizational processes have ensured excellent quality, effectiveness, and transparency through EHRs [2]. In order to create an integrated medical record between patients and medical professionals, the author of [3] (2008) established a patient portal called CONNECT (Care Online: Novel Networks to Enhance Communication and Treatment). The patients could easily access it from home, hospital, or doctor's office using the heterogeneous networks. CONNECT application was committed to improving patient-provider communication and assisting patients in understanding and managing their illnesses. In 2010, author of the paper [4] described that the use of EMRs in long-term care institutions at the time was neither widespread nor successful. The hospitals implementing EMRs were working on techniques to evaluate effectiveness; no statistical information was available on the efficacy of EMRs in the long-term care context. They opposed the widespread use of EMRs. Conversely, it was evident that when most patients carry their EHR, many chances were created for the industries. For instance, a "health record bank," proposed in [5], was a starting point for the enterprise's additional "value-added" services. An EHRs facilitates clinicians to view patients' data from numerous sources across a period for extensiveness and continuity of care. Research showed that the appropriate application of the EHRs could minimize medical errors, assist in diagnosing detrimental health events, facilitate the more appropriate use of healthcare services, and potentially reduce healthcare expenses [6]. An e-prescription system that assists mother and children with medicine data recording, retrieving, and reporting techniques have been developed by the author of [7]. Moreover, the possible medication errors were reduced by upgrading the prescribing process. In [8], the author claimed that the Implantable Device Cardiac Observation (IDCO) profile had been constructed at the Leiden University Medical Centre (LUMC) to integrate data from the external databases maintained by two device providers into the departmental Cardiology Information System. They continued by saying that the growing use of ICDs and remote monitoring at treatment stations will improve patient outcomes by lessening the burden of follow-up on clinics and personnel. From the patient's point of view, it was appealing due to better security and prevention of lengthy and time-consuming visits to the hospital. In particular, readmissions were minimized when patient history was viewed [9]. In the publication with reference [10], the author covered, The Internet of Things (IoT) and the digitization of medical records topics in a number of their presentations. They made portable web servers out of smart gadgets like smart discs and cards, making remote diagnostics record-keeping easier. The functional requirements call for a device-enabled system that could obtain and capture all the records with a backup on a few cloud-based platforms so that data is not lost, even though the device was lost. The system claimed to eliminate paper from patient transactions completely. Besides, storing patient reports, histories, visits, prescriptions, results, pharmacy bills, and emergency contact information was more straightforward. The patient had exclusive access to the gadget, and adequate verification was required for updating any information on the device. Additionally, in the same year, an advanced decentralized system, MedRec for managing Electronic Health Records (EHRs) through Blockchain technology was proposed. Their solution offered patients a comprehensive, tamper-proof log and easy access to their data [11]. In 2008, the basic concept of a distributed secure ledger was introduced. Blockchain technology is the approach to creating an immutable, secure, distributed database of transactions. Blockchains were trained to deliver a distributed ledger of financial transactions, not relying upon a central bank, credit company, or other financial institution. Additionally, it confirms a secure and permanent record of transactions. The significant progress in EMRs was the generation of a distributed ledger to productively transfer patients' records from the healthcare organization to the individual. One of the key features of blockchain technology is that it is way too transparent: transactions are processed by the network without requiring a single computer, database, or institution. Besides, the paper [12] offered insights into the use of blockchain in healthcare data management, with a specific focus on sharing EMRs among healthcare providers. Collaborating with Stony Brook University Hospital, a framework was developed for managing and sharing Electronic Medical Records (EMRs) of cancer patients. This framework involved connecting to the hospital's local database management system, which was specifically designed for storing oncology-related data. Patient data was then encrypted using each patient's unique key and stored on a cloud-based platform known as Varian Cloud. Additionally, a blockchain-based data-sharing system was proposed as part of this initiative. This system aimed to leverage the immutability and autonomy characteristics of blockchain technology to effectively address various medical challenges related to data security and sharing [13], [14]. In 2018, a blockchain-based data-sharing system called BPDS (Blockchain-based Privacy-Preserving Data Sharing) was introduced with the primary goal of safeguarding user privacy. By implementing BPDS, patients gained control over their Electronic Medical Records (EMRs), and both users and institutions could access this data without concerns about compromising patient privacy. In BPDS, the original EMRs were securely stored in the cloud-based BPDS system, while the indexes were maintained within a consortium that had stringent tamper-proof measures. This approach significantly reduced the risk of medical data leakage. During the same year, a novel method involving attribute-based signatures, with multiple authorities, was developed to ensure the authenticity of Electronic Health Records (EHRs) embedded within the blockchain. In this method, a patient could sign a message based on certain

attributes without revealing any additional personal information, thus providing proof of their attestation [15] [16]. To enhance privacy within the system, a method involving Personalized Radio Frequency Identification (RFID) cards was implemented. These RFID cards assigned unique IDs to different users, and modifications to these IDs could only be made by the relevant doctors using the medical database. This approach ensured the authentication of medical information, as only authorized personnel could manage and update the data. Furthermore, another blockchain-based approach was suggested to secure Electronic Medical Records (EMRs) for healthcare applications. In this approach, access control was patient-centric, meaning that patients would share the decryption key only with trusted doctors. These trusted doctors would then store the patient's encrypted electronic medical information securely on a blockchain, adding an additional layer of security to the handling of medical data. By utilizing the encryption technology, the authors of the publication [17] proposed a Sensitive and Energetic Access Control (SE-AC) mechanism for certifying fine-grained confidentiality of the patient's Electronic Health Record (EHR), wherein people given authorization could update or examine certain EHR. The authors emphasized that the security of EHRs seriously threatens the patient's privacy, and most third-party hosting platforms have some issues with data security and user privacy. Individuals having drug or alcohol addiction histories and HIV could request primary care physicians to access certain documents in the hospital [18] [19] [20]. The concept of remote healthcare was introduced with a focus on securing electronic records through encryption before they are transmitted to medical Blockchain networks. In this approach, the medical team members established a session key for authorized groups in a way that allowed authorized participants to decrypt and access patient information. Importantly, the content could only be deciphered using either the patient's private key or a group member's session key, and no personal identification password was involved in the entire process. In [21], the authors discussed the use of 'The InterPlanetary File System (IPFS)' for storing Electronic Medical Records (EMRs) and the Ethereum platform for replicating the blockchain. This proposed method enhanced the confidentiality of uploaded data, reduced the risk of errors due to data manipulation, and facilitated the secure digitization of medical records. The public key served as a universal identifier to segregate and maintain the complete medical record history for each patient [22]. In 2021, a comprehensive healthcare system was developed, encompassing an Android app, an iOS app, and a website, all designed to grant patients easy access to their medical records and history. Additionally, a novel feature was introduced: a health passport designed to compile and display patients' medical data from various healthcare providers. This encompassed a wide range of information, including medical images, clinical reports, lab results, dental records, and more. Users had the convenience of utilizing their smartphones to consolidate their medical history from diverse sources, including healthcare providers, imaging CDs, PDF notes, fitness trackers like Fitbit and Apple Watch, wellness forms, CDA Files, JPEG documents, and self-reported health records. Furthermore, they could maintain and update their health vitals such as Blood Pressure, Heart rate, height, weight, BMI, Body Temperature, Oxygen saturation, allergies, and BSA. Additionally, the application provided a user-friendly feature that enabled individuals to easily locate their nearest healthcare center and schedule appointments, making healthcare management more accessible and efficient [23]. A new architecture termed as Internet-of-Health-Care-Systems (IoHCS) was launched to supervise the EHRs accumulated from all partaking hospitals in a network; through Blockchain Technology developed on the Ethereum platform. Software agents via Message Queueing Telemetry Transport (MQTT) protocol were used to network the Health Information Systems (HIS) of all linked hospitals. A mobile app for accessing patient records using a Key Management System (KMS) and an Application Programming Interface (API) for reading and writing data for each hospital on the blockchain was also generated. Studies show that the prevailing HIS of 350 partaking hospitals was incorporated into the app. The EHRs of the patient have been formulated in 6 parts — personal information, visit information, laboratory test information, drug order information, and a hash value created by the encryption module. Amazon Key Management System and Amazon S3 (Simple Storage Service) guarantee data and process security. In 2022, a new web-based project was proposed with the aim of creating a centralized database to store patients' medical histories. The paper provided details about the software and technological architecture employed for this purpose. The project focused on enabling remote control and management of the integrity and documentation of medical histories, reports, and prescriptions. For the backend of the online application, PHP, AJAX, and jQuery were utilized, with MySQL serving as the database. On the frontend, HTML5, CSS, Bootstrap, and JavaScript were employed. Researchers developed a prototype platform that aimed to digitalize the healthcare system by offering digital appointments, artificial intelligence-driven recommendations, and a health-tracking blood bank. One notable challenge addressed in the context of the technological revolution was the storage of segmented Electronic Health Records (EHRs) in Healthcare Centers (HCs). To tackle this issue, a platform was introduced to establish a distributed electronic health record (EHR) ecosystem. This ecosystem aimed to integrate electronic medical reports securely within a private and permissioned blockchain, addressing the fragmentation of EHR data. The notable advantages of implementing the distributed network regarded clinical outcomes such as improved quality, and reduced medical errors; organizational outcomes like financial, and operational benefits; managerial outcomes, e.g., improved ability to conduct research, improved population health, and lower costs [24], [25]. To highlight, Blockchain is an emerging decentralized technology to validate and secure a patient's medical record efficiently. Several attempts were proposed to use Blockchain technology to address healthcare data privacy, security, and ownership issues. Blockchain-based digital platforms empower patients to communicate with their data providers, including healthcare providers, in a faultless, secure, and efficient approach [26], [27].

3 Features

Since the inception of the paper, the goal has been to deploy the history card for the comprehensive collection of an individual's crucial medical records to facilitate appropriate treatment

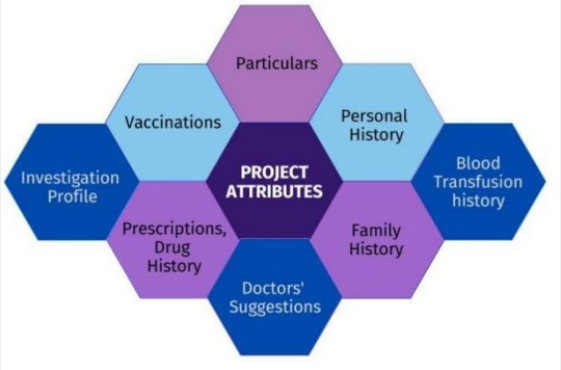


Figure 1: Features Diagram

To ensure the card's effective and optimized application, various features have been integrated, in Figure 1. Initially, 'Particulars' will contain the Name, Date of birth, blood group, parent's name and details, weight, height, classification of gender, address, and contact number. The 'Personal History' section provides comprehensive information concerning past allergies, sinusitis, any congenital anomalies, diabetes, Claustrophobia, thyroid conditions (hypothyroidism or hyperthyroidism), etc. Another vital segment is the 'Investigation Profile,' which grants access to all diagnostic reports, including cellular and chemical analyses (such as blood analysis, glucose tolerance test, serological test, enzyme analysis, kidney and liver function, protein-bound iodine test, gastric fluid analysis, and more); diagnostic imaging (such as mammography, ultrasonogram, brain scanning, and so forth); health examinations (like biopsy, laparoscopy, and others). Following that, the 'Vaccinations' section will contain a record of all vaccinations administered from birth onwards. Subsequently, the 'Drug History' segment provides a catalog of medication names, dosages, and the timing of each medication taken. In the 'Blood Transfusion History,' the record includes details about the quantity and blood type of blood donations made by the patient. Similarly, in the 'Prescriptions' section, all prescriptions provided by doctors are organized chronologically. Another essential component is the 'Family History,' which encompasses the genetic diseases present in the patient's family, such as Congenital deafness, Cancer, thyroid concerns, and so on. Finally, 'Doctors Suggestions' will outline the healthcare recommendations given by doctors regarding the patient's medical conditions.

4 Proposed Strategical Plan

The paper initially intends to use the health card to collect necessary information in order to eliminate the need for patients to repeatedly visit the history room next to the doctor's office. In summary, there are four sorts of users in this article's web application: patients, doctors, vaccine staff, and report staff. The priority of patient's health history has been analyzed as per the priority list of records in the card.

Following that, the attributes of the website such as Patients (Name, age, blood group, gender, etc..), Vaccination History, Personal History (Diabetics, High / Low blood pressure, thyroid, history of surgery, blood transfusion, allergy, congenital anomaly, Covid-19, etc.), Diagnosis Report, Prescriptions, Drug History, Family History (Kidney, Heart Disease, Cancer, Skin disease, etc.) and Doctors' Suggestions schemed.

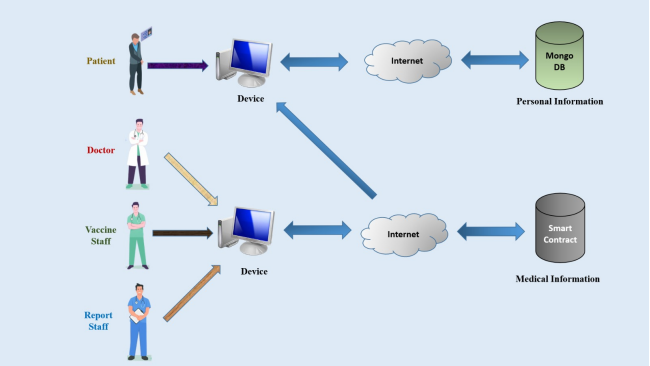


Figure 2: Network Diagram

The Network diagram depicts the fundamental work processes of the users for circumstances where accessibility to health records is prohibited for security reasons. (Figure 2). All of the patient's information can be read by both patients and doctors, but only they have the ability to make specific changes. The two more users, the vaccine and report staff are only permitted to look at the patient's personal background and relevant fields. Moreover, patients might view the automated EMR snapshots and update their personal information. Furthermore, MongoDB, a NoSQL database, was used to save the users' other traditional information, such as personal narratives, via the Internet, and the blockchain Smart Contract was chosen to store the medical data because there is always the risk of vital medical information being lost.

Smart Contract has been deployed as Blockchain Database to accommodate the patient's medical health history. To add, Smart Contract is a form of Ethereum account capable of sending transactions over the network. The agreement enables the execution of a contract between two parties through blockchain, without the involvement of any legal system.

5 Implementation

In this section, the web application's tools, frameworks and databases as well as the benefits of using these methods over more conventional methods. In addition, we shed light on the robust security measures and policies that have been put in place to safeguard the system.

5.1 Framework and Structural Design

The web application has been constructed using the MERN stack, which seamlessly combines four core technologies: React, and Node, MongoDB and Express. Furthermore, the application adheres to an architecture model that is built with 3-tier and it is characterized by three distinct layers: the front-end tier, represented by the web interface; the middle tier, composed of servers; and the back-end tier, covering the databases.

5.2 Approach to User Interface Development

To begin, the front-end infrastructure was constructed with HTML, CSS, React and JavaScript (JS). React is preferred for the project as it provides a front-end build pipeline making it useable with any back-end. Secondly, in order to develop dynamic HTML sites for taking user input into account and preserving permanent data using specialized objects, files, and relational databases, JS was utilized. Thirdly using CSS, the website's appearance and layout were updated.

5.3 Approach to Server-Side Development

The back-end server was developed with Node JS. In order to establish a connection between the front-end and the back-end, the Web3 application programming interface has been seamlessly integrated into the system.

During authentication, JSON Web Token (JWT) has been implemented for security purpose. JWT has been attached as "bearer" header on each login request to server. After that, it will be parsed by the server from header and the client identity will be verified. The JWT has been implemented because this authentication system is helpful to use for different type of users. To generate the JWT during every login HS256 signing algorithm has been used in the header that has been given below:

{
"alg": "HS256",
"typ": "JWT"
}

And the payload part of the JWT contains the claims of the particular users. The claims includes statements about the user and additional data which looks like this:

{
"hid": "123456789",
"exp": '5th June, 2023'
"iat": "2023-12:12 12:20:00"
}

Similarly, the refresh JWT is also being set as cookie so that, while the login access JWT will be invalidated the server could take set cookie of the refresh JWT and can generate a new refresh token for the logged in user. In the whole process, the expiration time of the access JWT has been set for 15 minutes and the refresh JWT has been set for 60 days.

In Figure 3, the displayed code snippet from our web application represents the implementation of the login process using JWT. As well as, specific validations has been incorporated to enhance security measures. Furthermore, to mitigate the risk of database injection, the use of the 'sanitize-html' package has been implemented.

5.4 Decentralized and Prominent Databases

The information will be stored in two separate databases. MongoDB will handle the orthodox data, while Smart Contract will store significant EMRs to enhance long-term protection. Additionally, MetaMask has been incorporated into the system as a wallet for overseeing and managing expenses related to ether.

```
// Login route to generate and return access and refresh tokens
app.post( path: '/clientLogin',
    handlers: (req :Request<P, ResBody, ReqBody, ReqQuery, LocalsObj>, res :Response<ResBody, LocalsObj> ) => {
    const { hid, password } = req.body;

    // Find the user based on the provided health ID and password
    const user = users.find((u :{...} ) => u.hid === hid && u.password === password);

    if (!user) {
        return res.status( code: 401).json( body: { message: 'Authentication failed' });
    }

    // Generate an access token with user information, exp, iat, and a secret key using HS256 algorithm
    const accessToken = jwt.sign( payload: {
        hid: user.hid,
        exp: Math.floor( x: Date.now() / 1000) + (15 * 60), // 15 minutes from now
        iat: Math.floor( x: Date.now() / 1000), // Issued at the current time
    }, secretKey,
        options: { algorithm: 'HS256' },
        callback: function(err, accessToken) {console.log(accessToken);});

    // Generate a refresh token with user information and a refresh secret key using HS256 algorithm
    const refreshToken = jwt.sign( payload: {
            hid: user.hid,
            exp: Math.floor( x: Date.now() / 1000) + (60 * 24 * 60 * 60), // 60 days from now
            iat: Math.floor( x: Date.now() / 1000),
        },
        refreshSecretKey,
        options: { expiresIn: '60d', algorithm: 'HS256' },
        callback: function(err, refreshToken) {console.log(refreshToken);});

    // Set the refresh token in a cookie with a secure flag and HttpOnly to make it more secure
    res.cookie( name: 'refreshToken', refreshToken, options: {
        maxAge: 60 * 24 * 60 * 60 * 1000, // 60 days
        httpOnly: true,
        secure: true,
        sameSite: 'none',
    });

    // Return the access token to the client
    res.json( body: {message: "Successfully Logged In", access_token : accessToken });
});
```

Figure 3: Code Snippet Illustrating User Authentication

```
ClientInfo.sol 1 ×   {} ClientInfo.json
contracts > ClientInfo.sol
1  // SPDX-License-Identifier: MIT
2  pragma solidity >=0.4.22 <0.9.0;
3
4  contract ClientInfo {
5      // Prescription Contract Start here.
6      uint256 public prescriptionCount = 0;
7
8      struct prescriptionStruct {
9          uint256 id;
10         string Dfirstname;
11         string Dlastname;
12         string Docregid;
13         string hospitalname;
14         string src;
15         string Time;
16         string hid;
17     }
18
```

Figure 4: Medical Information's Data Structure

Patient's medical data is kept in an Ethereum Smart Contract Database. Each specific category of medical information has its own data structure (Figure 3).

The patient's whole medical history was then entered into the smart contract using the set method. The data was lastly acquired using the get method (Figure 5). The EMRs are securely stored within a decentralized Blockchain database. The decision to implement blockchain technology was primarily driven by two key objectives: transparency and data integrity. Specifically, blockchain was chosen to ensure that both patients and doctors cannot conceal information and to prevent the falsification of critical medical data. The blockchain architecture operates as a decentralized digital ledger, recording distinct data or transactions across a network of computers in blocks. Each node within this network possesses a complete copy of the blockchain, ensuring security and transparency without reliance on a central authority. Characterized by decentralization

and distribution, blockchain networks facilitate secure peer-to-peer transactions. Blocks, comprised of validated transactions, contain a cryptographic hash of the preceding block. The database's security is enhanced through decentralization, immutability, consensus mechanisms, cryptographic security, transparency, redundancy, and controlled access.

```
function setPrescriptionData(
    string memory _Dfirstname,
    string memory _Dlastname,
    string memory _Docregid,
    string memory _hospitalname,
    string memory _src,
    string memory _Time,
    string memory _hid
) public {
    prescriptionCount++;
    newPrescriptionStruct[prescriptionCount] = prescriptionStruct(
        prescriptionCount,
        _Dfirstname,
        _Dlastname,
        _Docregid,
        _hospitalname,
        _src,
        _Time,
        _hid
    );
}

function getPrescriptionDataCount() public view returns (uint256) {
    return (prescriptionCount);
}

function getPrescriptionData(uint256 dataNum)
    public
    view
    returns (
        uint256,
        string memory,
        string memory,
        string memory,
        string memory,
        string memory,
        string memory,
        string memory
    )
{
    prescriptionStruct memory s = newPrescriptionStruct[dataNum];

    return (
        s.id,
        s.Dfirstname,
        s.Dlastname,
        s.Docregid,
        s.hospitalname,
        s.src,
        s.Time,
        s.hid
    );
}
```

Figure 5: Set and Get Method of Medical Information

An object-oriented, high-level Solidity language was used for implementing Smart Contracts. This language is designed to target the Ethereum Virtual Machine (EVM). Truffle, the development environment, asset pipeline, and testing framework for smart contracts was used for the framework. Smart contracts, functioning within these networks, execute agreements or transactions without intermediaries, using platform-specific languages like Solidity for Ethereum. Once deployed on the blockchain, Smart Contracts become immutable code, and decentralized verification prevents any single entity from controlling their execution. This technology holds transformative potential for revolutionizing medical history management by securely storing records, providing patient-controlled access, and triggering actions based on predefined conditions. The project utilized MetaMask as a wallet to purchase ether. To add, MetaMask is a free web crypto wallet that allows users to store and connect with the Ethereum blockchain ecosystem. Lastly, to host the Blockchain server locally Ganache was deployed to set up a local Ethereum blockchain to test the decentralized application in a safe environment. Additionally, the Blockchain design eliminates the risk of data loss through its distributed architecture, where each server on the network maintains a copy of the data, and no server has the capability to alter, amend, or delete any previously recorded information from the medical records. This combination of transparency and data immutability establishes a robust and secure foundation for the management and protection of sensitive medical information within the system.

6 Outcome and Prospects

In this section, the output of the web application has been demonstrated in order to get a clearer web experience. More especially, with the use of screenshots from the web application, the actions of the four different user types— doctors, patients, the vaccine team, and the reporting team—are elucidated.

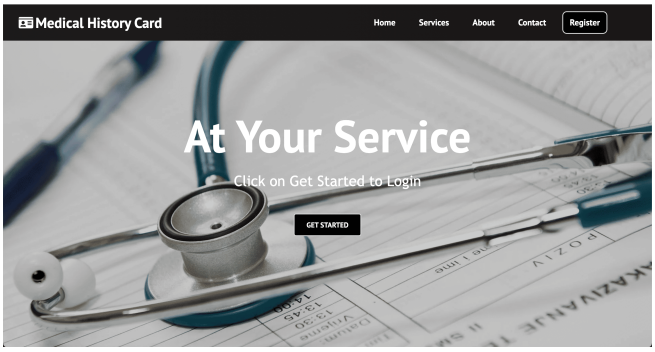


Figure 6: Home Page of the Website

The MHC starts with a Home page. This on boarding page will help the users learn how to get started and derive value from it (Figure 6). The home page contains a Navigation bar at the top to direct to a particular sector as per the user's requirements.

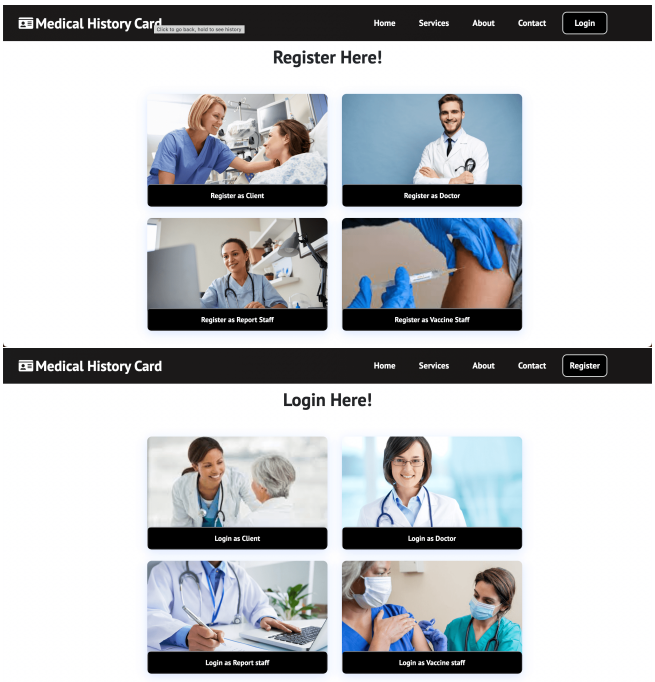


Figure 7: Registration and Login Dashboard

The registration and login pages contain four individual cards designed for each of our four users (Fig. 7). In the registration process, both the birth identity (BID) and national identification (NID) are collected as input, as the web application allows patients under the age of 18 to create accounts. In such cases, the BID becomes a mandatory requirement. After that, a success message will be received as well as the client could see the automatically generated HID in the message. Likewise, users can log in with the help of the login page after registration.

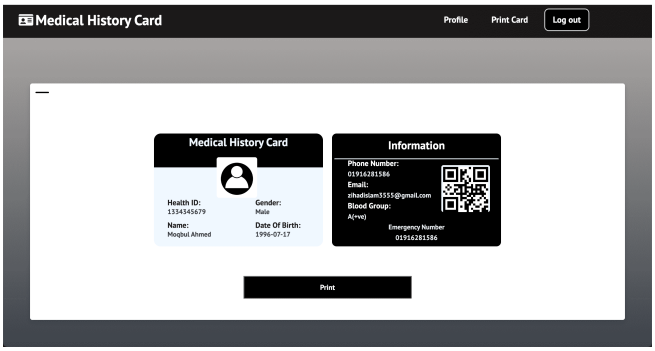


Figure 8: A Demo of the Medical History Card

The inclusion of a unique HID and QR code for each patient card adds an extra layer of convenience and security to the web application for both doctors and patients. It also adds an extra layer of security by ensuring that only authorized personnel (doctors and the respective patients) can access and view the associated medical records. Additionally, patients have the option to download and print their cards by clicking the ”Print” button within the website’s application, as illustrated in Figure 8.

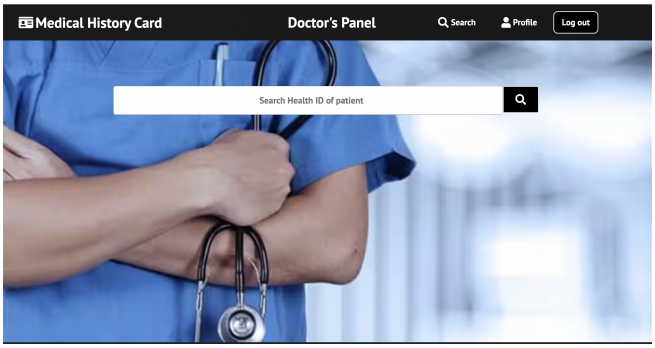


Figure 9: Doctor’s Searching Panel

The doctor can then search for a patient in the doctor searching panel by entering the patient’s HID and will be able to view his medical information (Figure 9). In a similar manner, the staff members who work with vaccines and reports have access to search. However, the patient’s medical information has been limited in accordance with the roles.

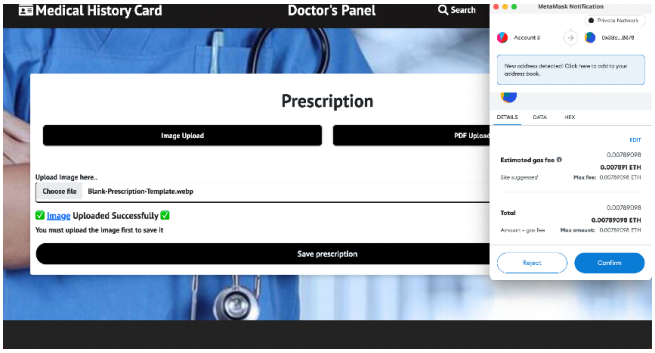


Figure 10: A Doctor Uploading a Patient’s Prescription

Doctors have the flexibility to upload patient prescriptions which prompts them to select the prescription file they want to upload. They can choose either an image file (e.g., JPG, PNG) or a PDF document, as shown in Figure 10.

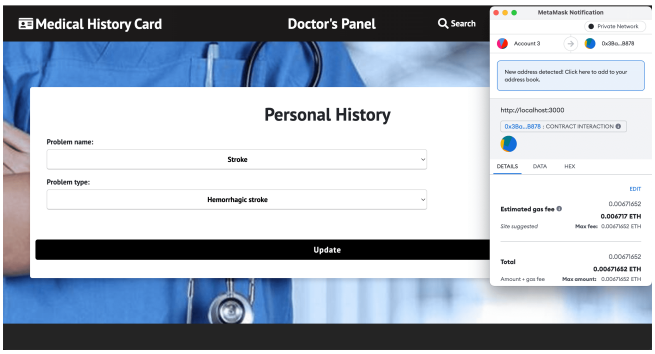


Figure 11: A Doctor Submitting Personal Health History of Patient

Moreover, they have the option to upload a patient’s personal medical background by picking the issue’s name and category from a drop-down menu, input familial medical histories by selecting the disease name and the family member’s relationship from a separate drop-down menu, and offer recommendations by entering information into a designated text field (Figure 11).

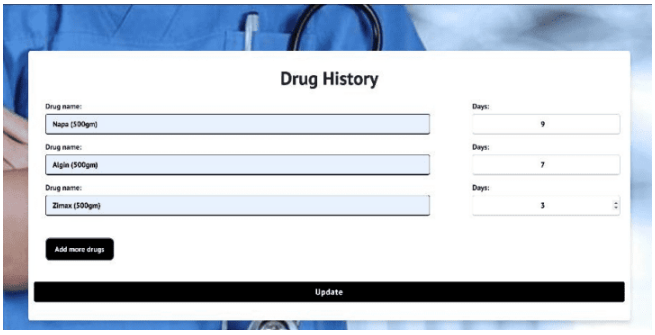


Figure 12: A Doctor Submitting a Patient’s Medication History

In the web application’s ’Drug History’ section, physicians can make updates to a patient’s prior medication record. To accomplish this, they need to enter the medication’s name manually into the text area and specify the duration of consumption by using the increment counter located on the right, as depicted in Figure 12.

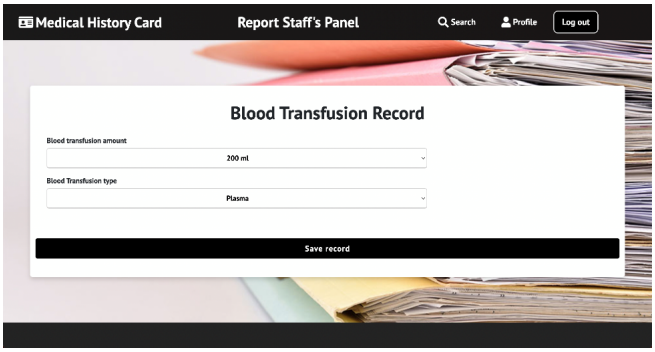


Figure 13: A Report Staff Uploading Blood Transfusion Record

On the other hand, the personnel responsible for managing vaccines and reports have the capability to not just review personal histories but also to make updates to vaccination histories and reports as needed. The "Blood Transfusion Record" might also be synced by the report staff, as shown in Figure 13. In this case, the drop-down boxes are utilized to choose the quantity and type of transfused blood.

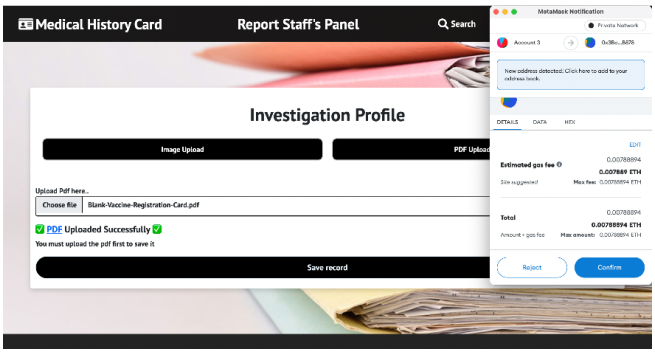


Figure 14: A Report Staff Uploading Patient's Investigation Profile

The report staff is also authorized to update a patient's "Investigation Profile," and the functionality for this task is similar to that of the prescription page. They have the ability to upload reports in both image and PDF file formats, providing flexibility in how medical information is recorded and shared.

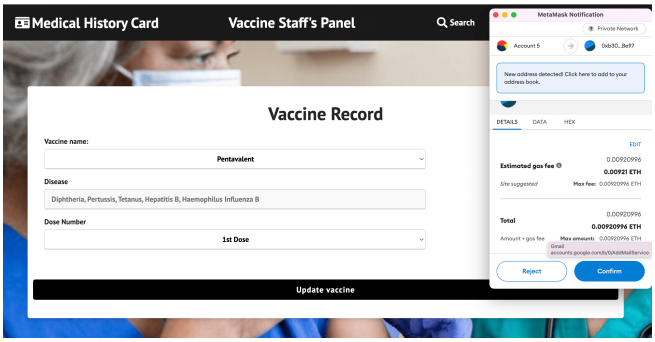


Figure 15: A Vaccine Staff Uploading Vaccination Record

It's also notable that, when updating vaccination histories, the vaccine staff will automatically see the name of the disease for which the vaccine is administered after selecting the vaccine's name from the drop-down box. Afterward, they must also select the quantity of vaccine doses (Figure 14).

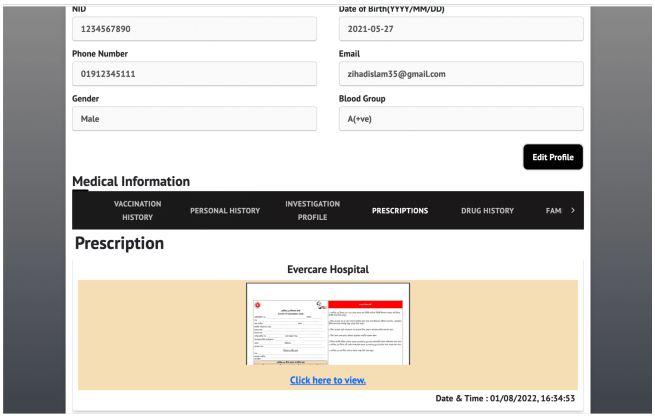


Figure 16: A Patient Viewing His Medical Information

Subsequently, patients are unable to upload their own data, but they can readily access their health-related information through the central navigation bar (as shown in Figure 15). Moreover, for enhanced monitoring, individuals have the option to download or print the medical reports and prescriptions, as well as capture screenshots of their medication history, personal health history, and vaccine history.

7 Contribution

The aforementioned card plays a crucial role in enhancing the security of patients' important medical data and preventing unauthorized access and manipulation of information. Along with preserving the authenticity of those high-priority data, this card can assist users during life-threatening conditions. The details enlisted in Personal History are of immense importance in the medical world. People with diabetes, high blood pressure, allergies, sinusitis, or asthma cannot afford every treatment or medicine. For this reason, doctors always ask these notable queries during consulting patients. If the patients are unaware of these facts, they are asked to undergo tests immediately. Besides, surgery history and thyroid (either hypothyroid or hyperthyroid) are also essential health indicators. During emergencies, the particulars provide essential information about a person. Specifically, the blood group and emergency contact number can significantly help during critical accidents. Concurrently, the investigation profile will remove the burden of carrying out reports, making diagnosis analysis easier for the doctors. They can glance through the reports to check whether the patient has any significant problems. To add, they can also check the severity of the disorder at that moment, which would instantly give them a more advanced view of the patient's present condition. They would understand if the present illness is somehow related to the previous history or not. Furthermore, people at a time tend to forget their vaccinations and the number of doses taken. This might create confusion in adulthood. Hence, this card can remove the hassle of memorizing all the vaccines. Drug history and prescriptions are the only ways to know about past treatments for any disease. With this card, the patients will no longer lose these valuable records while getting

treated. Consequently, they will get proper treatment on time. Congenital anomalies or birth defects develop prenatally or may be diagnosed later in life. These records help assess whether a disease is by birth. Afterward, the card will assist doctors in becoming familiar with the genetic disorders of a patient in such a way that they can recommend effective ways to reduce the risks. In the same way, blood transfusion history makes the task of knowing about the transfusion date and amount easier. So, patients can refrain from confirming any hypothetical dates to the medical officers.

It is worth mentioning that, Blockchain technology addresses healthcare security challenges in a broader range through its decentralized and tamper-resistant nature. Advanced encryption techniques and hash functions contribute to securing data within each block. The decentralized structure, distributed across a network of nodes, minimizes the risk of unauthorized manipulation. Smart contracts automate and enforce security protocols, reducing reliance on centralized databases and enhancing overall data protection. Above all, Blockchain presents a comprehensive solution to healthcare security by combining transparency, integrity, and patient-centric control in the management of sensitive health data.

Finally, advice given by doctors will be enlisted in this card. Suppose any patient has consulted two doctors for any particular disease over a time interval, and the former doctor has suggested something. Viewing the records allows the latter to analyze if it worked for the patient. Besides, he can also add his suggestions for future reference.

8 Observation

Between 1973 and 2023, there has been a significant evolution in the field of medical records. In the past, people relied on paper-based records, which were susceptible to data loss and damage, including the risk of being destroyed by fire or other accidents. As in [28] (2011) the authors have pointed out, EMRs have the potential to reduce diagnostic errors through various mechanisms. In a similar vein, we have structured our EMR system by categorizing information, aiming to create a more allocated, reliable, and information-rich system while maintaining brevity and efficiency. Furthermore, our web application is designed to alleviate the workload of hospital medical professionals and support staff.

Later, in 2016, the authors of the paper as refered [10] shed light on different aspects concerning the IoT and the digitalization of healthcare data. They emphasized that IoT mobile devices faced numerous challenges, including problems with integrating data, ensuring security and privacy, handling large volumes of data, maintaining performance, offering flexibility, dealing with the abundance of applications, and managing device diversity and interoperability. On the other hand, Blockchain databases were recognized for their flexibility and security. Additionally, our web application based on card technology was considered more convenient than relying on a device-oriented server.

Subsequently, we encountered web software which is stored information within a central database. However, in today's digital landscape, privacy issues have become prevalent, and concerns regarding hacking and data tampering with centralized databases have grown. Consequently, to enhance safety and visibility, we made the decision to adopt a decentralized database, specifically utilizing Blockchain technology. Moreover, we came across numerous applications that allowed patients to input their reports and other healthcare records. Nevertheless, on MHC, this information is automatically kept up to date by healthcare professionals and other affiliated staff members.

Following that, in [12] the authors provided valuable insights into the management of healthcare data utilizing blockchain technology, with a particular emphasis on sharing EMRs among medical providers. These researchers developed a prototype framework that aimed to guarantee privacy, security, availability, and precise access control over EMR data. It should be noted, however, that their framework had limitations, as it was designed exclusively for managing data related to cancer patients. In contrast, our paper extends this concept to encompass all patients, providing a more comprehensive approach to healthcare data management.

Next, in 2018, the researchers in [15] introduced an innovative attribute-based signature system that involved the participation of multiple authorities to authenticate the validity of EMRs integrated into the Blockchain. In this method, a patient signs a message using a specific attribute, revealing no extra information except for the verification that they have indeed endorsed it. It's important to emphasize that this signature method is rather complex and involves a higher associated cost. Conversely, our project is designed to be cost-effective, ensuring that it remains affordable for every citizen of a nation.

In summary, the Blockchain-based MHC system offers a solution where neither patients nor doctors can conceal or manipulate information, as records cannot be removed or altered. This innovative approach ensures the immutability of data, effectively putting an end to unreliable and erroneous information. The decentralized and distributed nature of Blockchain introduces a substantial level of complexity when it comes to any efforts to manipulate stored data, although it is not entirely impervious to such attempts. Moreover, the risk of data loss is eradicated because the servers of the network preserve a copy of the data.

Blockchain-based medical records systems in live healthcare environments face challenges such as user adoption, infrastructure integration, and regulatory compliance. These include complexity, resistance to change, compatibility with legacy systems, scalability, and compliance with data standards. Healthcare systems are highly regulated, and adhering to these regulations is crucial for patient data security. Additionally, ensuring interoperability and compliance with data standards across different healthcare providers' systems is essential. Initial investment in infrastructure, development, and training can be challenging, and resource allocation can be ongoing. To address these challenges, collaboration and education, pilot projects, interoperability standards, partnerships, and continuous evaluation and improvement are essential. A strategic approach, collaboration between stakeholders, and a clear understanding of technical and regulatory landscapes are necessary for successful implementation in a live healthcare environment.

The MHC system is not only cost-free but also more optimized than relying on any gadget-based server. Moreover, it is readily accessible to every citizen which makes it a valuable addition to the tech improvements of the medical field.

9 Upcoming Work

The medical history card is dedicated to advancing healthcare technology through the utilization of Blockchain technology. Its ultimate objective is to guarantee the safety and credibility of medical data. To enhance its use during emergencies, microchips will be incorporated in the future. This electronic, intelligent microchip-based card will store all patient data, enabling doctors to quickly access medical histories by scanning the card. Additionally, the implementation of biometric fingerprints will further enhance user authentication, providing significant benefits even for unconscious patients who may not have a physical card. Furthermore, MHC could serve as evidence of health insurance by storing records within a decentralized database. Medical history is essential in medical insurance, guiding risk assessment, underwriting, and claim validation. It informs insurers about health risks, aids policy customization, and verifies the legitimacy of treatments. Customizing policies based on health profiles and using medical history in actuarial analysis enhances overall insurance efficiency, ensuring tailored coverage and effective resource management. In essence, MHC is a cornerstone in shaping informed decisions throughout the insurance process.

In terms of family history, it would automatically establish a family history by cross-referencing an individual's HID with that of their parents. That said, the card that collects a patient's thorough history from birth will become inundated with a substantial amount of information, posing challenges and requiring a significant amount of time for doctors to navigate through. Therefore, several AI recommendations for assisting doctors would be added. When a patient visits a doctor, the system will prioritize displaying relevant information about the doctor's field of specialization. As an example, a cardiologist will be able to go through the patient's analysis reports, prescribed medicines, and diagnosis results of other consulted cardiologists first. This will not only optimize time and energy but also help in identifying the underlying causes more efficiently. The website will incorporate a search bar to simplify the process of searching for specific patient medical information by entering relevant keywords. In addition to this, it will introduce a filtering feature that allows doctors to search a patient's EMRs based on the year. The website will utilize AI to examine all the vital criteria provided by the client. If any potential risk factors are identified, the site will display a warning message and offer recommendations for the appropriate specialists. The integration of automated recommendation represents a paradigm shift in healthcare, employing AI, ML, and data analytics to offer personalized guidance to healthcare professionals and patients. These systems analyze patient data, create personalized treatment plans, predict health risks, monitor health remotely, prescribe medications, optimize resource allocation, and contribute to clinical trials. By leveraging extensive datasets, these automated recommendations enhance patient outcomes and streamline healthcare delivery, showcasing a substantial leap in utilizing data for improved medical practices. It also aspires to utilize AI to analyze user-provided information about their current health issues and provide recommendations for appropriate first-aid measures and relevant specialists. Besides, the users will be able to submit their urgent needs and preferred location for blood through a new section on the website. Anyone in the vicinity who is willing to donate blood will be promptly notified by the application.

10 Conclusion

The current paper implements a multi-functional history card system to be the best companion in one's healthcare world. In other words, it terminates the conventional papered system of preserving patients' medical history by introducing a portable card; that accumulates every tiny detail such as personal narrative, prescriptions, reports, drugs, vaccine history, etc. The card serves the dual purpose of consolidating Electronic Medical Records (EMRs) into a unified system and ensuring that users have controlled access to protect their privacy. Crucially, a brief review of the history card provides doctors with a more concise overview of the patient's medical background. Moreover, it makes it easy for the users to search for and get an auto-update of their required medical information within a trouble-free go. Besides, the understandable user interface can assist people in avoiding any last-minute stress while visiting doctors. Conversely, doctors and medical officers can also track patients' diagnoses properly. They can view, read, and update medical data as per their need and avoid the hassles of handling lots of reports and papers, eventually protecting their goodwill in the long run. Hence, the medical history card comes into the picture in every scenario of contributing to the flourishment of the healthcare world. All in all, Blockchain technology can revolutionize healthcare with secure, interoperable, and patient-centric data management. Research priorities include improving security, ensuring interoperability, implementing consent management, addressing legal challenges, and enhancing user experience and adoption. Continued research and development efforts have the potential to significantly elevate the capabilities and reliability of blockchain-based systems for medical records, transforming the healthcare industry.

References

[1] S. Fairooz, S. Y. Miti, Z. Islam, M. T. Zaman, "A medical history card utilizing the Blockchain technology," in 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 1–6, 2023, doi:10.1109/HORA58378.2023.10156689.

[2] R. Hillestad, J. Bigelow, A. Bower, F. Girosi, R. Meili, R. Scoville, R. Taylor, "Can Electronic Medical Record Systems Transform Health Care? potential health benefits, savings, and costs," Health Affairs, **24**(5), 1103–1117, 2005, doi:10.1377/hlthaff.24.5.1103.

[3] C. M. Ruland, H. Bryhni, R. Andersen, T. Bryhni, "Developing a Shared Electronic Health Record for Patients and Clinicians," Studies in health technology and informatics, **136**, 57–62, 2008.

[4] K. Phillips, C. Wheeler, J. Campbell, A. Coustasse, "Electronic medical records in long-term care," J. Hosp. Mark. Public Relations, **20**(2), 131–142, 2010.

[5] M. Gomi, Y. Nakayama, Y. Sakurai, R. Oyama, K. Iwasaki, M. Doi, Y. Liu, M. Hori, H. Watanabe, K. Hashimoto, H. Tanaka, K. Tange, Y. Nakai, H. Akita, "Tolerogenic lipid nanoparticles for delivering self-antigen mRNA for the treatment of experimental autoimmune encephalomyelitis," Pharmaceuticals (Basel), **16**(9), 1270, 2023.

[6] Y. C. Li, P. S. Lee, W. S. Jian, C. H. Kuo, "Electronic health record goes personal world-wide," Yearb. Med. Inform., **18**(01), 40–43, 2009.

[7] S. Soegijoko, I. Puspitasari, A. Aridarma, I. Jani, "e-health for improving community healthcare: Encouraging clinical experience of simple e-prescription system and m-health system development for mother and childcare," in 2011 IEEE 13th International Conference on e-Health Networking, Applications and Services, 102–105, 2011, doi:10.1109/HEALTH.2011.6026722.

[8] E. T. van der Velde, H. Foeken, T. A. Witteman, L. van Erven, M. J. Schalij, "Integration of data from remote monitoring systems and programmers into the hospital electronic health record system based on international standards," Neth. Heart J., **20**(2), 66–70, 2012.

[9] O. Ben-Assuli, I. Shabtai, M. Leshno, "Using electronic health record systems to optimize admission decisions: the Creatinine case study," Health Informatics J., **21**(1), 73–88, 2015.

[10] A. Chhatlani, A. Dadlani, M. Gidwani, M. Keswani, P. Kanade, "Portable Medical Records Using Internet of Things for Medical Devices," in 2016 8th International Conference on Computational Intelligence and Communication Networks (CICN), 446–451, 2016, doi:10.1109/CICN.2016.93.

[11] A. Ekblaw, A. Azaria, J. Halamka, A. Lippman, "A case study for blockchain in healthcare:"MedRec" prototype for electronic health records and medical research data. White Paper," Retrievedfrom https://static1. squarespace. com/static/59aae5e9a803bb10bedeb03e, **5**, 1517278000381, 2016.

[12] A. Dubovitskaya, Z. Xu, S. Ryu, M. Schumacher, F. Wang, "Secure and trustable electronic medical records sharing using blockchain," in AMIA annual symposium proceedings, volume 2017, 650, American Medical Informatics Association, 2017.

[13] "Why Blockchain Technology Is Important for Healthcare Professionals — Mendeley," https://www.mendeley.com/search/?page=1&query=Why%20Blockchain%20Technology%20Is%20Important%20for%20Healthcare%20Professionals&sortBy=relevance, (Accessed on 11/30/2023).

[14] A. F. Hussein, N. ArunKumar, G. Ramirez-Gonzalez, E. Abdulhay, J. M. R. Tavares, V. H. C. de Albuquerque, "A medical records managing and securing blockchain based system supported by a Genetic Algorithm and Discrete Wavelet Transform," Cognitive Systems Research, **52**, 1–11, 2018, doi: https://doi.org/10.1016/j.cogsys.2018.05.004.

[15] R. Guo, H. Shi, Q. Zhao, D. Zheng, "Secure Attribute-Based Signature Scheme With Multiple Authorities for Blockchain in Electronic Health Records Systems," IEEE Access, **6**, 11676–11686, 2018, doi:10.1109/ACCESS.2018.2801266.

[16] J. Liu, X. Li, L. Ye, H. Zhang, X. Du, M. Guizani, "BPDS: A Blockchain Based Privacy-Preserving Data Sharing for Electronic Medical Records," in 2018 IEEE Global Communications Conference (GLOBECOM), 1–6, 2018, doi:10.1109/GLOCOM.2018.8647713.

[17] K. Riad, R. Hamza, H. Yan, "Sensitive and Energetic IoT Access Control for Managing Cloud Electronic Health Records," IEEE Access, **7**, 86384–86393, 2019, doi:10.1109/ACCESS.2019.2926354.

[18] Y. Zhao, M. Cui, L. Zheng, R. Zhang, L. Meng, D. Gao, Y. Zhang, "Research on electronic medical record access control based on blockchain," International Journal of Distributed Sensor Networks, **15**(11), 1550147719889330, 2019.

[19] V. Mehra, P. Sarvari, N. Ruban, "RFID Based Secured, Remotely Accessible Personal Medical Data Base Including the Medicinal History," in 2018 4th International Conference on Computing Communication and Automation (ICCCA), 1–4, 2018, doi:10.1109/CCAA.2018.8777617.

[20] M. T. de Oliveira, L. H. A. Reis, R. C. Carrano, F. L. Seixas, D. C. M. Saade, C. V. Albuquerque, N. C. Fernandes, S. D. Olabarriaga, D. S. V. Medeiros, D. M. F. Mattos, "Towards a Blockchain-Based Secure Electronic Medical Record for Healthcare Applications," in ICC 2019 - 2019 IEEE International Conference on Communications (ICC), 1–6, 2019, doi: 10.1109/ICC.2019.8761307.

[21] S. Siva Rama Krishnan, M. Manoj, T. R. Gadekallu, N. Kumar, P. K. R. Maddikunta, S. Bhattacharya, D. Y. Suh, M. J. Piran, "A Blockchain-Based Credibility Scoring Framework for Electronic Medical Records," in 2020 IEEE Globecom Workshops (GC Wkshps, 1–6, 2020, doi:10.1109/GCWkshps50303.2020.9367459.

[22] C.-T. Li, D.-H. Shih, C.-C. Wang, C.-L. Chen, C.-C. Lee, "A Blockchain Based Data Aggregation and Group Authentication Scheme for Electronic Medical System," IEEE Access, **8**, 173904–173917, 2020, doi:10.1109/ACCESS.2020.3025898.

[23] M. T. Mahmud, F. Soroni, M. M. Khan, "Development of a Mobile Application for Patient's Medical Record and History," in 2021 IEEE World AI IoT Congress (AIIoT), 0081–0085, 2021, doi:10.1109/AIIoT52608.2021.9454227.

[24] M. Nasim, S. Abdullah-Al-Noman, A. Ragib Hasan, A. Sattar, "Digitalization and Centralization of Medical Information and Patient History in Bangladesh," in 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), 153–158, 2022, doi:10.1109/ICCMC53470.2022.9753690.

[25] R. Cerchione, P. Centobelli, E. Riccio, S. Abbate, E. Oropallo, "Blockchain's coming to hospital to digitalize healthcare services: Designing a distributed electronic health record ecosystem," Technovation, **120**, 102480, 2023, doi: https://doi.org/10.1016/j.technovation.2022.102480.

[26] S. Singh, S. Kumar Sharma, P. Mehrotra, P. Bhatt, M. Kaurav, "Blockchain technology for efficient data management in healthcare system: Opportunity, challenges and future perspectives," Materials Today: Proceedings, **62**, 5042–5046, 2022, doi:https://doi.org/10.1016/j.matpr.2022.04.998, international Conference on Innovative Technology for Sustainable Development.

[27] M. Haidar, S. Kumar, "Smart Healthcare System for Biomedical and Health Care Applications using Aadhaar and Blockchain," in 2021 5th International Conference on Information Systems and Computer Networks (ISCON), 1–5, 2021, doi:10.1109/ISCON52037.2021.9702306.

[28] N. Malhotra, M. Lassiter, et al., "The coming age of electronic medical records: From paper to electronic," International Journal of Management & Information Systems (IJMIS), **18**(2), 117–122, 2014.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 3, 100-107 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Landmarking Technique for Improving YOLOv4 Fish Recognition in Various Background Conditions

Sutham Satthamsakul, Ari Kuswantori, Witsarut Sriratana, Worapong Tangsrirat*, Taweepol Suesut

Department of Instrumentation and Control Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, 10520, Thailand

ARTICLE INFO

Article history:
Received: 20 February, 2023
Accepted: 11 May, 2023
Online: 21 May, 2023

Keywords:
Fish classification
Fish recognition
YOLO
Landmarking technique
Computer vision

ABSTRACT

The detection and classification of fish is a prevalent and fascinating area of study. Numerous researchers develop skills in fish recognition in both aquatic and non-aquatic environments, which is beneficial for population control and the aquaculture industry, respectively. Rarely is research conducted to optimize the recognition of fish with diverse backgrounds. This paper proposes a method for fish recognition that uses the landmarking technique to optimize YOLO version 4 to detect and classify fish with varying background conditions, making it applicable for both underwater and terrestrial fish recognition. The proposed procedure was evaluated on the Bringham Young University (BYU) dataset containing four different fish species. The final test results indicate that the detection accuracy had reached 96.60% with an average confidence score of 99.67% at the 60% threshold. The outcome is 4,94% better than the most common traditional labeling method.

1. Introduction

Fish detection and classification, or fish recognition, are exciting and popular research topics [1]. Numerous researchers devise techniques for recognizing fish in underwater and out-of-water environments [2]. Recognition of fish in their underwater condition is beneficial for fish population control and their long-term viability [3-8], especially in light of climate change and global warming [5, 8-10]. It is advantageous for the aquaculture industry to recognize fish in conditions outside the water, such as through automatic classification processes, fish quality control, or other processes [2, 11].

Fish recognition has quite complex and varied challenges, both in conditions in the water and outside the water [1]. The challenges of recognizing fish in underwater conditions are; resolution of the image, swimming freely, camouflaging with other objects and other fish, in various background conditions and lighting conditions, and other challenges such as noise [3, 5, 7-9, 12]. For fish recognition in out-of-water conditions, the challenges that arise are structural deformations of the fish, such as the condition of the eyes, scales, abdomen, and fins that are damaged lightly, moderately, or severely. The condition of random fish positions is also a challenge in this recognition process [11].

Many image processing techniques are used to overcome the challenges, such as; noise filtering by the median filter [11], fish head detection using convex-hull deficiency [11], Color Multi-Scale Retinex (MSR) to overcome water turbidity [13], orientation correction using MSEE [11], contrast enhancement to find a fish disease [14], and auto segmentation of fish objects with various techniques [9, 11]. Deep learning methods or models are also widely used for fish detection and classification, such as Convolutional Neural Network (CNN) [3, 4, 15-18], YOLO [5, 13, 19-23], few-shot learning for limited training images [6], Alex-Net, ResNet-18, ResNet-50, Inception V3, GoogleLeNet [11], etc.

Numerous strategies have been implemented to overcome the various background challenges and optimize the fish recognition results. CNN was used to classify fish in [3] by training it with the number of species and their environments, including reef substrates and water. They applied their proposed method to 116 videos of fish captured underwater with a GoPro camera. The greatest results were obtained when classifying nine of the twenty fish species that appear most frequently in the videos. Using an ensemble of seven CNN components and two Region Proposal Networks (RPNs) coupled by sequentially jointly trained Long Short-Term Memory (LSTM) units, a multi-cascade object detection network was constructed as described in [4].

They used a collection of 18 underwater fish recordings recorded with a GoPro underwater camera for training and testing. Even though their proposed method can reliably identify and count fish objects in a variety of benthic backgrounds and lighting conditions, it is not used to classify fish. Rather, it is used to locate

*Corresponding Author: Worapong Tangsrirat, worapong.ta@kmitl.ac.th

https://dx.doi.org/10.25046/aj080312

fish. Using classic CNNs such as these has benefits when applied to other industries, such as agriculture [24] or in other general situations, such as detecting micro scratches [25]. In [12], the foreground extraction was implemented and a CNN-based algorithm with an SVM classifier was created. The instant segmentation on underwater fish images was proposed in [9] and then developed an algorithm based on Res-Net. The recognition of moving fish in [5] used Optical flow, GMM (Gaussian Mixture Models), and ResNet-50, and then combined the output with YOLOv3. The combination of these techniques allowed for the robust detection and classification of fish, which was applied to the LifeCLEF 2015 benchmark dataset from the Fish4Knowledge repository [26] and a dataset collected by the University of Western Australia (UWA) that was described in detail [27]. The GMM and Pixel-wise Posteriors were proposed in [8], and then combined with CNN in [28] for further development. In addition, they utilized a fish dataset extracted from the Fish4Knowledge repository. Similar to the work [4], the approaches proposed in their papers were limited to fish detection without classification.

The fish was divided into three sections: the head, the torso, and the scales [11]. Alex-Net was then utilized to categorize each of these components. The ultimate classification results were then determined utilizing Naive Bayesian Fusion (NBF). This method performed admirably when applied to the Fish-Pak [29] and Bringham Young University (BYU) datasets [20]. Even though the narrative of this work was intended for an automatic classification system, the fish images used were static. The method for detecting fish in aquaculture ponds was developed in [13]. YOLOv3 was utilized to detect the fish after image enhancement was used to improve fish detection in murky water conditions. This method is not intended for classification, but rather for counting and monitoring the trajectory of fish. Faster R-CNN was used to detect and record fish trajectories and compared with YOLOv3 to investigate its behavior and relationship to ammonia levels in pond water [30].

However, the works cited in [3-5, 8, 28] identified moving fish for underwater (ocean) environments, whereas other works merely detected fish without classifying them. In aquaculture, the works of [13] and [30] discovered swimming fish, but it was not used for classification. The work described in [11] classified fish using narration for an automatic classification system, but the datasets used were images that were not dynamic. According to the knowledge of the authors, there is no public dataset for cultured fish that travel on conveyors, and there is no published work to detect and classify moving fish for the fish culture industry, particularly for automatic sorting based on fish species using deep learning and computer vision.

There are numerous versions of YOLO (You Only Look Once) algorithms, such as YOLOv3 [19], YOLOv4 [23], and YOLOv5 [21], that are suitable for real-time applications in terms of detection speed and precision. The optimal algorithm is YOLOv5, but we opted for YOLOv4 on the CiRA-Core platform, which yielded satisfactory results. This study proposes a straightforward method that incorporates the landmarking technique and YOLOv4 for identifying and classifying fish in various background conditions. In the most prevalent labeling technique, the accuracy results will be compared to the conventional "box-squeezing-object" method.

This paper proposes and evaluates a method for detecting and classifying fish using deep learning and computer vision on real videos of aquaculture freshwater fish moving along a conveyor belt for automatic sorting. Consequently, the proposed methodology originates from the following major contributions.

- We created our own database of eight species of cultivated fish. The dataset includes both static images and videos of fish moving at random with two distinct speeds (i.e., low and high) on a conveyor belt.
- This method used YOLOv4, a widely-used recognition algorithm that was optimized using a novel labeling method.
- Additionally, trial research evaluating the efficacy of multiple strategies was conducted. These are training schemes, variants of YOLOv4, and labeling technique comparisons.

The paper is therefore divided into four sections. Section 1 introduces the background of the problem, and approaches from previous researchers. In Section 2, the input datasets, augmentation techniques, landmarking, and algorithms used in this study are discussed. Section 3 contains the outcomes of the experiment. In Section 4, a conclusion and future work are summarized.

2. Material and Methodology

2.1. Dataset and Image Augmentation

This study aimss to detect and categorize fish in a variety of background conditions, such as objects, hues, and others. For this reason, the BYU dataset is deemed highly suitable for use in this work. This dataset was developed by the robotic vision group at Brigham Young University [20], which contains sixteen species of fish photographed under various conditions. Multiple images of fish are captured in the processing unit on a table equipped with an image calibration plate, which provides guidelines for preprocessing and color correction. Other fish photographs were captured in nature and water. Some images have also been altered by means of cropping, recoloring, etc.

Additionally, each species' image from the sixteen different fish species is quite imbalanced. Due to these factors, only four species were obtained in this study despite the large quantity of image data, unprocessed raw images, and diverse background conditions, i.e., natural and underwater backgrounds. The four fish species listed in Table 1 are B.C. Trout, Kokanee, UT Sucker, and Steal Head. Figure 1 contains illustrations of image data from the BYU dataset. Table 1 shows that the amount of images for the three classes, Kokanee, UT Sucker, and Steal Head, is insufficient and out of proportion to the amount of image data for the entire class. Less than 100 data points reduce validation, and unbalanced data inhibits the algorithm from having the same training chance. Table 2 displays the statistical data for the BYU dataset and image augmentation. This calls for a procedure of augmentation [11]. In this study, the augmentation methods employed include flip, rotation, and translation. This method is thought to be appropriate for fish augmentation [21].

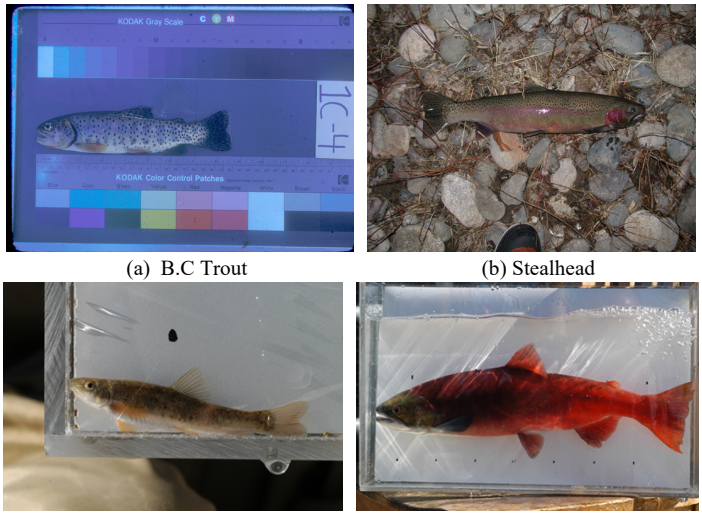
Table 1: BYU dataset and image augmentation

Fish class	No. of images	Multiplication factor	No. of augmented images	New image dataset	For training (80%)	For testing (20%)

B.C. Trout	191	0	0	191	153	38
Kokanee	60	1	60	120	96	24
UT Sucker	87	1	87	174	139	35
Steal Head	25	3	75	100	80	20
Total	363	-	222	585	468	117

Table 2: Statistical data of BYU dataset and image augmentation

No. of images	Fish class				Total	Standard Deviation Value (SDV)	Average
	B.C. Trout	Kokanee	UT Sucker	Steal Head			
Images dataset	191	60	87	25	363	61.91	91
New images dataset	191	120	174	100	585	37.42	146



(a) B.C Trout (b) Stealhead

(c) UT Sucker (d) Kokanee

Figure 1: Sample images of BYU dataset

Initially, the number of images (N_c) for each class of fish images (I_f^c) is determined as shown in (1). The multiplication factor (m_f^c) is obtained by comparing the target (N_T) with the number of images in each class (N_c) as in (2). In this work, the target (N_T) is set at 100 for each class. The set of multiplication factors (m_f) is obtained by repeating the procedure for each class of fish. The multiplication factor indicates the number of augmented images that need to be created from each image (n) in each class of fish. Flip (F_a), rotation (θ_a), and translation vector (T_a) are selected randomly and generate an augmented image of the I_{af}^c set obtained from a single image (I_f^{Nc}) as in (3). The variances of each of the flip (F_a), rotation (θ_a), and translation vectors (T_a) were predetermined. Finally, all the augmented image sets (I_{af}^{Nc}) are merged with the original image set (I_f^c), and a new data set (I_{Nf}^c) is formed as in (4).

$$I_f^C = \left\{ I_f^{1C}, I_f^{2C}, \ldots, I_f^{N_c C} \right\} \tag{1}$$

$$m_f = \left\{ m_f^c = \left| 1 - \frac{N_f}{N_c} \right|, C \in [1,2,\ldots C], N_c < N_T \right\} \tag{2}$$

$$I_{af}^{nc} = \left\{ H\left(I_f^{nc}, F_a, \theta_a, T_a \right) \middle| a \in \left[1,2,\ldots, m_f^c \right] \right\} \tag{3}$$

and

$$I_{Nf}^C = \left\{ I_f^{1C}, I_f^{2C}, \ldots, I_f^{N_c C}, I_{af}^{1C}, I_{af}^{2C}, \ldots, I_{af}^{N_c C} \right\} \tag{4}$$

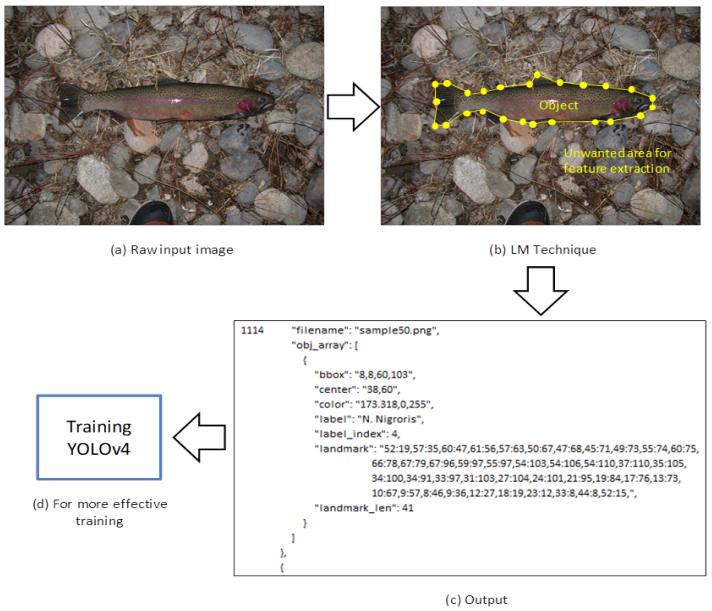
After the augmentation procedure, the number of images in the new dataset, which ranges from 363 to 585, has increased significantly. The average number of images in each class increased from 91 to 146. In addition, the dataset is now more evenly distributed, as indicated by the decrease in standard deviation from 61.91 to 37.42. The training (80%) and testing (20%) processes were then randomly applied to the images [7].

2.2. Occupancy Ratio and Landmarking Technique

The occupancy ratio (*OR*) compares the object area to the overall image area or bounding box, which includes the background, and is measured in $M \times N$. The high occupancy ratio reduces the likelihood of an algorithm for deep learning capturing extraneous backgrounds during training. It will increase the effectiveness of deep learning in identifying features unique to objects. The occupancy percentage is represented by the following expression [11]:

$$OR_{bb} = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} I_{bb}(i,j)}{M \times N} \tag{5}$$

The landmarking technique is a relatively new labeling method that is still infrequently employed, particularly on fish objects. This technique is utilized immediately prior to the image being utilized in the training procedure. With this technique, any portion of the object may be marked and become an area that is processed only during training. The *OR* will increase to a very high level or even reach unity, resulting in more accurate and efficient extracted feature maps. As a result, the training process will be more efficient, and the recognition results are anticipated to be more accurate and robust.



(a) Raw input image (b) LM Technique

(d) For more effective training (c) Output

Figure 2: Landmarking technique.

In theory, the landmarking technique will arrange the object's coordinate coordinates into a polygon and create an area on the object. Only objects or portions of objects that have been marked as areas will be extracted later for the training process; disregard other objects outside of the area. This method is utilized concurrently during the labeling procedure (class recognition). In addition to bounding box (bbox), center point, color, label, and index label information, the output of this process includes landmark points and landmark len, which will be utilized to improve the effectiveness of the training process. Figure 2 illustrates the concept of this method of landmarking.

Figure 3 illustrates procedure of the proposed scheme. First, the image data from the BYU dataset were subjected to an image augmentation procedure in order to increase the number of images by means of twisting, rotating, and transitioning. The new fish image dataset was then partitioned into two parts: 80% for the training dataset, which was subjected to the landmarking technique and trained with YOLO v4, and 20% for the testing dataset. Accuracy, precision, recall, specificity, and the F1 score were the outcomes of the classification procedure. Using the CiRA-Core software, this method of landmarking is implemented.

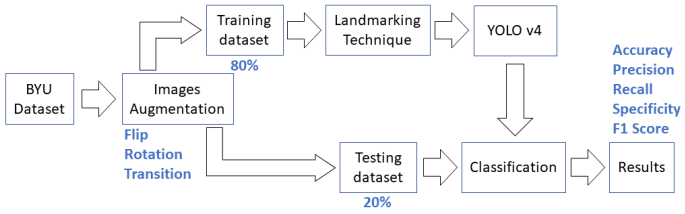


Figure 3: Overview of the proposed scheme.

(a) (b)

Figure 4: (a) labeling using the landmarking technique (b) conventional and the most commonly labeling technique.

This paper also discusses fish recognition using the most prevalent conventional labeling method, namely the "box squeezing object" method. Figure 4 illustrates the distinction between the landmarking method and the conventional method.

2.3. YOLOv4

In this experiment, the detection and classification algorithm is YOLO version 4 (YOLOv4). YOLO is an algorithm for detecting and classifying viral objects. Due to its reasonable detection rate, accuracy, and quick detection time, YOLO has become popular and extensively employed. YOLO operates differently than the majority of other object detection algorithms. A singular neural network is applied to the entire image by YOLO. The network will segment the image into regions, identify the bounding box, and designate a probability value. The probability of classifying each unit within the boundary as an object or not is measured. Finally, the bounding box with the highest value is chosen as the object separator [23]. It makes the YOLO algorithm suitable for detecting fish as objects in a variety of background conditions, even when the background is in motion or when there are multiple fish in a single image. As depicted in Figure 5 [23], the initial YOLO structure and working system consisted of 24 convolutional layers and two completely connected layers.

The YOLOv4 utilized in this work is the April 2020 release of the fourth development version of YOLO [23]. In this research, YOLOv4 achieves state-of-the-art performance with 43.5% AP operating at 65 FPS in real time on a Tesla V100 with the MS COCO dataset. To achieve these outcomes, they combine Cross-Stage-Partial-connections (CSP), DropBlock regularization, Weighted-Residual-Connections (WRC), Self-adversarial-training (SAT) and Mish -activation, CIoU loss, Cross mini-Batch Normalization (CmBN), and Mosaic data augmentation [31].

As shown in Table 1, YOLOv4 was trained using training data from the BYU dataset and was able to achieve acceptable results with an average loss of 0.01. Training is conducted with 32 data points and 32 subdivisions per iteration. Data enhancement is performed by rotating a threshold between -180° and 180° with 90 steps and comparing it to a threshold between 0.4 and 1.1 with 0.2 steps. Simulating noise and blur during the training procedure is not used to improve data.

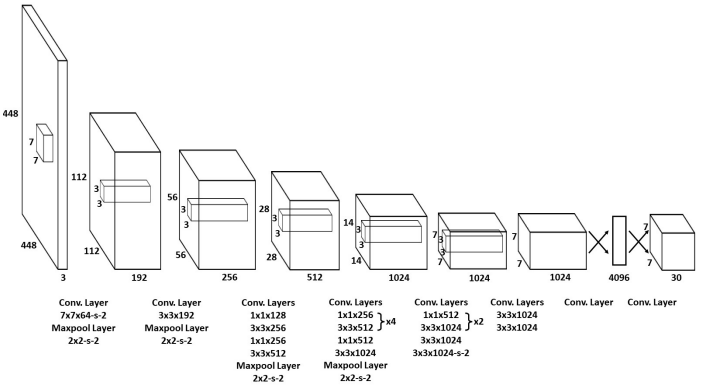


Figure 5: YOLO architecture [22].

2.4. Validation Matrix

The confusion matrix is utilized to characterize the model's performance. The confusion matrix is composed of four components: True Positive (*TP*) and True Negative (*TN*) for fundamental truth, and False Positive (*FP*) and False Negative (*FN*) for fundamental false. *TP* is when the model correctly detects the object, *TN* is when the model does not detect because the object does not exist, *FP* is when the model detects an object incorrectly, including double detecting, and *FN* is when the model does not detect objects [32].

- Accuracy

One of the evaluation metrics is accuracy. It is defined as the proportion of correctly classified fish to the total number of fish in the test set.

$$\text{Accuracy} = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) \times 100\% \cdot \qquad (6)$$

Alternately, the following relation represents a comparison between the total number of predictions and the number of accurate predictions as a measure of the model's precision.

$$\text{Accuracy} = \left(\frac{\sum_{i}^{N} P_i}{\sum_{i}^{N} |Q_i|} \right) \times 100\% \cdot \quad (7)$$

In (7), $\sum_{i}^{N} P_i$ is the number of correct predictions, and $\sum_{i}^{N} |Q_i|$ is the total number of predictions.

- Precision

Precision is the ratio of correctly classified fish (TP) to positive detection ($TP + FP$). It determines the percentage of precisely classified fish as specified by:

$$\text{Precision} = \left(\frac{TP}{TP + FP} \right) \times 100\% \cdot \quad (8)$$

- Sensitivity

Sensitivity is the ratio of correctly classified fish (TP) to fundamental truth fish ($TP + FN$), as defined by:

$$\text{Sensitivity} = \left(\frac{TP}{TP + FN} \right) \times 100\% \cdot \quad (9)$$

- Specificity

Specificity is determined by the ratio of TN to the sum of FP and TN, as stated below:

$$\text{Specificity} = \left(\frac{TN}{TN + FP} \right) \times 100\% \cdot \quad (10)$$

- F1 Score

F1 Score (Measure F) is a metric derived from the average of precision symphony and memory [32], based on the following relationship:

$$\text{F1 Score} = 2 \left(\frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \right) \times 100\% \cdot \quad (11)$$

3. Experimental Results

The proposed method has been evaluated based on the BYU testing data set. The testing data consists of 20% of each image of four fish classes (species) that have been enhanced, as shown in Table 1. At the 50% threshold, the average ultimate detection accuracy is 94.22% and the average classification score for correct detection is 99.42%. The UT Sucker class had the lowest accuracy at 88.57%, while the Kokanee class had the highest accuracy at 100%. Table 3 provides a summary of the results, while Figure 6 depicts the confusion matrix.

Interestingly, the detection continues to run well even though there are actually different types of fish in one class (class BC Trout, Kokanee, and UT Sucker). There are several types of BC Trout, including BC Trout Provo, BC Trout Little Dell, and BC Trout H Creek in the BC Trout class. These varieties of BC Trout have different colors and scales. As depicted in Figure 7, there are variants of Kokanee Male and Kokanee Female within 1 Kokanee class that have slightly distinct morphologies. Similarly, the UT Suckerfish class includes the fish species UT Sucker, UT Sucker Weber, and UT Sucker Imm Weber, each of which has a slightly unique scale pattern.

Table 3: Experimental results of YOLOv4 with landmarking technique at the 50% threshold.

Threshold: 50%								
Class	**Testing data**	**Correct classification**	**Wrong classification**	**Not detect**	**Double classification**	**Total prediction**	**Accuracy (%)**	**Avg. Classification score (%)**
BC Trout	38	35	0	3	0	38	92.11	99.67
Kokanee	24	24	0	0	0	24	100.00	99.93
UT Sucker	35	31	1	0	3	35	88.57	99.20
Steelhead	20	19	0	0	1	20	95.00	99.09
Total	117	-	-	-	-	-	-	-
Average	-	-	-	-	-	-	93.92	99.47

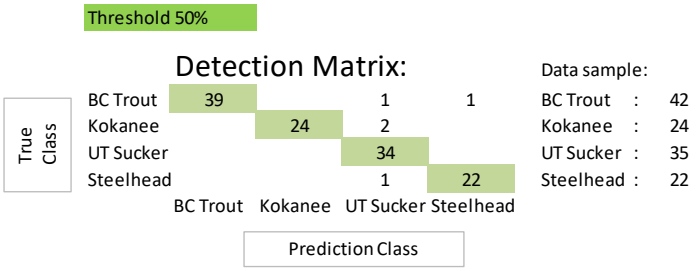
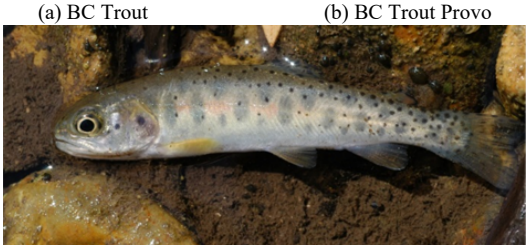


Figure 6: Confusion matrix for experiment results of YOLOv4 with landmarking technique at the 50% threshold.

(a) BC Trout

(b) BC Trout Provo

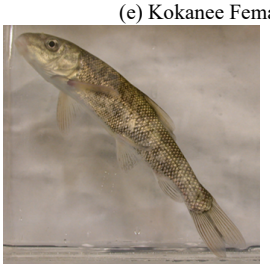


(c) BC Trout Little Dell

(d) BC Trout H Creek

(e) Kokanee Female

(f) Kokanee Male



(g) UT Sucker

(h) UT Sucker Weber

Figure 7: Examples of different fish in one class.

The test also revealed that double detection occurs frequently. One fish object was incorrectly identified as two different fish objects with a confidence level below 60% for the incorrect classi fication. It optimizes the accuracy level by increasing the threshold to 60%, enabling the final average detection accuracy level to reach 96.79% and the average confidence score for correct detection to reach 99.42%. As summarized in Table 4 and Figure 8, the lowest detection accuracy reached 92.86% for the BC Trout class, while the maximum detection accuracy reached 100% for the Kokanee and Steelhead classes. Fish structural deformities are responsible for detection failure. This issue was considered during the development of an algorithm utilizing the Fish-Pak dataset [33].

This work also evaluates the YOLOv4 algorithm model using the conventional "box squeezing object" labeling technique, which is most commonly applied to the same set of test data. With the conventional labeling technique, the final average accuracy is 90.09 % with a confidence score of 99.19 % at the 50% threshold, and the final average accuracy is 92.56 % with a confidence score of 99.19 % at the 60% threshold. At the 50% threshold, the accuracy is 4.59% lower than with the landmarking technique, and at the 60% threshold, it is 4.57% lower. The results of YOLOv4 detection using this conventional labeling technique are summarized in Tables 5 and 6, and Figure 9 depicts the confusion matrix.

Table 4: Experimental results of YOLOv4 with landmarking technique at the 60% threshold.

Threshold: 60%								
Class	**Testing data**	**Correct classification**	**Wrong classification**	**Not detect**	**Double classification**	**Total prediction**	**Accuracy (%)**	**Avg. Classification score (%)**
BC Trout	38	35	0	3	0	38	92.11	99.67
Kokanee	24	24	0	0	0	24	100.00	99.93
UT Sucker	35	33	1	0	1	35	94.29	99.24
Steelhead	20	20	0	0	0	20	100.00	99.83
Total	117	-	-	-	-	-	-	-
Average	-	-	-	-	-	-	96.60	99.67

Table 5: YOLOv4 detection results with conventional labeling techniques at the 50% threshold.

Threshold: 50%								
Class	**Testing data**	**Correct classification**	**Wrong classification**	**Not detect**	**Double classification**	**Total prediction**	**Accuracy (%)**	**Avg. Classification score (%)**
BC Trout	38	35	1	2	0	38	92.11	99.12
Kokanee	24	21	0	0	3	24	87.50	99.99
UT Sucker	35	27	0	0	8	35	77.14	99.96
Steelhead	20	20	0	0	0	20	100.00	99.87
Total	117	-	-	-	-	-	-	-
Average	-	-	-	-	-	-	89.19	99.74

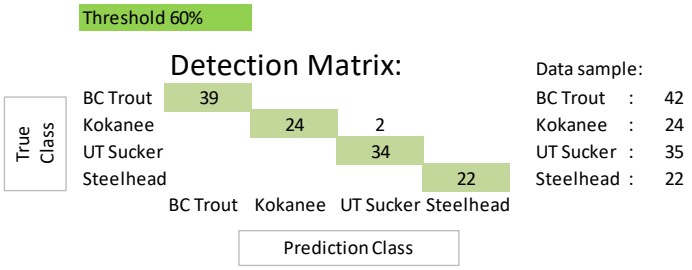


Figure 8: Confusion matrix for experiment result of YOLOv4 with landmarking technique at the 60% threshold.

Table 6: YOLOv4 detection results with conventional labeling techniques at the 60% threshold.

Threshold: 60%								
Class	**Testing data**	**Correct classification**	**Wrong classification**	**Not detect**	**Double classification**	**Total prediction**	**Accuracy (%)**	**Avg. Classification score (%)**
BC Trout	38	35	1	2	0	38	92.11	99.12
Kokanee	24	22	0	0	2	24	91.67	99.99
UT Sucker	35	29	0	0	6	35	82.86	99.96
Steelhead	20	20	0	0	0	20	100.00	99.87
Total	117	-	-	-	-	-	-	-
Average	-	-	-	-	-	-	91.66	99.74

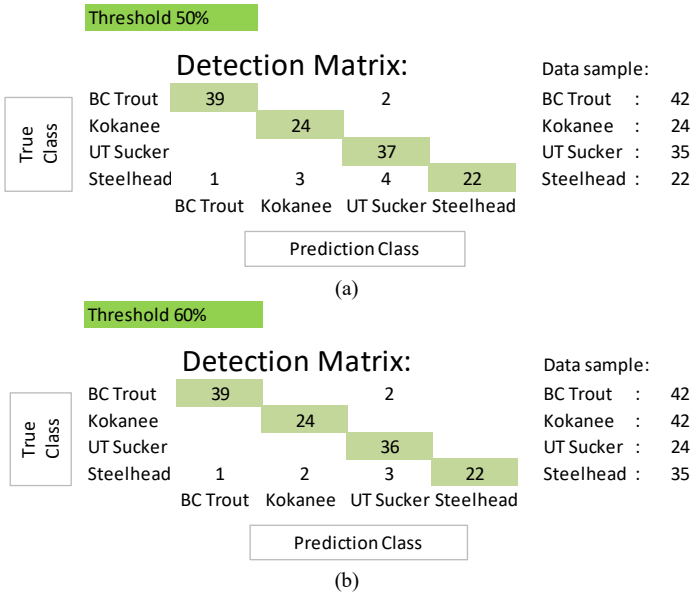


Figure 9: Confusion matrix for experiment result of YOLOv4 with conventional labeling technique
(a) at the 50% threshold (b) at the 60% threshold.

4. Conclusion

The purpose of this paper is to propose a method for detecting fish in a variety of background conditions. Combining the landmarking technique with YOLO version 4 is the method. The proposed method was evaluated using the BYU dataset, which contained four different fish species. According to the experimental test results, the detection accuracy is 94.22% at a 50% threshold and 96.79 % at a 60% threshold. This result is greater than the 4.59% at the 50% threshold and 4.57% at the 60% threshold obtained using the conventional labeling method. The results of this trial are considered quite favorable, and the method used is straightforward. The proposed method does not require complicated work steps and is available for implementation.

In future research, it may be possible to automate the landmarking technique in fish using BLOB. Through the implementation of image processing techniques, it is possible to increase precision by adjusting the threshold value. The challenge that will be faced is the ability to detect and separate fish objects from their backgrounds, find the edges of fish objects, and export the coordinate points on the fish edges needed to define the required area.

Acknowledgments

This work was financially supported by King Mongkut's Institute of Technology Ladkrabang, Contract no. 2562-02-01-004. The authors would also like to thank AMI-KMITL for granting the CiRA-Core software license.

References

[1] M. K. Alsmadi, I. Almarashdeh, "A survey on fish classification techniques," Journal of King Saud University - Computer and Information Sciences, **34**, 1625-1638, 2020, doi: https://doi.org/10.1016/j.jksuci.2020.07.005.

[2] S. Zhao, S. Zhang, J. Liu, H. Wang, J. Zhu, D. Li, R. Zhao, "Application of machine learning in intelligent fish aquaculture: A review, Aquaculture, **540**, 1-19, 2021, doi:https://doi.org/10.1016/j.aquaculture.2021.736724.

[3] S. Villon, D. Mouillot, M. Chaumont, E.S. Darling, G. Subsol, T. Claverie, S. Villéger, "A deep learning method for accurate and fast identification of coral reef fishes in underwater images," Ecological Informatics, **48**, 238-244, 2018,doi: https://doi.org/10.1016/j.ecoinf.2018.09.007.

[4] A. B. Labao, P.C. Naval, "Cascaded deep network systems with linked ensemble components for underwater fish detection in the wild," Ecological Informatics, **52**, 103-121, 2019, doi:https://doi.org/10.1016/j.ecoinf.2019.05.004.

[5] A. Jalal, A. Salman, A. Mian, M. Shortis, F. Shafait, "Fish detection and species classification in underwater environments using deep learning with temporal information," Ecological Informatics, **57**, 1-13, 2020, doi:https://doi.org/10.1016/j.ecoinf.2020.101088.

[6] S. Villon, C. Iovan, M. Mangeas, T. Claverie, D. Mouillot, S. Villéger, L. Vigliola, "Automatic underwater fish species classification with limited data using few-shot learning," Ecological Informatics, **63**, 1-6, 2021, doi:https://doi.org/10.1016/j.ecoinf.2021.101320.

[7] Z. Ju, Y. Xue, "Fish species recognition using an improved alexnet model," Optik, **223**, 1-9, 2020, doi:https://doi.org/10.1016/j.ijleo.2020.165499.

[8] A. Salman, S. Maqbool, A.H. Khan, A. Jalal, F. Shafait, "Real-time fish detection in complex backgrounds using probabilistic background modelling," Ecological Informatics, **51**, 44-51, 2019, doi:https://doi.org/10.1016/j.ecoinf.2019.02.011

[9] N. F. F. Alshdaifat, A. Z. Talib, M. A. Osman, "Improved deep learning framework for fish segmentation in underwater videos, "Ecological Informatics, **59**, 1-11, 2020, doi:https://doi.org/10.1016/j.ecoinf.2020.101121.

[10] A. B. Hollowed, M. Barange, R. J. Beamish, K. Brander, K. Cochrane, K. Drinkwater, M. G. G. Foreman , J. A. Hare, J. Holt, S. Ito, S. Kim, J. R. King, H. Loeng, B. R. MacKenzie, F. J. Mueter, T. A. Okey, M. A. Peck, V. I. Radchenko, J. C. Rice, M. J. Schirripa, A. Yatsu, Y. Yamanaka, "Projected impacts of climate change on marine fish and fisheries," ICES Journal of Marine Science, **70**, 1023–1037, 2013, doi:https://doi.org/10.1093/icesjms/fst081.

[11] N. S. Abinaya, D. Susan, R. K. Sidharthan, "Naive bayesian fusion based deep learning networks for multisegmented classification of fishes in aquaculture industries," Ecological Informatics, **61**, 1-13, 2021, doi:https://doi.org/10.1016/j.ecoinf.2021.101248.

[12] H. Qin, X. Li, J. Liang, Y. Peng, C. Zhang, "Deepfish: Accurate underwater live fish recognition with a deep architecture," Neurocomputing, **187**, 49-58, 2016, doi:https://doi.org/10.1016/j.neucom.2015.10.122.

[13] H. E. D. Mohamed, A. Fadl, O. Anas, Y. Wageeh, N. ElMasry, A. Nabil, A. Atia, "Msr-yolo: Method to enhance fish detection and tracking in fish farms," in The 11th International Conference on Ambient Systems, Networks and Technologies (ANT), 539-546, 2020, doi:https://doi.org/10.1016/j.procs.2020.03.123.

[14] M. S. Ahmed, T.T. Aurpa, M.A.K. Azad, "Fish disease detection using image based machine learning technique in aquaculture, "Journal of King Saud University - Computer and Information Sciences, **34**, 5170-5182, 2021, doi:https://doi.org/10.1016/j.jksuci.2021.05.003.

[15] A. A. D. Santos, W. N. Gonçalves, "Improving pantanal fish species recognition through taxonomic ranks in convolutional neural networks," Ecological Informatics, **53**, 1-11, 2019, doi:https://doi.org/10.1016/j.ecoinf.2019.100977.

[16] T. Miyazono, T. Saitoh, "Fish species recognition based on CNN using annotated image," in IT Convergence and Security 2017, Lecture Notes in Electrical Engineering, 449, 156-163, 2018, doi:https://doi.org/10.1007/978-981-10-6451-7_19.

[17] B. S. Rekha, G. N. Srinivasan, S. K. Reddy, D. Kakwani, N. Bhattad, "Fish detection and classification using convolutional neural networks," in International Conference On Computational Vision and Bio Inspired Computing, 1221-1231, 2019, doi:https://doi.org/10.1007/978-3-030-37218-7_128.

[18] A. Taheri-Garavand, A. Nasiri, A. Banan, Y.-D. Zhang, "Smart deep learning-based approach for non-destructive freshness diagnosis of common carp fish," Journal of Food Engineering, **278**, 1-9, 2020, doi:https://doi.org/10.1016/j.jfoodeng.2020.109930.

[19] K. Cai, X. Miao, W. Wang, H. Pang, Y. Liu, J. Song, "A modified yolov3 model for fish detection based on mobilenetv1 as backbone," Aquacultural Engineering, **91**, 1-9, 2020, doi:https://doi.org/10.1016/j.aquaeng.2020.102117.

[20] K. D. Lillywhite, D.J. Lee, Robotic vision lab, brigham young university, fish dataset, 2013, Available:http://roboticvision.groups.et.byu.net/Machine_Vision/BYUFish/BYU_Fish.html.

[21] Z. Liu, X. Jia, X. Xu, "Study of shrimp recognition methods using smart

networks," Computers and Electronics in Agriculture, **165**, 1-10, 2019, doi:https://doi.org/10.1016/j.compag.2019.104926.

[22] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You only look once: Unified, real-time object detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779-788, 2016, doi:https://doi.org/10.48550/arXiv.1506.02640.

[23] A. Bochkovskiy, C. Y. Wang, H. Y.M. Liao, "Yolov4: Optimal speed and accuracy of object detection," ArXiv:2004.10934, 1-17, 2020, doi:https://doi.org/10.48550/arXiv.2004.10934.

[24] C. Li, T. Zhen, Z. Li, "Image classification of pests with residual neural network based on transfer learning", Appl. Sci., **12**, 4356, 2022, doi:https://doi.org//10.3390/app12094356

[25] W. Li, L. Zhang, C. Wu, Z. Cui, C. Niu, "A new lightweight deep neural network for surface scratch detection", Int. J. Adv. Manuf. Technol., **123**, 1999–2015, 2022, doi:https://doi.org/10.1007/s00170-022-10335-8

[26] R. B. Fisher, Y. H. Chen-Burger, D. Giordano, L. Hardman, F. P. Lin, Fish4Knowledge: Collecting and Analyzing Massive Coral Reef Fish Video Data; Springer: Berlin/Heidelberg, Germany, 2016.

[27] S. A. Siddiqui, A. Salman, M. I. Malik, F. Shafait, A. Mian, M. R. Shortis, E. S. Harvey, "Automatic fish species classification in underwater videos: Exploiting pre-trained deep neural network models to compensate for limited labelled data", ICES J. Mar. Sci., **75**, 374–389, 2018, doi:https://doi.org/10.1093/icesjms/fsx109

[28] A. Salman, S. A. Siddiqui, F. Shafait, A. Mian, M. R. Shortis, K. Khurshid, A. Ulges, U. Schwanecke, "Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system", ICES J. Mar. Sci., **77**, 1295–1307, 2020, doi:https://doi.org/10.1093/icesjms/fsz025

[29] S. Z. H. Shah, H. T. Rauf, M. Ikram Ullah, M. S. Khalid, M. Farooq, M. Fatima, S. A. C. Bukhari, "Fish-pak: Fish species dataset from Pakistan for visual features based classification", Data in Brief, **27**, 104565, 2019, doi:https://doi.org/10.1016/j.dib.2019.104565

[30] W. Xu, Z. Zhu, F. Ge, Z. Han, J. Li, "Analysis of behavior trajectory based on deep learning in ammonia environment for fish", Sensors, 20, 4425, 2020, doi:https://doi.org/10.3390/s20164425

[31] N. D. Marom, L. Rokach, A. Shmilovici, "Using the confusion matrix for improving ensemble classifiers," in 2010 IEEE 26-th Convention of Electrical and Electronics Engineers in Israel, 555-559, 2010, doi:https://doi.org/10.1109/EEEI.2010.5662159.

[32] D. H. T. Minh, D. Ienco, R. Gaetano, N. Lalande, E. Ndikumana, F. Osman, P. Maurel, "Deep recurrent neural networks for winter vegetation quality mapping via multitemporal sar sentinel-1," IEEE Geoscience and Remote Sensing Letters, **15**(3), 464-468, 2018, doi:https://doi.org/10.1109/LGRS.2018.2794581.

[33] A. Kuswantori, T. Suesut, W. Tangsrirat, N. Nunak, "Development of object detection and classification with YOLOv4 for similar and structural deformed fish," EUREKA: Physics and Engineering, **2**, 154-165, 2022, doi: 10.21303/2461-4262.2022.002345

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 9, No. 1, 104-113 (2024)

www.astesj.com

Special Issue on Computing, Engineering and Multidisciplinary Sciences

ASTESJ

ISSN: 2415-6698

Bridging Culture and Care: A Mobile App for Diabetes Self-Care Honoring Native American Cultural Practices

Wordh Ul Hasan[1], Kimia Tuz Zaman[1], Shadi Alian[1], Tianyi Liang[1], Vikram Pandey[1], Jun Kong[1], Cui Tao[2], Juan Li[1,*]

[1]*Department of Computer Science, North Dakota State University, Fargo, 58102, USA*

[2]*School of Biomedical Informatics, University of Texas Health Science Center, Houston, 77030, USA*

ARTICLE INFO

Article history:
Received: 14 November, 2023
Revised: 06 January, 2024
Accepted: 06 January, 2024
Online: 26 January, 2024

Keywords:
Diabetes
Self-management
Native American
Cultural Sensitivity
Participatory Design
Mobile health

ABSTRACT

Diabetes presents a significant public health issue for Native Americans, exacerbated by cultural nuances often ignored by conventional healthcare. To address this, we introduce a mobile app designed with the cultural context of Native American populations in mind. The app's development followed participatory design principles, with direct input from Native American stakeholders through focus groups and interviews. The app features culturally tailored nutrition plans incorporating traditional foods, community-based support systems, and engagement with tribal health resources. An interface highlighting Native American heritage, along with gamification using cultural storytelling, aims to enhance user engagement and educational content is provided within a culturally relevant framework. This innovative integration of technology and cultural heritage in health management is anticipated to improve engagement, self-efficacy, and health outcomes for Native Americans with diabetes, serving as a blueprint for culturally sensitive health interventions.

1. Introduction

The prevalence of diabetes among Native American (NA) populations is not only a pressing public health issue but a stark representation of existing health disparities in the United States. Data from the American Diabetes Association reveal that Native Americans endure the highest age-adjusted prevalence rates of diabetes across all racial and ethnic groups nationwide, with figures nearly twice as high as those observed in non-Hispanic whites [1]. This staggering statistic underscores the urgency for specialized healthcare approaches tailored to the distinct needs of NA communities.

One of the most pronounced obstacles in diabetes management within these populations is the limited accessibility to healthcare [2]. Geographic isolation compounds this challenge, with many NA communities situated in remote areas where healthcare facilities are sparse, under-resourced, and often miles away from those in need [3]. This situation is further complicated by socio-economic barriers such as poverty, which is prevalent in many NA communities, and lower rates of health literacy that impede the uptake and implementation of diabetes management practices [4].

Moreover, the cultural dimensions of NA life — encompassing food, lifestyle, and health beliefs — are deeply ingrained and central to their identity. However, these cultural aspects are frequently overlooked in mainstream diabetes care strategies. The resultant disconnect renders many existing diabetes management applications ineffective for NA patients, as they fail to align with the patients' cultural contexts and lifestyle preferences [5].

Language and educational disparities between healthcare providers and NA patients can also lead to misunderstandings and misapplications of diabetes management plans, further inhibiting effective care. It is therefore imperative to embrace a diabetes management strategy that not only recognizes but integrates the cultural, socio-economic, and geographic particularities of NA patients.

Our study advocates for a diabetes self-management system that is culturally congruent with the lives and values of NA patients. This system is designed in close collaboration with the NA community, particularly engaging with members from the Lower Sioux Indian Community in Morton, MN, to gather requirements, iterate design solutions, and garner feedback on the prototype. Our methodology is inclusive, drawing upon the unique cultural traits of the NA population in our ontological framework, user interface, nutrition guidance, social networking features, and compilation of local diabetes care resources. This paper delineates the process of developing a mobile application that is not merely a tool but a culturally resonant companion for NA patients, offering tailored nutritional and lifestyle advice, fostering a strong sense of community, and enhancing the accessibility of local diabetes management resources.

*Corresponding Author: Dr. Juan Li, Department of Computer Science, North Dakota State University Fargo 58102, USA, j.li@ndsu.edu
This work was supported by the National Science Foundation (NSF) with award numbers: 1722913 and 2218046.

By situating the lived experiences and cultural fabric of NA individuals at the forefront, we endeavor to contribute a meaningful and functional solution to the challenges posed by diabetes management in NA communities, thereby aiming to bridge the gap between healthcare availability and cultural appropriateness.

2. Related Work

The concept of cultural appropriateness in healthcare interventions is not new; it has been thoroughly investigated by numerous scholars who advocate for the integration of cultural elements to foster positive and enduring health outcomes. The consensus is clear: to be effective, health interventions must resonate culturally with the target population. Airhihenbuwa [6] posits that cultural considerations should supersede behavioral ones, suggesting that aligning health interventions with cultural norms and values is paramount. Kreuter [7] echoes this stance, highlighting the enhanced efficacy of health programs that are culturally tailored to meet the specific needs of a given group. Further expanding on this idea, Barrera [8] devised a strategic framework for the development of culturally enriched interventions, demonstrating their superiority in bolstering health outcomes in contrast with conventional approaches.

The tripartite model of culture proposed by Asad [9], which classifies culture into knowledge, practice, and change, underscores the multifaceted influence of culture on health intervention success. This model suggests that understanding cultural knowledge, respecting cultural practices, and being open to cultural change are essential for the effectiveness of health interventions. The synthesis of these scholarly insights supports the premise that embedding cultural traditions and practices within health interventions is a pivotal strategy in diminishing health disparities and enhancing the acceptability and impact of such interventions.

Despite the known benefits of cultural tailoring, there exists a conspicuous deficiency in the availability and quality of diabetes management apps that cater to culturally diverse groups. Rodríguez [10] uncovered a scarcity of diabetes management applications available in local languages and noted a lack of credible certification or scientific backing within those that are available. Brzan [11] conducted an extensive analysis of 65 mobile self-management apps for diabetes and concluded that a mere nine met the criteria to be considered comprehensive and effective for diabetes management. This finding accentuates the need for a more rigorous and patient-centric approach to app development.

Highlighting the necessity for oversight, Sneha [12] advocates for regulation, standardization, and quality control within the burgeoning market of diabetes management applications. The current landscape is marked by an influx of apps with varying degrees of reliability and efficacy, necessitating a structured framework to assure their quality. Conversely, the research conducted by Kebede [13] offers a counterpoint, illustrating the positive impact of popular diabetes management applications like mySugr, as well as continuous glucose monitoring systems, on the self-care behaviors of individuals with type 1 and type 2 diabetes.

The intersection of culture and healthcare technology is a critical focal point in enhancing the efficacy of diabetes management applications. The influence of cultural and traditional considerations on the success of such interventions is increasingly recognized in the literature. Petersen[14] underscores the profound effect of cultural practices on diabetes self-management, suggesting that these cultural influences often have a greater impact on patient behavior than the technological features of mobile applications themselves. This insight is pivotal, as it directs attention to the nuanced ways in which cultural contexts shape health-related behaviors and choices.

Echoing this sentiment, Millan-Ferro [15] advocates for the development of interventions that are culturally resonant, particularly for marginalized groups such as the Latino community in the United States. The call is for culturally sensitive tools that are not merely translations of existing resources but are built from the ground up with cultural relevance in mind. This approach ensures that interventions are more than just medically accurate—they are culturally meaningful and thus more likely to be embraced by the communities they intend to serve.

The work of Chacko [16] sheds light on the intricate ways individuals with Type 2 diabetes navigate their healthcare options, often blending modern biomedicine with traditional practices. In Kerala, India, for instance, patients incorporate Ayurvedic medicine and folk herbal remedies alongside conventional treatments, reflecting a holistic approach to health that honors cultural heritage. This amalgamation of healthcare practices highlights the necessity of acknowledging and integrating traditional health paradigms into modern healthcare solutions.

Furthermore, Goody [17] articulates the critical role that cultural food practices play in diabetes care and education. The application of frameworks such as the Campinha-Bacote Model of cultural competency is instrumental in enabling healthcare professionals to understand and respect the intricate relationship between cultural dietary habits and diabetes management. By recognizing and incorporating these cultural food practices into diabetes care, healthcare providers can develop more effective education and intervention strategies that align with patients' lifestyles and beliefs.

These studies collectively underscore the potential for culturally tailored diabetes management applications to not only enhance their practical utility but also to resonate deeply with the daily lives and values of diverse patient populations. It is evident that there is a pressing need for such culturally informed tools, yet the current landscape reveals a scarcity of interventions that successfully integrate these critical cultural and traditional elements.

Considering this, our research is poised to fill a significant void by designing a healthcare intervention that not only acknowledges but prioritizes the cultural and traditional aspects of the Native American community. Our aim is to develop a diabetes management app that is not just functionally robust but also culturally attuned, ensuring that the Native American traditions and perspectives are interwoven into every facet of the app's design and functionality.

3. Design

The development of a culturally sensitive diabetes self-management application for Native Americans necessitates a methodological approach that is both systematic and adaptive to the unique needs of the community. This study employed a

community-based participatory research (CBPR) framework, which emphasizes collaboration and partnership with the target community at all stages of the app's development. This approach aligns with the principles of respect, relevance, and reciprocity, ensuring that the research is culturally grounded and community driven.

Our methodology integrates qualitative data collection methods to inform the design and functionality of the mobile application. The process was iterative and responsive, allowing for continuous feedback and refinement. This section details the comprehensive steps taken—from participant recruitment, design considerations to app development. Each phase was conducted with rigorous attention to maintaining cultural integrity and fostering a sense of ownership and validation within the Native American community.

The core objective of our methodological design was to create an application that is not only a technological tool for diabetes management but also a culturally resonant resource that empowers Native American users and acknowledges their unique perspectives on health and well-being.

Our engagement with the Lower Sioux Indian Community indicated a sufficient level of access to smartphones and internet services, making the deployment of a mobile health app a viable option. This was corroborated by community feedback during the initial phase of our study, where participants expressed familiarity with and access to modern mobile technology. Thus, while acknowledging the existing challenges, our study operates on the premise that a significant portion of the target population has the means to access and benefit from the app, making it a relevant tool for diabetes self-care in these communities.

3.1. Background on Native American Traditions and Perspectives

In developing our study and the subsequent mobile application for diabetes self-care, we delved deeply into the rich tapestry of Native American traditions and perspectives. The Native American communities, with their diverse and vibrant cultural practices, offer a unique lens through which healthcare solutions can be viewed and tailored.

At the heart of our narrative is an understanding that health and wellness in Native American communities are not merely physical states but are deeply intertwined with spiritual, community, and environmental well-being. These aspects, often overlooked in conventional healthcare approaches, are central to the Native American perspective on health. Food is much more than sustenance in Native American cultures; it carries historical significance and is a medium for community bonding and cultural expression. Traditional foods, preparation methods, and communal eating practices are not only nutritional choices but also a means to preserve heritage and identity. We integrated these aspects to ensure our app respects and supports these dietary traditions. The role of family and community is paramount in Native American societies. Healthcare interventions, to be effective, must recognize and leverage these social structures. Our design process involved consultations with community members to understand how family dynamics and community interactions influence health behaviors and diabetes management. Storytelling, a revered tradition in Native American cultures, serves as an educational and cultural preservation tool. By incorporating elements of storytelling and native symbolism into our app, we aimed to connect users with their heritage, making the app not just a health tool but also a medium of cultural engagement.

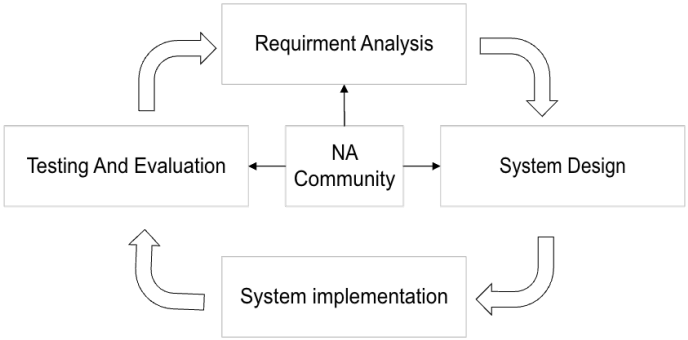


Figure 1: CBPR Process for Developing a Culturally Tailored Diabetes Management App

This narrative underscores our commitment to developing a diabetes self-management tool that is not only functionally effective but also culturally congruent. By incorporating these cultural insights, we aim to bridge the gap between technology and tradition, offering a healthcare solution that resonates deeply with Native American communities.

3.2. NA Community Based Participatory Research

Our research is deeply rooted in the principles of Community-Based Participatory Research (CBPR), an approach that ensures the development of our mobile application is closely aligned with the distinct needs and cultural context of the Native American (NA) community. Figure 1 illustrates the cyclical and collaborative nature of the CBPR process, which integrates the Lower Sioux Indian Community throughout the system development cycle.

This study has been granted ethical approval by Lower Sioux Indian Community Tribe's governance body under resolution number 16-153.

3.2.1. Requirement Analysis

In the initial phase of our study, a detailed requirement analysis was conducted, as illustrated by the top box of Figure 1. This phase was twofold: it involved individual interviews with domain experts and focused group sessions with the community.

We engaged in knowledge elicitation sessions with an assembly of domain experts, including tribal leaders from the Native American reservation community in the upper Midwest, local diabetes physicians, nutritionists specializing in diabetes care, and public health faculty. Preliminary meetings were convened to gather coarse-grained, general knowledge and set the foundation for a more nuanced understanding of the domain. The insights from these meetings were meticulously documented, serving as a vital reference for developing a domain-relevant ontology and for informing subsequent, more detailed inquiries.

Parallel to expert interviews, focused group discussions were held with members of the Lower Sioux Indian Community to explore the lived experiences of managing diabetes within their socio-cultural context. These sessions were pivotal in extracting nuanced insights into the community's cultural traditions, resource accessibility, socio-economic factors, and the specific challenges encountered in diabetes self-management. The iterative nature of these discussions is signified by the two-way arrows in Figure 1, reflecting the dynamic exchange between the research team and

the community, ensuring that their voices directly influenced the requirement analysis.

The amalgamation of knowledge from domain experts and the experiential wisdom of the community formed a comprehensive foundation for our requirement analysis. This collaborative approach ensured that the resulting diabetes self-management app was not only grounded in expert knowledge but also deeply embedded in the cultural and practical realities of the Native American users it seeks to serve.

3.2.2. Cultural Diversity in App Design

Our mobile app embraces the cultural diversity of Native American tribes and communities through an adaptable and inclusive design. The following features have been integrated:

- **Dynamic Ontology-Based Foundation**: Built on an extensible ontology, the app's knowledge base is tailored to the Lower Sioux tribe but is designed to expand to include the cultural nuances of other tribes.
- **Modular Design**: This allows for the agile integration of tribe-specific content, ensuring that the app remains relevant across a spectrum of Native American cultures.
- **Community Involvement**: We leverage a community forum chatgroup function for direct input from users, ensuring the app's cultural accuracy and responsiveness.
- **Cultural Education and Sharing**: The app promotes inter-tribal cultural exchange, enhancing understanding and unity among its users.

By incorporating these elements, our app is designed to be a culturally intelligent tool for diabetes self-management, adaptable to the varied practices and traditions of Native American communities.

3.2.3. System Design

With the input garnered from the community, we progressed to the System Design phase, which is shown on the right side of Figure 1. In this stage, we employed a participatory design methodology to mold the system's interface, functionality, and features. Our design choices were culturally informed, integrating tribal logos and NA celebrity voices to ensure the application resonates with the community's cultural heritage and preferences. The participatory nature of this phase is illustrated by the continuous loop back to the central community collaboration element in the figure, signifying an ongoing exchange of ideas and feedback.

3.2.3.1. Accessibility and Communication Strategy

Our mobile application has been meticulously crafted to facilitate effective communication and comprehension of health management strategies within the Native American community. During the design phase, input from tribal community members was sought to identify straightforward and easily understood terms, thereby avoiding intricate medical terminology.

Recognizing the potential obstacle that language can present, we have strategically reduced dependence on text. The application makes extensive use of intuitive visuals, symbols, and icons that are widely recognizable and can bridge linguistic gaps. These graphical elements are instrumental in effectively imparting health-related information.

In addition to visual aids, the application features an integrated natural voice assistant. This tool extends beyond mere navigation aid; it serves as an interactive medium for users who might favor auditory learning or those who find it more accessible. The voice assistant is programmed to communicate in a clear and comprehensible manner, catering to users with diverse literacy levels and language abilities.

In essence, our application overcomes language challenges by combining straightforward language, illustrative aids, and a user-friendly voice assistant, ensuring that all users, irrespective of their language proficiency, can easily access and understand health management strategies.

3.2.3.2. Socioeconomic Factors in App Design

Our mobile app is strategically designed to mitigate the socioeconomic challenges faced by Native American populations in managing diabetes. It provides personalized nutrition plans that are sensitive to users' financial situations and the local availability of food, ensuring that the dietary recommendations are both affordable and accessible. The app respects and supports the inclusion of traditional Native American foods, fostering sustainable eating practices deeply rooted in cultural heritage. Recognizing the variety in health literacy levels, the app conveys information using straightforward language and visual aids to ensure comprehensibility for all users. A curated database within the app guides users to local, cost-effective options for diabetes-friendly meals and healthcare services. Additionally, community forums and groups within the app serve as platforms for sharing resources and strategies, reinforcing a network of support attuned to the users' socioeconomic realities. This holistic approach to app design underscores our commitment to providing a supportive and accessible diabetes self-management tool for Native American communities.

3.2.3.3. Cultural User Centered Interface Design

Our approach adopts a Cultural-User-Centered Design (CUCD) approach to construct an interface meticulously crafted to resonate with Native American (NA) users. This design methodology is deeply rooted in cultural sensitivity, ensuring that the interface serves not just as a tool, but also as a cultural bridge. We've integrated a spectrum of culturally significant elements, each chosen for its profound connection to the NA community

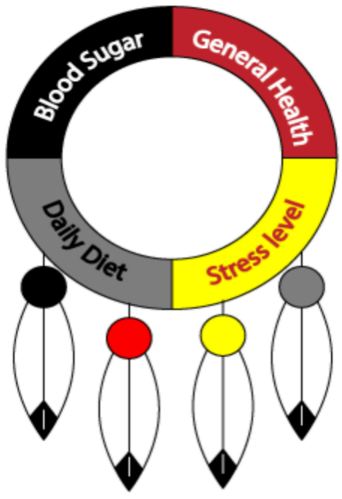


Figure 2: Tribal Emblem as UI Component

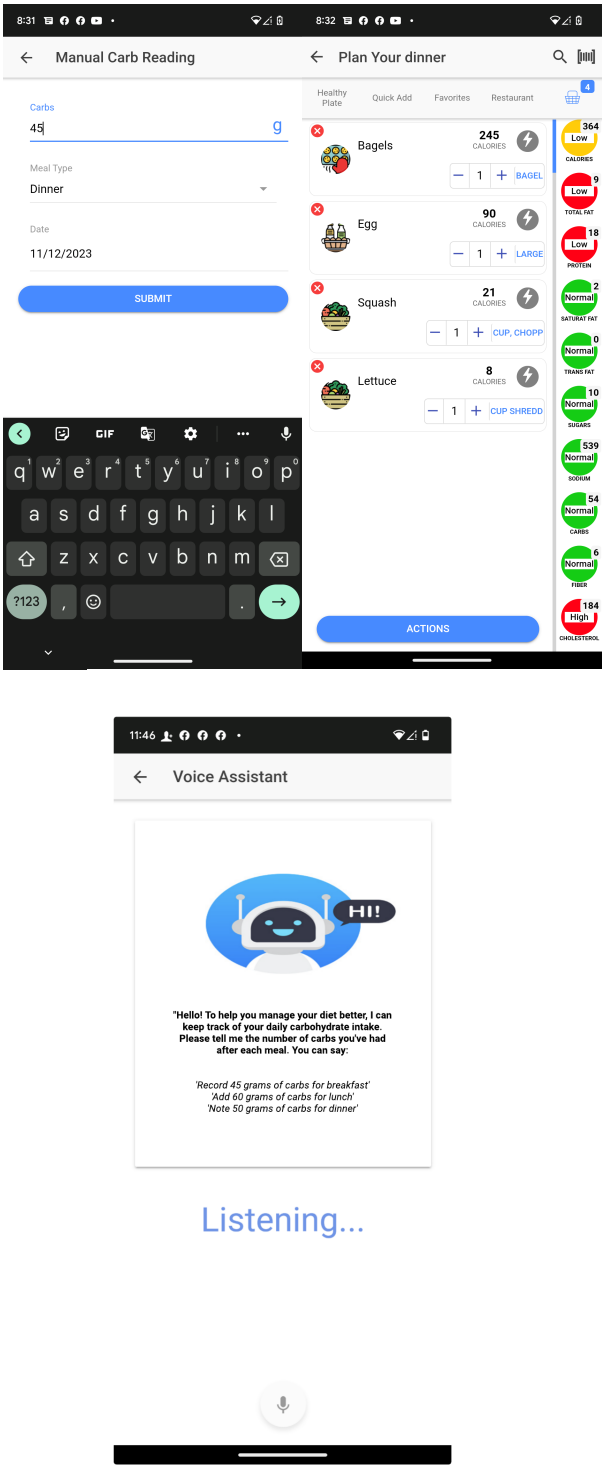


Figure 3: (a) Manual carb input Interface. (b) Calculated carb input through meal selection Interface (c) Voice assistant based Carb input

- *Tribal Emblem Integration*

The integration of the tribal emblem within our user interface serves as a cornerstone of our culturally tailored design strategy. This emblem shown in figure 2, segmented into four interactive quadrants, does more than provide aesthetic value; it functions as a dynamic interface element. Each quadrant is meticulously crafted to represent an essential component of the system's functionality and to serve as a navigational aid, transforming into buttons that, upon interaction, reveal detailed logs and further information pertaining to the system's various modules. The emblem itself is deliberately chosen for its cultural significance, embodying the rich heritage and shared history of the community. It acts as a visual anchor, grounding the user experience in familiar iconography that resonates on a personal level with the community members. Its inclusion is a thoughtful acknowledgment of the community's values, traditions, and identity, intended to engender a sense of belonging and ownership over the technological experience.

This emblematic interface promotes accessibility and ease of use, with each interactive portion designed to guide users through the system's capabilities intuitively. For instance, selecting one quadrant might reveal insights into the daily diet analysis data, while another might delve into blood sugar logs and trends over time. This method of organization not only simplifies navigation but also aligns with the community's cognitive patterns and cultural context, making the system's operation inherently intuitive for the intended NA users. It's a harmonious blend of function and tradition, ensuring that the system is not only efficient but also culturally congruent, facilitating a seamless and resonant user experience.

- *Celebrity Vocal Engagements*

To enrich the interface's auditory experience, we have procured voice recordings from Native American luminaries and venerated figures, like Evan Adams, known for his pivotal portrayal of Thomas Builds-the-Fire in the seminal film "Smoke Signals." His voice serves as a thread weaving through our interface, providing guidance and feedback, imbuing the user experience with authenticity and a deeper cultural linkage. This initiative goes beyond mere functionality—it is a homage to a voice that echoes the collective memories and experiences of the NA community.

- *Linguistic Tailoring*

Acknowledging the diverse linguistic capabilities within the NA communities, our interface offers customizable language options. It allows users to select their preferred language and adjust the complexity of the language to match their literacy level, ensuring accessibility and comprehension for all. Insights into this aspect of our design process are detailed in publication [18].

- *Multi Modal Interaction*

Our application offers a multifaceted interface designed to accommodate the distinctive communication styles within the community. As depicted in Figure 3(a), the app includes a straightforward graphical user interface for manual carbohydrate entry. This tactile method is uncomplicated, facilitating quick and direct input of carbohydrate consumption through simple screen interactions.

In another approach, demonstrated in Figure 3(b), the app leverages a sophisticated algorithm to estimate carbohydrate intake from meal selections. Users can describe their meals, and the app intelligently calculates and logs the carbohydrate content, streamlining the tracking process.

Additionally, Figure 3(c) introduces the convenience of a voice assistant. This modality allows users to verbally log their

carbohydrate intake, accommodating those who might find manual entry cumbersome or prefer the immediacy of speech.

These three modalities reflect our dedication to inclusive design, offering both visual and auditory options to manage health data in the most intuitive way possible. By embracing diverse methods of interaction, the app not only improves user experience but also respects the cultural and individual preferences of our wide-ranging user demographic. This inclusive approach positions our application as a comprehensive tool for managing diabetes through user-friendly carbohydrate tracking.

- *Cultural Narratives and Context*

The interface weaves in narratives, traditions, and practices that are the fabric of NA cultural identity. These elements are not mere adornments; they serve to create a digital experience that is culturally congruent, fostering a sense of community and relevance.

The CUCD methodology champions the idea that an interface should be a cultural artifact, one that NA users can see themselves reflected in. By weaving these culturally rich threads into the interface, we aim to deliver an experience that is not only functionally superior but also culturally enriching and empowering for NA users.

3.2.3.4. Cultural Tailored System Design

In the development of our system, we prioritize a nuanced approach tailored to Native American (NA) users, focusing on elements that cater to their unique dietary needs and cultural practices. This includes:

Our system is finely tuned to offer personalized nutrition advice, considering the distinct dietary preferences and needs of NA users. We understand the importance of incorporating traditional and culturally significant foods in our recommendations. This includes a special emphasis on game food, which is not only a dietary staple but also holds deep cultural significance in many NA communities.

Recognizing the diverse culinary landscape across NA communities, our app considers the availability of local foods. This involves integrating traditional ingredients and dishes that are integral to various NA cultures, ensuring that our nutritional guidance is not only health-focused but also culturally resonant.

The socioeconomic context of NA communities is a critical aspect of our system design. We strive to make our nutritional recommendations affordable, considering the varying economic backgrounds of our users. This ensures that our guidance is realistic and accessible, not just theoretically ideal.

Our app goes beyond basic nutritional advice by embracing the rich cultural heritage of NA communities. We respect and incorporate traditional food practices and preferences, ensuring that our recommendations do not conflict with cultural norms and practices. Our app also acknowledges the diverse cultural practices and beliefs of tribal members, offering personalized lifestyle recommendations. For instance, we suggest physical activities like traditional dancing or nature walks on ancestral trails, which are not only enjoyable but also culturally significant. These suggestions honor the traditions and enhance the physical well-being of NA users.

To bridge the gap between nutritional advice and practical meal options, we analyze local restaurant menus. This analysis is geared towards ensuring that the food options available to NA users at local eateries are in line with their dietary requirements and cultural preferences. Previously our work [19–21]on restaurant menu and personized meal planning describes how health conditions can be used to recommend foods. This feature not only supports healthy eating habits but also promotes cultural integrity and relevance.

Understanding the centrality of family and community in NA culture, our app encourages social activities that strengthen these bonds. This could include family-based support systems or participation in cultural events and gatherings. By doing so, we aim to foster a supportive environment that is conducive to positive health outcomes and sustained user engagement.

Our social network feature is designed with a deep understanding of the importance of community connections in NA culture. It enables users to connect, share experiences, and support each other in their journey of diabetes self-management. This virtual community space is a platform for exchanging ideas, sharing cultural practices pertinent to diabetes management, and providing emotional support. It's more than a network; it's a communal space that strengthens the sense of belonging and encourages positive health behaviors.

We facilitate easy and meaningful communication between NA users and healthcare professionals who are attuned to their cultural context. This feature ensures that healthcare delivery is culturally competent, addressing the unique needs and preferences of NA users. It’s a bridge connecting users with healthcare providers in a way that respects and understands their cultural background.

The app provides educational content specifically tailored to the cultural sensitivities and needs of NA users. Covering various aspects of diabetes management, including diet, physical activity, medication, and self-monitoring, this content is not just informative but also relatable. It's designed to be a culturally relevant guide in managing diabetes effectively.

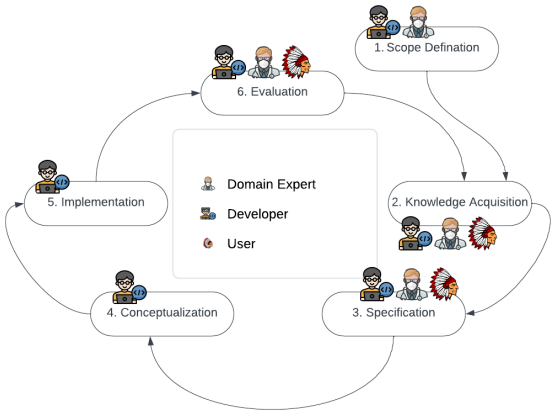


Figure 4: Ontology design and development life cycle

By incorporating features that facilitate effective communication and tailored education, our app plays a pivotal role in promoting culturally competent healthcare. It respects and acknowledges the influence of cultural context on health beliefs

and behaviors, aiming to bridge any gaps between healthcare providers and users.

Overall, our system is meticulously designed to provide a relevant, inclusive, and culturally sensitive user experience for NA users. It aims to promote healthy lifestyle changes, foster community support, and ultimately contribute to improved health outcomes within the NA community. Our approach is not just about managing diabetes; it's about embracing and integrating the rich cultural heritage of NA users into every aspect of their healthcare journey.

3.3. NA Biocultural Profile Ontology Development

In designing self-management solutions for diabetes within Native American (NA) populations, it is essential to acknowledge and integrate their distinct biocultural attributes. Our approach introduces a sophisticated biocultural user profile ontology, NAOnto, tailored to encapsulate the intricate tapestry of biological, cultural, socio-economic, and environmental factors that characterize NA diabetes patients. This ontology is a rich repository of data points, including but not limited to cultural influences, geographic locations, prevalent health beliefs and attitudes, varying literacy and numeracy levels, physical capabilities, familial support structures, and economic conditions. Together, these variables form a holistic view of a patient's well-being, firmly rooted in their unique biocultural landscape.

The creation of NAOnto was orchestrated through a structured six-phase process, as depicted in Figure 4. Beginning with scope definition, our team delineated the boundaries and objectives of the ontology, ensuring alignment with the specific self-management needs of NA diabetes patients. Subsequent knowledge acquisition involved meticulous gathering of information from a host of resources, including direct inputs from the community and domain experts. Specification followed, where the acquired knowledge was translated into a detailed framework, leading to the conceptualization phase, which established the ontology's foundational structures and relationships.

In the course of development, we have collaborated with the Community Health & Social Service Directors, the Clinic CEO, and two of their assistants from the Lower Sioux Indian Community. While these professionals did not participate in the broader user validation process, they played a crucial role during the design phase. Their insights and contributions were pivotal in tailoring our app's design to be culturally sensitive and to address the specific healthcare needs of the Native American community.

To ensure a wider relevance and applicability, a survey-based user study with nine participants was conducted across Native American communities in North Dakota and Minnesota.

The implementation phase saw the ontology being constructed and populated with data, while the evaluation phase subjected it to rigorous community and expert review, allowing for validation and calibration against real-world scenarios. This cyclical development process—iterative in nature—permitted ongoing refinement and updating of the ontology, ensuring that it remained dynamic and reflective of the evolving understanding of the NA community's

needs and the latest medical insights. The result is a dynamic, living model that is as responsive as it is comprehensive.

The utility of NAOnto extends beyond a mere database; it is a critical tool for crafting personalized diabetes self-management strategies that honor the biocultural context of NA patients. Through NAOnto, we enable the creation of interventions that are not only medically precise but also culturally consonant, fostering greater acceptance and effectiveness.

Figure 5 presents a partial view of the NAOnto that is defined for this research. Further details of the ontology's structure and its application within our CBPR framework are elaborated in the accompanying publications [22, 23], where we delve into the granularities of NAOnto and its operational impact.

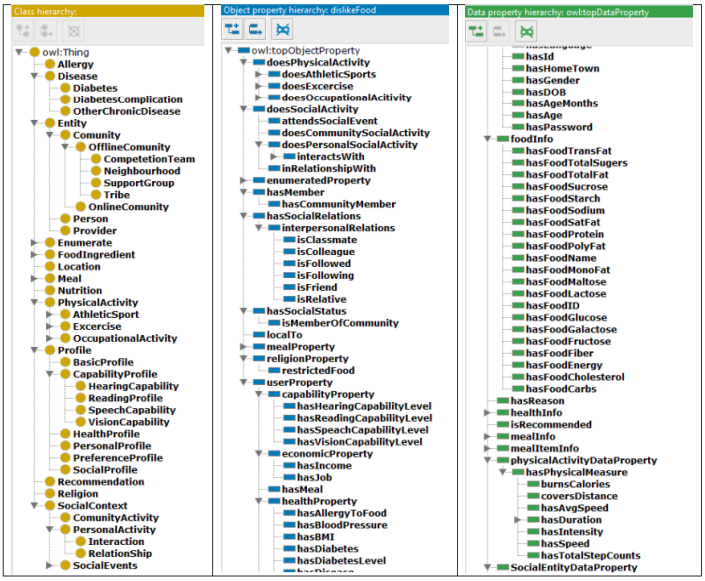


Figure 5: Partial view of ontology NAOnto

4. Implementation

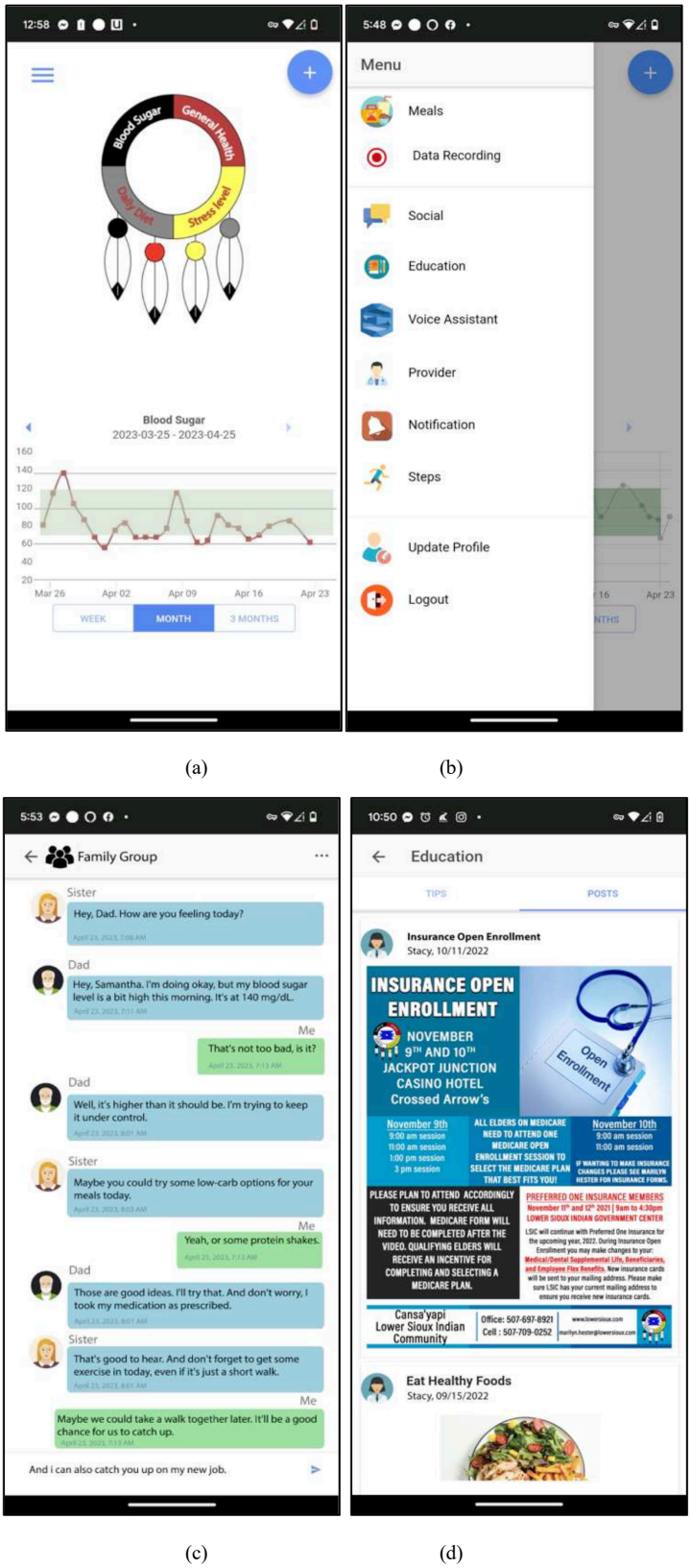
The development and deployment of our diabetes self-management mobile app leveraged the capabilities of the Ionic framework and Javalin server, optimizing it for both functionality and security.

Ionic was selected for its exceptional cross-platform capabilities, allowing us to create a single codebase that functions seamlessly on both iOS and Android platforms. This choice significantly streamlined the development process, ensuring a consistent user experience across different devices.

Javalin, a lightweight Java and Kotlin web framework, was employed for handling HTTP requests between the app and our backend server. It efficiently routes these requests to the appropriate handlers, ensuring swift and accurate data processing.

To safeguard user data, all communications between the mobile app and the Javalin server were encrypted using HTTPS protocols. Additionally, data stored on the SQL server was secured using RSA encryption, reflecting our commitment to data security and user privacy.

The app features several unique interfaces, each designed with specific user needs in mind. In figure 6(a) shows the integration of the Lower Sioux tribal logo into our diabetes management system offers a culturally sensitive and intuitive interface. The health



(a) (b)

(c) (d)

Figure 6: Screenshots of the mobile app.

wheel, divided into sections representing blood sugar levels, general health, stress levels, and diet, allows users to input and track their data efficiently. In figure 6(b) menu encompasses a broad range of features including meal recommendations, health data recording, social networking, educational resources, and direct connectivity with healthcare providers. It's a holistic state of tools designed to facilitate diabetes self-management. Family Group Communication Feature is shown in figure 6(c) Acknowledging the importance of family in NA culture, this feature enables family members to support each other in diabetes management, promoting a collaborative care approach. Culturally feature enables family members to support each other in diabetes management, promoting a collaborative care approach. Culturally Tailored Education and Events is shown in figure 6(d). This aspect of the app keeps users informed about cultural events and educational resources, fostering community engagement and participation.

Each app function underwent rigorous testing to ensure optimal performance and usability. We conducted multiple use case studies to assess the feasibility and effectiveness of features such as meal planning and restaurant menu recommendations. Specific app functions like personalized meal planning and restaurant menu screening were evaluated through user studies. These studies provided insights into the usability and impact of these features on the self-management of diabetes among NA users.

While individual functions have been evaluated, a comprehensive assessment by the tribal community is pending. We plan to engage with the Lower Sioux community to gather feedback and assess the overall effectiveness of the app. This evaluation will offer vital insights into the app's acceptability, practicality, and its impact on health outcomes in the NA community.

To ensure sustained engagement with our mobile app, we have developed a comprehensive strategy focusing on personalization, gamification, community involvement, and continuous improvement. Personalization is key; the app learns and adapts to user interactions, offering tailored recommendations to maintain relevance. Gamification techniques, such as challenges and rewards, are introduced to engage users and encourage consistent app usage. As part of our cyclical Community-Based Participatory Research (CBPR) process, depicted in Figure 1, we commit to regular content updates, providing fresh information and resources that invite users to keep engaging with the app. Community features like forums and support groups are incorporated to cultivate a supportive network, enhancing user motivation through shared experiences and successes. Additionally, we have established a feedback mechanism that enables users to contribute to the app's evolution, ensuring it continuously aligns with their preferences and needs. This strategic approach aims to integrate the app into users' daily routines, making it a valuable, long-term asset in diabetes self-management.

In summary, our diabetes self-management app combines technological innovation with cultural sensitivity. It's designed not just as a tool for health management, but as a culturally attuned companion for Native Americans. Our goal is to empower users through a blend of modern technology and respect for traditional values and practices, thereby facilitating effective diabetes management in the NA community.

5. Conclusion

In the concluding section of our paper, we emphasize the successful design and development of a mobile application that is

specifically tailored to the needs and cultural nuances of Native Americans (NAs) with diabetes. Our approach, characterized by a deeply collaborative and iterative design process, incorporated valuable insights from the Lower Sioux Indian Community. This collaboration was instrumental in shaping an app that is not only technologically sound but also deeply respectful and reflective of the unique cultural traditions, local resources, and socio-economic realities of the NA population.

The app offers bespoke dietary recommendations, aligning with the traditional dietary habits and available local food resources of NAs. This personalized approach ensures that the guidance is not only health-conscious but also culturally appropriate. Acknowledging the integral role of family in NA communities, the app incorporates features that enable and encourage family involvement in diabetes management, promoting a collective approach to health and wellness. The app provides users with information on local resources, including traditional foods and community events, facilitating a connection between users and their local environment. By embedding cultural elements into the app's interface and functionality, we ensure that the technology resonates with the users on a cultural level, enhancing engagement and relevance.

Moving forward we aim to conduct extensive user studies within the Lower Sioux community. These studies will involve a larger sample size to ensure a more diverse and representative feedback. A key focus of our upcoming studies will be to evaluate the app's effectiveness in improving health outcomes among NAs with diabetes. This will involve tracking various health indicators and user engagement metrics over time. The feedback obtained from these comprehensive studies will be pivotal in further refining the app. Our goal is to continuously evolve the app's features and usability to better align with the needs and preferences of NA users. While our initial focus has been on the Lower Sioux community, we envisage expanding the reach of the app to other NA communities, adapting it to different tribal traditions and needs.

In summary, our paper not only details the creation of a culturally tailored mobile app for NAs with diabetes but also sets a roadmap for its future development and enhancement. By continuing to engage with the community and rigorously assessing the app's impact, we aim to create a dynamic tool that not only supports diabetes management but also fosters a deeper connection between health technology and cultural identity. This endeavor reflects our commitment to developing solutions that are not only technologically advanced but also culturally empathetic and responsive to the unique needs of diverse populations.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors express their gratitude to Dr. Donald Warne, Darin Prescott, Stacy Hammer and the Lower Sioux Indian community for their invaluable support, collaboration, valuable suggestions, and feedback. Their contribution has been instrumental in the development and improvement of our app, and we are thankful for their ongoing partnership and commitment to improving healthcare outcomes for Native American communities affected by diabetes. This work was supported by the National Science Foundation (NSF) with award numbers: 1722913 and 2218046.

References

[1] National Diabetes Statistics Report | Diabetes | CDC, Jan. 2024.
[2] Disparities | Fact Sheets, Jan. 2024.
[3] A.H. Ariel-Donges, E.L. Gordon, B.N. Dixon, A.J. Eastman, V. Bauman, K.M. Ross, M.G. Perri, "Rural/urban disparities in access to the National Diabetes Prevention Program," Translational Behavioral Medicine, **10**(6), 1554–1558, 2020, doi:10.1093/TBM/IBZ098.
[4] M. Sarche, P. Spicer, "Poverty and health disparities for American Indian and Alaska native children: Current knowledge and future prospects," Annals of the New York Academy of Sciences, **1136**, 126–136, 2008, doi:10.1196/ANNALS.1425.017.
[5] R. Struthers, F.S. Hodge, B. Geishirt-Cantrell, L. De Cora, "Participant Experiences of Talking Circles on Type 2 Diabetes in Two Northern Plains American Indian Tribes," Http://Dx.Doi.Org/10.1177/1049732303256357, **13**(8), 1094–1115, 2003, doi:10.1177/1049732303256357.
[6] C.O. Airhihenbuwa, C.L. Ford, J.I. Iwelunmor, "Why Culture Matters in Health Interventions," Http://Dx.Doi.Org/10.1177/1090198113487199, **41**(1), 78–84, 2013, doi:10.1177/1090198113487199.
[7] M.W. Kreuter, S.N. Lukwago, D.C. Bucholtz, E.M. Clark, V. Sanders-Thompson, "Achieving Cultural Appropriateness in Health Promotion Programs: Targeted and Tailored Approaches," Http://Dx.Doi.Org/10.1177/1090198102251021, **30**(2), 133–146, 2003, doi:10.1177/1090198102251021.
[8] M. Barrera, F.G. Castro, L.A. Strycker, D.J. Toobert, "Cultural adaptations of behavioral health interventions: a progress report," Journal of Consulting and Clinical Psychology, **81**(2), 196–205, 2013, doi:10.1037/A0027085.
[9] A.L. Asad, T. Kay, "Toward a multidimensional understanding of culture for health interventions," Social Science & Medicine, **144**, 79–87, 2015, doi:10.1016/J.SOCSCIMED.2015.09.013.
[10] A. Quevedo Rodríguez, A.M. Wägner, "Mobile phone applications for diabetes management: A systematic review," Endocrinología, Diabetes y Nutrición, **66**(5), 330–337, 2019, doi:10.1016/J.ENDINU.2018.11.005.
[11] P.P. Brzan, E. Rotman, M. Pajnkihar, P. Klanjsek, "Mobile Applications for Control and Self Management of Diabetes: A Systematic Review," Journal of Medical Systems, **40**(9), 1–10, 2016, doi:10.1007/S10916-016-0564-8/METRICS.
[12] S. Sneha, S. Thalla, I. Rischie, H. Shahriar, "Health Internet Technology for Chronic Conditions: Review of Diabetes Management Apps," Jan. 2024, doi:10.2196/17431.
[13] M.M. Kebede, C.R. Pischke, "Popular diabetes apps and the impact of diabetes app use on self-care behaviour: A survey among the digital community of persons with diabetes on social media," Frontiers in Endocrinology, **10**(MAR), 135, 2019, doi:10.3389/FENDO.2019.00135/BIBTEX.
[14] F. Petersen, "Impact of Culture on the Adoption of Diabetes Self-Management Applications: Cape Flats, South Africa," ArXiv.Org, 2021.
[15] A. Millan-Ferro, A.E. Caballero, "Cultural approaches to diabetes self-management programs for the Latino community," Current Diabetes Reports, **7**(5), 391–397, 2007, doi:10.1007/S11892-007-0064-9/METRICS.
[16] E. Chacko, "Culture and therapy: complementary strategies for the treatment of type-2 diabetes in an urban setting in Kerala, India," Social Science & Medicine, **56**(5), 1087–1098, 2003, doi:10.1016/S0277-9536(02)00105-3.
[17] C.M. Goody, L. Drago, "Using Cultural Competence Constructs to Understand Food Practices and Provide Diabetes Care and Education," Diabetes Spectrum, **22**(1), 43–47, 2009, doi:10.2337/DIASPECT.22.1.43.
[18] R. Hendawi, S. Alian, J. Li, "A Smart Mobile App to Simplify Medical Documents and Improve Health Literacy: System Design and Feasibility Validation," Jan. 2024, doi:10.2196/35069.
[19] B. Maharjan, J. Li, J. Kong, C. Tao, "Alexa, What Should I Eat? : A Personalized Virtual Nutrition Coach for Native American Diabetes Patients Using Amazon's Smart Speaker Technology," 2019 IEEE International Conference on E-Health Networking, Application and Services, HealthCom 2019, 2019, doi:10.1109/HEALTHCOM46333.2019.9009613.
[20] M.S.A.T. Zadeh, J. Li, S. Alian, "Personalized Meal Planning for Diabetic Patients Using a Multi-Criteria Decision- Making Approach," 2019 IEEE International Conference on E-Health Networking, Application and

Services, HealthCom 2019, 2019, doi:10.1109/HEALTHCOM46333.2019.9009593.

[21] W. Ul Hasan, K. Tuz Zaman, M.S.A.T. Zadeh, J. Li, "Eat This, Not That! - a Personalised Restaurant Menu Decoder That Helps You Pick the Right Food," 2022 IEEE International Conference on E-Health Networking, Application and Services, HealthCom 2022, 43–48, 2022, doi:10.1109/HEALTHCOM54947.2022.9982770.

[22] V. Pandey, J. Li, S. Alian, "Evaluation and Evolution of NAOnto - An Ontology for Personalized Diabetes Management for Native Americans," 2021 7th International Conference on Computer and Communications, ICCC 2021, 1635–1641, 2021, doi:10.1109/ICCC54389.2021.9674339.

[23] J. Li, S. Alian, "Design and development of a biocultural ontology for personalized diabetes self-management of American Indians," 2018 IEEE 20th International Conference on E-Health Networking, Applications and Services, Healthcom 2018, 2018, doi:10.1109/HEALTHCOM.2018.8531107.

Copyright: This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 8, No. 6, 107-119 (2023)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Enhancing Cloud Security: A Comprehensive Framework for Real-Time Detection, Analysis and Cyber Threat Intelligence Sharing

Fazalur Rehman[*,1], Safwan Hashmi[2]

[1]*Department of Cybersecurity, Air University, PAF Complex, E-9, Islamabad, Pakistan*

[2]*Department of Cybersecurity, Riphah International University, Evacuee Trust Complex, F-5, Islamabad, Pakistan*

ARTICLE INFO

Article history:
Received: 30 September, 2023
Accepted: 05 December, 2023
Online: 30 December, 2023

Keywords:
Cloud Security
Virtual Machine Introspection
Cloud Computing
Incident Management
Cyber Threat Intelligence
Incident Response
Threat Sharing
Cloud Infrastructure Security

ABSTRACT

Cloud computing has emerged as a pivotal component of contemporary IT systems, affording organizations the agility and scalability required to meet the ever-changing demands of business. However, this technological evolution has introduced a new era of cybersecurity challenges, as attackers employ increasingly sophisticated strategies to breach cloud networks. Such breaches can have far-reaching consequences, including data loss, financial repercussions, reputational damage, and legal liabilities. In response to these challenges, developing a robust security framework is imperative for effectively safeguarding cloud infrastructure. This paper proposes a novel Hypervisor-based Virtual Machine Introspection (HVMI) for real-time detection and runtime forensic analysis of cyberattacks targeting cloud platforms. The framework proposed comprises several essential components, including a forensic application empowered by Virtual Machine Introspection (VMI) for real-time memory analysis, a centralized Cloud Forensic Tool (CFT) portal for streamlined incident management, and a data transmission and integration web service. Notably, this framework is founded upon a commitment to continuous optimization and enhancement. This iterative process is facilitated through a collaboration approach. It involves fine-tuning various aspects of the framework, such as adjusting settings for VMI, refining criteria for classifying incidents, and updating security controls. Enhancing the forensic application represents a proactive measure aimed at improving the efficiency and effectiveness of VMI and forensic analysis capabilities. The iterative refinement process integrates incident analysis, threat intelligence infusion, and collaborative efforts to adapt to emerging challenges. This dynamic approach fosters a flexible security posture capable of detecting, analyzing, and responding to emerging attacks within cloud platforms. In summary, the proposed framework embodies a comprehensive approach to cloud security, integrating advanced technology with continuous refinement to protect cloud infrastructure, mitigate risks, and navigate the ever-evolving cybersecurity threat landscape effectively.

1 Introduction

Cloud computing has emerged as a fundamental element within modern computing systems, providing organizations with the capacity for on-demand resources and adaptable scalability. It is the on-demand delivery of IT resources. Organizations instead of buying or owning their physical data centers and servers depend on cloud service providers. Securing IT infrastructure is a major challenge nowadays. Within the domain of cloud computing, cybersecurity is a critical concern, given that data breaches have far-reaching impacts on organizations. This paper presents a novel technique and framework based on virtual machine introspection to detect and perform runtime forensic analysis of attacks on cloud platforms. This paper is an extension of the work originally presented at the IEEE 3rd International Conference on Artificial Intelligence (ICAI) 2023 [1].

With the expansion of cloud infrastructure adoption, there has been a corresponding amplification in the array of techniques and approaches employed by malicious entities to initiate attacks targeting these network environments [2]. Such attacks present a substantial threat to the confidentiality, integrity, and availability of cloud-based systems, and can have significant consequences for organizations depending on these systems for their operational functions [3]. These breaches can lead to the exposure of sensitive information, financial

[*]Corresponding Author: Fazalur Rehman, Air University, PAF Complex, E-9, Islamabad, Pakistan, 181219@students.au.edu.pk

losses, harm to an organization's reputation, and legal consequences [4]. To mitigate these risks, organizations need to implement effective cybersecurity measures to ensure the security of their cloud infrastructure [5].

Conventional methods employed for securing cloud infrastructures, such as virtual-level segregation, intrusion detection prevention systems (IDS/IPS), cloud access and security brokers (CASB), and endpoint detection & response, frequently prove inadequate in countering the progressively sophisticated techniques deployed by attackers targeting cloud networks [6]. Even when efforts are made to analyze or thwart these attacks, attackers often exhibit adaptability, which enables them to resist detection and mitigation. Furthermore, these protective measures often operate within virtualized environments shared across interconnected networks, rendering them susceptible to deceptive tactics, insider threats, and network-level attacks [7, 8].

In the event of a security breach, the proposed HVMI solution notifies the cloud service provider while concurrently initiating a forensic analysis to identify the root cause and extent of the breach. This real-time detection and analysis capability inherent in the HVMI tool represents a pivotal enhancement to cloud system security, offering substantial protection against the adverse consequences stemming from data breaches. Furthermore, by incorporating this solution within a web service framework, it attains cross-platform compatibility, irrespective of the underlying hardware and software infrastructure. In a bid to maximize the utility of this research, we have conceptualized a web portal where instances of attack patterns can be uploaded. This portal serves as a means to disseminate information to security organizations and cloud service providers globally, effectively notifying them of prevailing cloud-based attack trends and facilitating the formulation of defensive strategies against such threats. This research endeavor not only contributes to the collective knowledge in the field but also holds practical significance for industry stakeholders, security professionals, and cloud service providers. It offers a novel and real-time approach to detect and analyze attacks within cloud environments, with the overall objective of minimizing their adverse impact on the confidentiality, integrity, and availability of cloud systems.

1.1 Research Contributions

1. The HVMI solution comprises a client application, specifically a forensic application, that operates on the cloud service provider's host. Its primary function is to identify and mitigate the impact of security breaches, making it a valuable resource for organizations seeking to shield themselves from the cyberattacks.

2. Within this framework, malicious artifacts are systematically gathered from virtual machines (VMs), yielding critical insights into the techniques and methods employed by attackers. These artifacts are subsequently transformed into a comprehensible, organized, and shareable format. HVMI adopts the Structured Threat Information eXpression (STIX) standard [9] to generate consistent threat details. These standardized threat details are invaluable to security organizations as they facilitate the development of defensive strategies against specific types of cyberattacks in the cloud, thereby enhancing the overall security posture of cloud systems.

3. The integration of a web service framework ensures cross-platform compatibility, rendering the HVMI solution independent of the underlying hardware and software. Additionally, the accompanying web portal serves as a dedicated platform for the dissemination of cyber threat intelligence. It caters to the needs of both security organizations and cloud service providers, facilitating the exchange of vital information to bolster cybersecurity efforts in cloud environments.

1.2 Organization of the paper

The paper is organized as: section 2 provides literature review. Section 3 provides details of the proposed framework. Section 4 provides details of results, and attack simulations. Section 5 is conclusion & future directions.

2 Review of the Literature

This section offers a comprehensive summary of the existing research within the selected domain. It serves to elucidate the present state of knowledge regarding the subject matter, encapsulating the key findings and insights derived from prior studies. Furthermore, it identifies gaps in existing research or areas where current research falls short, signifying opportunities for further investigation and research.

Virtualization stands as a pivotal element within the IT industry, greatly augmenting management capabilities and unlocking the full potential of hardware infrastructure. In essence, Virtualization provides the capability to the impression of multiple physical systems of a single system. It provides a virtual version of the underlying hardware platform, storage media, network devices etc. Virtualization has numerous advantages over traditional physical system architectures, encompassing cost-efficiency, reduced hardware resource requirements, and optimal utilization of available hardware resources. It enables the utilization of hardware to its maximum potential, a feat unattainable within the confines of conventional infrastructure [10]. A hypervisor, also referred to as a virtual machine manager (VMM), is a software application designed to facilitate the creation and administration of virtual machines (VMs) on a physical host. These hypervisors introduce a layer of abstraction between the physical hardware and the virtual machines, enabling multiple VMs to efficiently utilize the resources of a single physical machine. Hypervisors are of paramount importance in cloud computing settings, as they assume a central role in the establishment and supervision of numerous virtual machines (VMs) on a single physical host. This functionality endows organizations with the flexibility to dynamically adjust their computational resources in tandem with fluctuations in demand, thereby optimizing resource allocation in the cloud environment. There are two primary categories of hypervisors:

Type I (or native or bare metal) hypervisors: These hypervisors operate directly on the host's hardware, creating a virtualization layer that interfaces between the hardware and the operating sys-

tem [11]. Prominent examples of Type 1 hypervisors encompass VMware ESXi, KVM, Microsoft Hyper-V, and Xen.

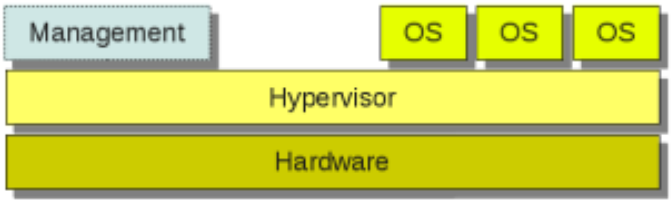


Figure 1: VMM Type I [12]

Type II (or hosted) hypervisors: These hypervisors run atop a host operating system, delivering virtualization capabilities for guest operating systems. Well-known instances of Type 2 hypervisors include VMware and VirtualBox.

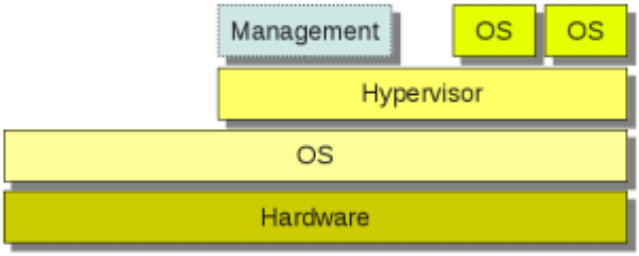


Figure 2: VMM Type II [12]

There has been a growing interest in the adoption of hypervisor-based introspection techniques for the identification and analysis of security breaches within cloud platforms. Notable contributions in this field include the work of Brian [13] have developed a toolbox known as Virtual Contemplation for Xen (VIX). This toolbox has demonstrated its capacity to conduct live examinations and gather volatile data from virtual machines. Their approach involves pausing the virtual machine, extracting the necessary data, and then resuming the virtual machine. In a similar vein, another toolset akin to VIX is XenAccess, created and maintained in [14]. XenAccess is designed to offer a library of functions for constructing a monitoring architecture. This includes employing virtual memory introspection to monitor applications and access the memory state of the virtual machine. This architecture has been leveraged to implement Virtual Machine Introspection (VMI), enabling forensic investigations within virtual machine environments.

In [15], the author introduced Memory Forensics Analysis (MFA). They achieved this by deploying an in-guest agent within the virtual machine (VM) to collect detailed information regarding the processes running on that VM. In a similar vein [16], the author proposed a comprehensive security framework that encompasses both a network-based intrusion detection system (NIDS) and a virtual machine introspection-based intrusion detection system (VMIIDS). The NIDS acts as the initial line of defense, monitoring network traffic before it reaches the virtualization layer. Meanwhile, the VMI-IDS operates at the hypervisor level, where it focuses on detecting intrusions occurring within the VMs themselves. This dual-layered approach enhances the security posture of cloud environments by addressing threats at different levels of the infrastructure.

In [17], the author introduced an anomaly-based detection system designed for the detection of malware at the hypervisor level. Nonetheless, the effectiveness of this approach is circumscribed, and it may not prove sufficiently robust against highly sophisticated and complex attacks. In [18], the author advocated for the utilization of hypervisor-based introspection to identify malware by analyzing information gathered from guest machines and network-level devices. Although this approach achieved a detection accuracy of approximately 90 percent under optimal conditions, its performance declined when confronted with more complicated and advanced attack vectors.

In [19], the author introduced an innovative hardware-based monitoring system named HyperMon, designed for the detection and surveillance of attacks within cloud platforms. HyperMon was implemented within the Xen hypervisor and used machine learning techniques to inspect low-level hardware data at the virtual machine manager (VMM) layer. This process was employed to construct statistical models for programs. In [20], the author put forward a Virtual Machine Introspection (VMI) approach, focusing on anomaly-based detection of keyloggers. This approach accomplished by tracking a spectrum of events, including memory reads and writes, interrupts, and network logs. Subsequently, the collected data underwent analysis employing an artificial immune system (AIS)-based intrusion detection system (IDS) to identify anomalies. However, it's noteworthy that this approach has limitations as it is primarily applicable to Linux-based systems and virtual machines.

In [21], the author proposed a security framework for cloud environments, centered around Virtual Machine Introspection (VMI). This framework is geared towards monitoring the active processes within virtual machines (VMs) by tracking system call traces. The method employs Nitro, a hardware-based tool, to capture these system call traces, which are subsequently relayed to a centralized analyzer. This analyzer utilizes a classification model to distinguish between legitimate and malicious processes. If a process is classified as malicious, an alert is generated. In alignment with the notion that security and monitoring should be implemented at the hypervisor level. In [22], the author proposed a malware analysis technique employing VMI. Their research, however, predominantly focuses on the domain of malware analysis, virtualization techniques, and specific malware families.

The existing solutions discussed in the literature review, while effective at detecting certain types of attacks like rootkits, keyloggers, or malware, exhibit limitations in their capacity to defend against emerging cybersecurity threats. These limitations arise due to their inability to comprehensively address the evolving tactics, techniques, and procedures employed by sophisticated attackers. Furthermore, many of these solutions lack a specific focus on the unique challenges posed by the cloud environment, thus constraining their effectiveness in safeguarding against cloud-based attacks. Traditional methods for analyzing virtual machine (VM) memory and conducting forensic investigations are becoming increasingly very complicated and resource-intensive, rendering them impractical for everyday use. There is a pressing need for an automated solution capable of handling the concurrent security of numerous

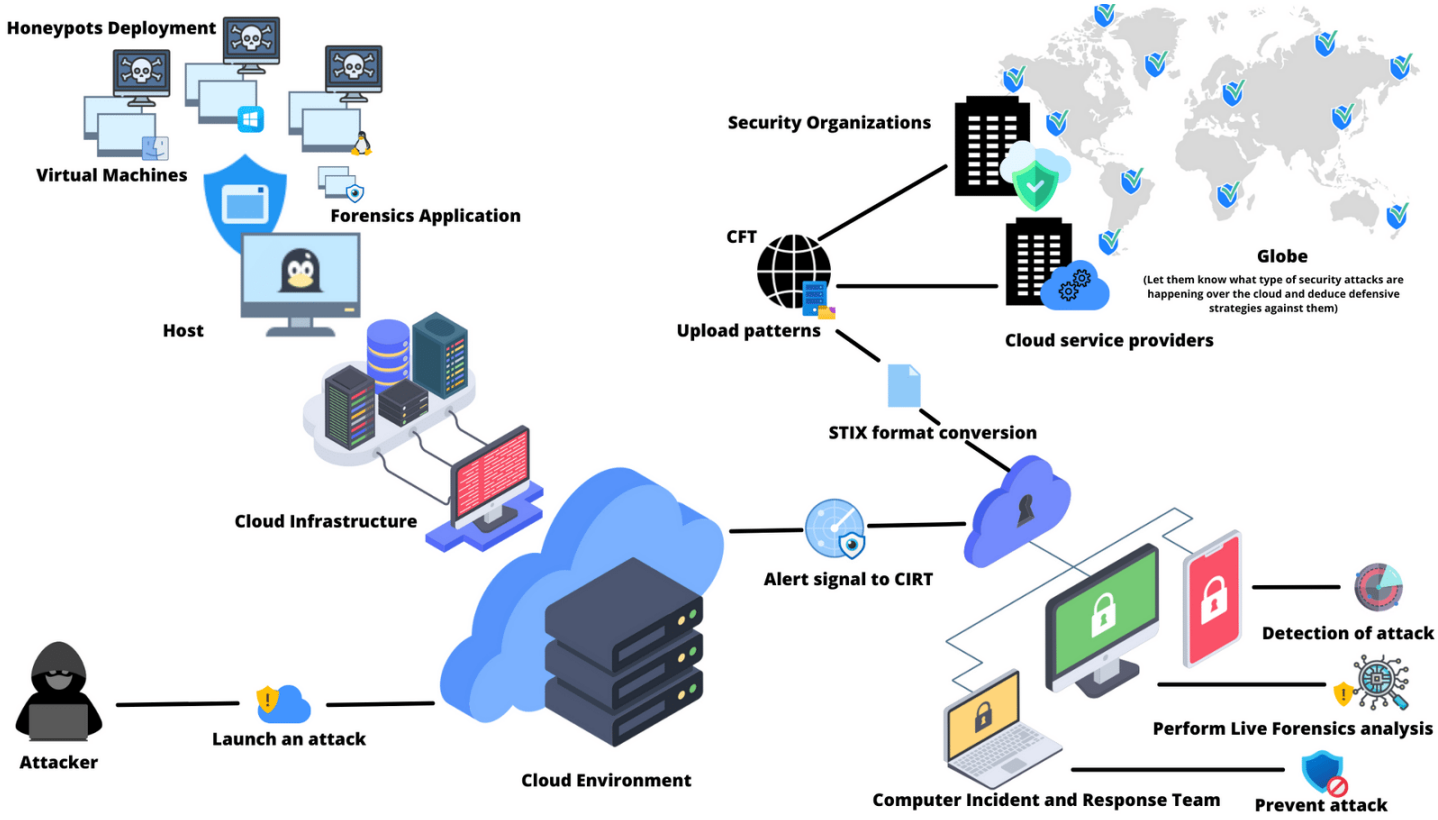


Figure 3: Provides a comprehensive overview of the proposed framework.

VMs, effectively guarding against a broad spectrum of emerging cyber threats, and possessing the scalability required to meet the demands of extensive cloud environments.

3 Proposed Framework: Design and Architecture

Our proposed framework for detecting and analyzing attacks on cloud platforms adopts a proactive monitoring approach encompassing both the host and virtual machine (VM) levels. Within this framework, a dedicated forensic application is deployed to inspect VM memory for valuable artifacts and to execute forensic analysis in response to potential attacks on VMs. In the event of a security breach, the system generates alerts, which are promptly disseminated to a Computer Incident Response Team (CIRT) portal. At the CIRT portal, an in-depth forensic analysis is initiated to identify the specific tactics, techniques, and procedures employed in the attack. The findings from this analysis are subsequently translated into a structured threat information expression (STIX) format, enabling their integration into the realm of cyber threat intelligence. Importantly, this framework emphasizes rapid incident reporting, ensuring that security incidents are communicated as soon as they are detected and thoroughly analyzed.

The CIRT possesses access to an extensive knowledge repository comprising analytical reports, procedural guidelines, service records, security logs, and applications installed within the targeted virtual machine (VM). This repository serves as a pivotal resource for identifying any indicators of malicious activity during forensic analysis. Furthermore, the forensic application integrated into the framework exhibits continuous operational verification by dispatching SYN (synchronize) messages to the user interface, thereby confirming its operational status and ensuring the continued activity of the VM. This application also generates diverse types of alerts that are duly documented within the activity log featured on the dashboard. The underlying objective of our proposed framework is to deliver a holistic approach for the detection and analysis of security breaches within cloud platforms. This entails facilitating the exchange of threat intelligence, enriching the knowledge base, and generating alerts, all while maintaining cross-platform compatibility—a facet often absent in traditional solutions. The fundamental aim is to mitigate the repercussions of security incidents and empower the formulation of effective defensive strategies [23].

Fig. 3 offers a comprehensive schematic representation of our proposed framework, providing a holistic view of its operational dynamics. The diagram illustrates the combination of (VMs) operating within the host machine, located in the upper left corner. Concurrently, it offers an illustrative depiction of an attack scenario orchestrated by an aggressor. On the right-hand side of the diagram, the visualized attack scenario seamlessly transitions into the stages of detection, forensic analysis, and the subsequent generation of attack patterns. Notably, these attack patterns are made accessible to the public domain, facilitating proactive measures for defense against forthcoming cyber threats. This visual representation encapsulates the essence of our proposed framework's operation clearly and concisely, offering a cohesive understanding of its functionali-

ties.

In the realm of cybersecurity, the acquisition and utilization of information hold paramount importance. The contemporary surge in cyberattacks has given rise to an encyclopedia of data related to these attacks, effectively constituting a comprehensive repository of information-gathering assaults. Safeguarding systems and IT infrastructure necessitates a two-fold approach. Firstly, a thorough understanding of system vulnerabilities and weaknesses is imperative. Simultaneously, it is crucial to possess insights into the tactics and techniques employed by malicious actors to conduct attacks. This dual comprehension forms the base for making informed decisions geared towards fortifying system defenses. Acknowledging the critical need for robust information exchange within the cybersecurity community, a standardized language has been devised to share threat indicators. This standardization provides flexible and effective dissemination of vital information among cybersecurity stakeholders. In tandem, it complements the proactive stance of gathering comprehensive knowledge about cybersecurity threats, thereby fortifying our capacity to defend against adversarial actions targeted at IT and digital infrastructure. Facilitating this information sharing and knowledge enrichment is done by the Structured Threat Information eXpression (STIX). STIX is an industry-standard programming initiative, developed from the collaborative efforts of multiple organizations, tailored to cater to the need of cyber threat intelligence. This standardized framework is designed to enable the most effective, flexible, and uniform exchange of information in the realm of cybersecurity. The data sharing encompassed within this framework covers threat indicators, incident reports, cyberattack campaigns, profiles of threat actors, courses of action, cyber observables, adversary Tactics, Techniques, and Procedures (TTPs), and exploited targets.

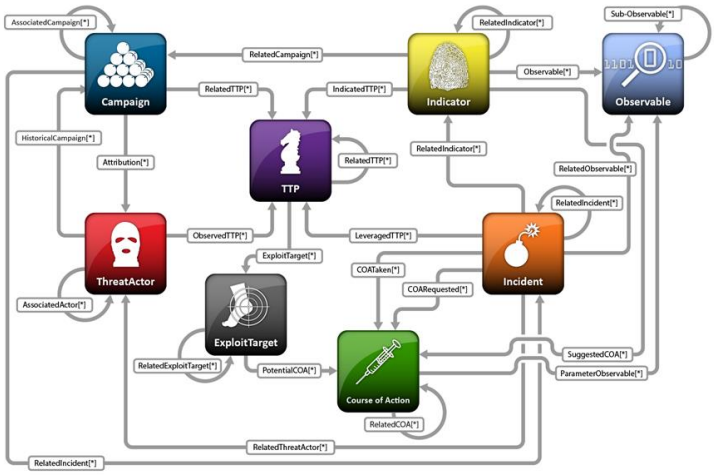


Figure 4: Figure 4 outlines STIX's architecture [24] for human-machine cyber threat understanding.

Fig. 4 provides a brief and basic overview of STIX's architecture consisting of cyber threat information which enables human and machine understandable format. Furthermore, our proactive approach extends to the development of a web portal that invites collaborative contributions from Computer Incident Response Teams (CIRTs), cloud service providers, and security organizations. This collaborative endeavor is predicated on the principle that sharing knowledge and expertise is instrumental in enhancing cybersecurity measures. In the aftermath of successfully mitigating an attack, the resulting attack patterns are transmuted into the STIX format and uploaded onto this web portal. This affirms the pivotal role played by STIX in the domain of cyber threat intelligence, as an indispensable tool for strengthening cybersecurity measures and fostering collective resilience against evolving threats.

Within this solution, we have strategically integrated the NIST 800-53 security control families, with particular emphasis on the following key elements: **Audit and Accountability (3.3):** This component is vital for maintaining comprehensive records of system activity to facilitate effective monitoring and analysis. **Assessment, Authorization, and Monitoring (3.4):** This aspect encompasses the critical phases of assessing, authorizing, and continually monitoring system security. **Vulnerability Monitoring and Scanning (3.16 RA-5):** The solution places significant focus on continuously monitoring and scanning for vulnerabilities, ensuring proactive threat mitigation. **Technical Surveillance Countermeasures Survey (3.16 RA-6):** This control addresses the imperative task of conducting technical surveillance countermeasures surveys to safeguard against potential threats. **Incident Response (3.8):** Incident response is a core facet, with particular emphasis on IR-4, which encompasses incident handling. This includes detection, analysis, containment, and eradication of security incidents. These NIST 800-53 control families are integral to the comprehensive complete solution, contributing to its robust security posture by addressing various aspects of security control, monitoring, and incident response, thereby fortifying the overall cybersecurity framework.

3.1 Deployment and Simulation

The proposed framework for the detection and analysis of attacks on cloud platforms is adaptable to diverse cloud environments. Its flexibility is exemplified by the ability to deploy the forensic application both on the host and within virtual machines (VMs). Furthermore, the CIRT portal is designed to be accessible remotely, permitting authorized users to engage with it from a distance. To enhance the efficacy of the proposed framework, rigorous testing can be conducted through the execution of simulated attack scenarios encompassing various attack types and security contexts. The outcomes derived from these simulations serve as a means to comprehensively assess the framework's performance. Subsequently, any identified areas for potential enhancement or improvement can be systematically addressed, ensuring that the framework attains optimal effectiveness in real-world deployments. This iterative process of testing and refinement contributes to the continual evolution and robustness of the framework, aligning it with the dynamic landscape of cybersecurity in cloud environments.

To illustrate, the forensic application within the framework is amenable to configuration for the emulation of diverse attack types, encompassing scenarios involving malware, rootkits, and keyloggers. Subsequently, the CIRT can conduct an in-depth analysis of the data generated by these simulated attacks to assess the framework's efficacy in both detection and analysis. The outcomes of these simulations serve a dual purpose. Firstly, they enable the fine-tuning of the forensic application's configuration and the optimization of

the CIRT portal's functionality, enhancing their performance in the context of attack detection and analysis. Secondly, the simulation results can illuminate potential vulnerabilities or deficiencies within the framework, thereby providing critical insights for further enhancement. By utilizing the power of simulations, the proposed framework stands to undergo iterative refinement, resulting in an elevated level of effectiveness in the detection and analysis of attacks within cloud platforms. This iterative process, rooted in empirical testing and analysis, lends itself to the continuous improvement of the framework's capabilities, ensuring its relevance and robustness in an ever-evolving cybersecurity landscape within the cloud. To facilitate comprehension, an illustrative workflow is presented in Fig. 5, providing a simplified depiction of the sequential steps inherent to the outlined rational approach. This provides a clear and accessible representation of the procedural sequence.

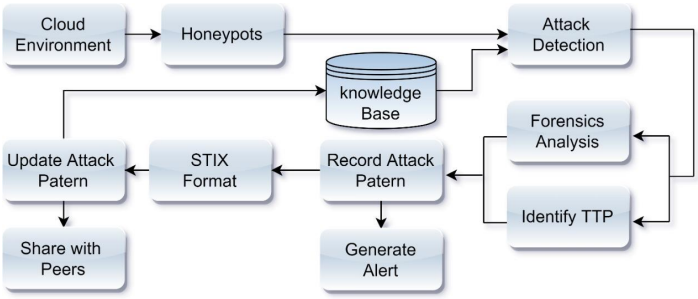


Figure 5: A visual representation of the attack detection workflow process.

The establishment of a cloud system entails the configuration of multiple virtual machines (VMs), each equipped with distinct honeypots designed to entice potential attackers into initiating attacks. Upon activation of these honeypots, they initiate a signaling process that notifies the Computer Incident Response Team via the dedicated CIRT portal. This orchestration allows the framework to conduct forensic analysis of the attack stealthily, rendering the attacker unaware of the ongoing detection, monitoring, and analysis. This stealthy approach is instrumental in uncovering the Tactics, Techniques, and Procedures (TTP) deployed by the attacker during the course of the attack. By surreptitiously observing and dissecting the attack, the framework can glean critical insights into the attacker's modus operandi, thereby enhancing our understanding of cyber threats and fortifying our defenses. Detailed information regarding the subcomponents, essential features, and functional domains is elaborated upon in the subsequent sections.

3.1.1 *Forensic application*

The core functionality of the forensic application encompasses the critical task of executing virtual machine introspection (VMI) and subsequently transmitting the acquired data to the web service [25]. It is important to highlight that the forensic application is designed to function flawlessly on both the host and the virtual machines (VMs), highlighting its critical function in enhancing cloud infrastructure security against persistent cyber threats. To ensure the distinct identity and traceability of each instance of the forensic application, a unique identification (ID) is assigned to every running instance, whether on the host or VMs. This rigorous identification system serves as a base for precise command execution during VMI operations. Notably, the framework's inherent cross-platform compatibility enables the seamless implementation of updates and patches for the forensic application, thereby ensuring its continual alignment with evolving security requirements. It is crucial to keep the forensic application hidden from possible attackers. It is safeguarded through an array of comprehensive countermeasures that are meticulously enforced to preserve the application's integrity. Furthermore, the forensic application, seamlessly integrated within the framework, maintains continuous operational verification. This is achieved through the systematic dispatch of SYN messages to the user interface, thereby affirming its operational status and ensuring the sustained functionality of the associated VM.

3.1.2 *Web services*

The web service plays a pivotal role within the architecture, serving as a vital intermediary between the forensic application and a centralized database repository. This repository houses an extensive range of data, which is subsequently made accessible through a dedicated cloud forensic tool (CFT) or portal. The primary function of the web service revolves around the seamless transmission of data from the forensic application to the database, a process executed in real-time upon data reception. The data collected by the forensic application is rich and multifaceted, encompassing crucial system information such as operating system details, processor specifications, core configurations, RAM allocations, storage characteristics (both hard drive and solid-state drive attributes), serial identification markers, geographical location data, active processes, system services, security logs, and a comprehensive list of installed applications. Before integration into the database, this collected data undergoes rigorous parsing procedures to ensure its uniformity and compatibility. Once parsed, the data is methodically inserted into the database, subsequently becoming readily accessible through the designated portal. The incorporation of the web service stands as a fundamental component within this framework, endowing the proposed solution with cross-platform compatibility. This strategic architecture ensures that the core infrastructure remains consistent, with adaptations made solely to the forensic application running on the host and VMs to align with the specific programming languages and hardware of the underlying systems. Consequently, the framework boasts versatility, capable of accommodating diverse operating environments ranging from UNIX, Linux, MAC, and IOS, to Android, among others. This cross-platform compatibility underscores the framework's adaptability and applicability across a wide spectrum of technology ecosystems.

3.1.3 *Cloud Forensic Tool/Portal (CFT)*

Cloud Forensic Tool (CFT) is a special portal designed to make it easier for Cloud Service Providers (CSPs) to interact with the framework. Through this site, CSPs can easily register both their host machines and virtual machines (VMs). Furthermore, the CFT operates as a centralized hub for receiving real-time security breach alerts, a critical function that underpins the framework's proactive stance in cybersecurity. Integral to the CFT's functionality is its role as the platform for orchestrating Virtual Machine Introspec-

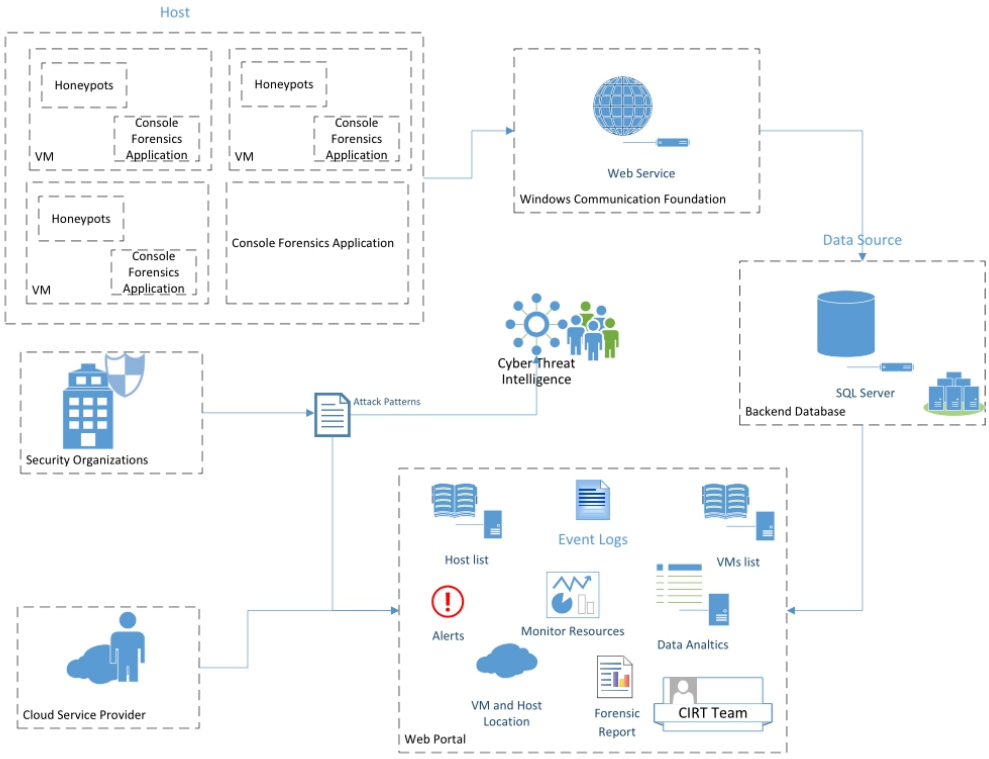


Figure 6: Provides a complete system architecture overview of the proposed framework.

tion (VMI). This capability empowers the framework to conduct deep and real-time examinations of VMs, enhancing its capacity to swiftly detect and analyze security breaches. Additionally, the CFT is the medium through which attack patterns are uploaded in a structured threat information expression (STIX) format after an attack has been effectively contained. This mechanism facilitates the dissemination of vital cyber threat intelligence, thereby bolstering the collective knowledge base and enhancing preparedness against future threats. Fig 7 offers a glimpse into the CFT's user interface, providing an insightful single-screen overview encompassing all integral components of the framework. This visual representation serves as an invaluable tool, affording users a consolidated perspective of the framework's operations, enhancing user experience, and streamlining administrative functions.

Figure 7: A dashboard screen snapshot offering a comprehensive overview of CFT components.

The CFT dashboard screen serves as a pivotal control center within the framework, offering a comprehensive array of critical information and functionalities. It serves as a dynamic repository of indispensable data, including reported cybersecurity incidents, real-time threat alerts, updates regarding newly onboarded Cloud Service Provider (CSP) members, resource utilization metrics pertaining to host machines, and geospatial location data pinpointing the whereabouts of both host machines and their hosted virtual machines (VMs). Notably, CSPs are endowed with the versatile capability to access and extract data in various formats, including the option to download data in CSV files or generate hardcopy printouts. This flexibility extends to encompass data retrieval from host machines, VMs, as well as comprehensive logs and process records.

In terms of incident management and response, both CSPs and administrators wield the authority to initiate alert transmissions to the Computer Incident Response Team (CIRT). This role complements the foundational commitment of the framework to prompt and efficient incident response and plays a crucial role in launching forensic investigations. Furthermore, the dashboard's functionality extends to the systematic dispatch of SYN (synchronize) messages from VMs to validate the operational status of the forensic application and affirm the active state of the VM. Together with the various alert types found in the activity log, these messages help to strengthen the dashboard's real-time monitoring and notification features.

Optimizing the framework is a crucial aspect of maintaining a robust security posture in cloud environments. The optimization process involves identifying vulnerabilities or weaknesses in the framework's configuration. This can be achieved through rigorous testing, incident simulations, and real-world incident analysis. The primary goals of optimizing the framework are to enhance its resilience against emerging cyber threats, improve its efficiency in detecting and analyzing attacks, and minimize false positives. During the optimization process, cloud service providers (CSPs) and administrators collaborate to fine-tune various components of the framework. This may include adjusting settings for Virtual Machine Introspection (VMI), refining incident classification criteria, and updating security controls. By continually optimizing the framework, organizations can adapt to evolving threat landscapes and ensure that their cloud infrastructure remains secure and resilient. Enhancing the forensic application is a proactive measure aimed at improving the effectiveness of virtual machine introspection (VMI) and forensic analysis capabilities. Enhancements to the forensic application may encompass several aspects for example performance improvement, feature updates, security enhancement, and cross-platform compatibility. Regular updates and enhancements to the framework help to stay ahead of emerging cyber attacks.

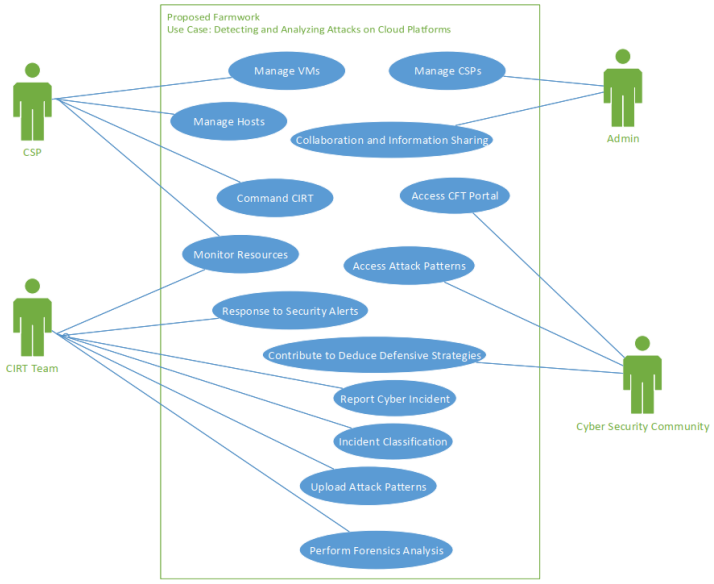


Figure 8: Provides a concise overview of use case modeling.

Fig 8 provides a very brief and precise use case modeling encompassing the proposed framework, and Fig 6 depicts the system architecture implemented within the proposed framework.

4 Results

The outcomes derived from the utilization of our proposed Hypervisor-based Virtual Machine Introspection (HVMI) tool exhibited promising results in detecting and analyzing attacks on cloud platforms. Rigorous testing was conducted through a series of simulations to comprehensively evaluate HVMI's efficacy in detecting a spectrum of attack types and dissecting the complexity of their tactics, techniques, and procedures (TTPs).

In the initial simulation, our evaluation focused on assessing HVMI's capability to identify a straightforward brute-force attack directed toward a virtual machine. The findings from this simulation revealed that HVMI exhibited prompt and effective detection of the attack, transmitting timely notifications to the CSP mere seconds after the attack's commencement. Furthermore, HVMI furnished a detailed and comprehensive analysis of the attack's Tactics, Techniques, and Procedures (TTPs). This encompassed an in-depth examination of the attack's nature, its originating source, and the specific system that was the subject of the attack.

Within the scope of the second simulation, our assessment aimed to evaluate HVMI's proficiency in detecting and analyzing a complicated and highly sophisticated malware attack targeted at a virtual machine. The findings from this simulation affirmed HVMI's adeptness in not only promptly identifying the attack but also in conducting a comprehensive analysis of the attack's intricacies. This analysis culminated in the generation of an exhaustive report detailing the Tactics, Techniques, and Procedures (TTPs) employed by the attack. Furthermore, the report delved into the potential ramifications of the attack, particularly in terms of its potential impact on the confidentiality, integrity, and availability of the system. To facilitate a holistic understanding of the entire process, we have included a detailed case study within this context.

4.1 Case Study: Launching a reverse TCP attack

Conducting the case study necessitated the initial setup of a Kali Linux machine, designated as the attacker entity. This machine underwent meticulous configuration to enable the initiation of a reverse TCP attack against a virtual machine (VM) hosted within the cloud infrastructure, ultimately aiming to illicitly establish reverse shell access. For the execution of this attack, the utility was employed to craft a payload tailored to facilitate the invasion into the VM. A pivotal component of this payload was an executable (Exe) file, meticulously constructed to be delivered to the target VM. Upon execution of this payload, a chain of events was set into motion. It prompted the VM to initiate the opening of a reverse shell, thus enabling the attacker to gain remote access through reverse TCP connectivity. Within the payload's configuration, crucial details such as the IP address and port on which the attacker's machine would be listening were meticulously specified. This case study served as a practical exercise to comprehensively assess HVMI's capabilities in detecting and analyzing an incursion of this nature, ultimately yielding valuable insights into the attack's methodologies and potential impact on the targeted system's security and integrity.

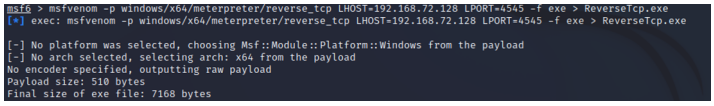


Figure 9: A Snapshot of an Exe payload generation for targeted VM.

Upon configuring the requisite parameters, the exploit was executed. This initiation prompted the attacker's machine to transition into a listening state, awaiting the incoming reverse TCP connection

from the target virtual machine (VM). This allowed an attacker to gain unauthorized access to the victim machine's shell, thereby affording an array of capabilities for potentially malicious activities. Within the scope of this gained access, activities encompassed the potential for privilege escalations and the execution of DLL injection, thereby underscoring the severity and breadth of the security breach. Fig 10 shows a snapshot of the reverse shell from the attack simulation.

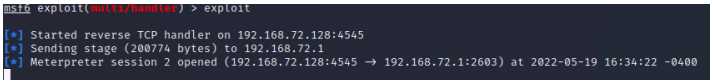


Figure 10: A snapshot of the accessed reverse TCP shell in the attack simulation.

Following the successful execution of the attack, our subsequent actions encompassed the installation of remote access software with the intent of securing persistent access while simultaneously deleting all traces within the event logs to obfuscate our activities and minimize detection. However, as soon as this attack occurred, our forensic application sprang into action. It promptly transmitted a series of alerts to the designated web portal and initiated the critical forensic analysis procedure. Fig 11 provides a comprehensive insight into the detailed real-time alerts generated as part of this incident response mechanism, emphasizing the real-time nature of our detection and analysis capabilities.

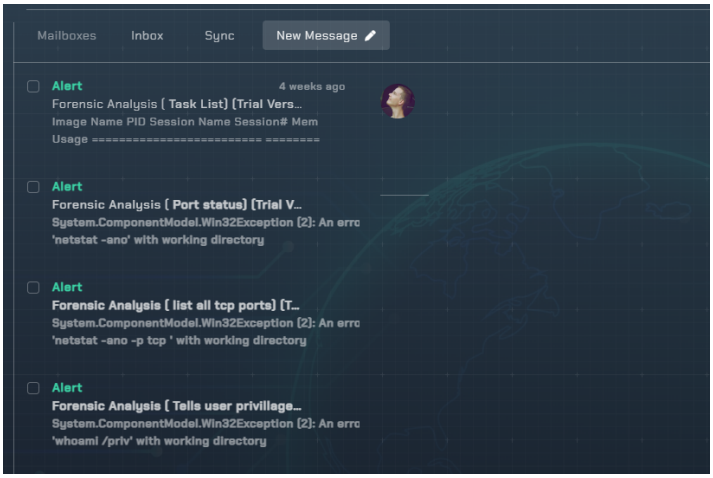


Figure 11: A snapshot of real-time alert prompts displayed on the portal.

This point underscores the pivotal role played by HVMI in resolving these security challenges. The proactive security and monitoring functionalities embodied within HVMI serve as a potent deterrent against such malicious incursions.

Before embarking on the identification of anomalous entities within VM's memory, several key inquiries come to the forefront. These include: Process Legitimacy: Are there any indications of suspicious processes deviating from their expected execution paths? Is the suspicious process running in tandem with its legitimate parent, or does it originate from an unrelated source? Temporal Aspects: What is the temporal profile of the process in question? When did it commence and conclude its execution? What are the implications of its execution timeline, and do they align with the expected behavior? Behavioral Profiling: What behavioral advantages can be attributed to the observed process? Does its behavior conform to the anticipated outcomes, or does it exhibit aberrant characteristics that warrant scrutiny? Process Name Variance: Does the process name exhibit any anomalies? Is it masquerading as a valid Windows process, thus evading detection? Comprehensive Analysis: Beyond process name scrutiny, an exhaustive analysis extends to threads, mutexes, Dynamic Link Libraries (DLLs), process-to-file mappings, Resident Memory (RAM) sections, and any associated sockets and open ports attributed to the process. These attributes are critical in discerning the origin of connections and the entities initiating them, as well as their associated dependencies. Our proposed solution (HVMI) empowers cybersecurity professionals with a potent arsenal of capabilities, enabling a proactive stance in countering, identifying, and comprehensively analyzing security threats.

HandleCount	Name	Priority	ProcessId	ThreadCount	WorkingSe
0	System Idle Process	0	0	8	8192
4139	System	8	4	115	57344
0	Registry	8	124	4	79228928
60	smss.exe	11	352	2	573440
645	csrss.exe	13	452	11	2678784
161	wininit.exe	13	540	3	1728512
169	csrss.exe	13	548	10	2392064
214	winlogon.exe	13	640	3	4018176
644	services.exe	9	684	9	8208384
4304	lsass.exe	9	704	8	19484672
1033	svchost.exe	8	824	11	15077376
33	fontdrvhost.exe	8	852	5	1613824
33	fontdrvhost.exe	8	860	5	1523712
1232	svchost.exe	8	936	10	20365312
1911	svchost.exe	8	988	5	9269248
2881	svchost.exe	8	500	127	33820672
503	LogonUI.exe	13	440	9	45416448
752	dwm.exe	13	552	14	21630976
211	svchost.exe	8	984	1	2277376
340	svchost.exe	8	1080	1	4386816
611	svchost.exe	8	1152	6	30240768
158	svchost.exe	8	1272	3	3686400
138	svchost.exe	8	1300	2	3866624
213	svchost.exe	8	1364	2	7499776
179	svchost.exe	8	1372	3	2637824

Figure 12: A snapshot of real-time processes running on VM.

The forensic application, as part of its operational protocol, diligently transmitted an array of pertinent information regarding the processes actively running within the targeted virtual machine (VM). This comprehensive data set encompassed critical attributes such as process priority, unique process identification (PID), thread counts, and resource consumption metrics, among other relevant parameters. For more clarity and reference, Fig 12 provides a visual representation of the enumerated processes actively executing within the targeted VM, thereby affording a comprehensive overview of the system's operational state.

sethc.exe	12800		2	720 K
TabTip.exe	13360		2	9,588 K
TabTip32.exe	14272		2	1,376 K
svchost.exe	12740	Services	0	5,996 K
SecurityHealthService.exe	12572	Services	0	9,932 K
msedge.exe	2816		2	86,100 K
msedge.exe	6444		2	9,116 K
msedge.exe	8892		2	31,396 K
msedge.exe	5396		2	35,784 K
msedge.exe	6552		2	19,728 K
msedge.exe	12692		2	72,888 K
TabTip.exe	11060	Console	1	13,500 K
msedge.exe	9140		2	134,240 K
MusNotifyIcon.exe	11844		2	3,420 K
w3wp.exe	8252	Services	0	322,612 K
Ssms.exe	6084		2	306,216 K
WmiPrvSE.exe	11760	Services	0	20,896 K
sppsvc.exe	11236	Services	0	12,688 K
devenv.exe	11052		2	1,257,808 K
ReverseTcp.exe	11056	Services	2	9,924 K
PerfWatson2.exe	8160		2	77,848 K
Microsoft.ServiceHub.Cont	10704		2	73,008 K
ServiceHub.VSDetouredHost	14648		2	101,360 K
ServiceHub.IdentityHost.e	14548		2	71,900 K
ServiceHub.ThreadedWaitDi	15692		2	80,312 K
ServiceHub.RoslynCodeAnal	700		2	434,764 K
ServiceHub.Host.CLR.x86.e	9676		2	73,828 K
ServiceHub.SettingsHost.e	14584		2	76,404 K
ServiceHub.Host.CLR.x64.e	12860		2	827,164 K
ServiceHub.Host.CLR.exe	6148		2	117,212 K
ServiceHub.TestWindowStor	14172		2	80,300 K
Microsoft.Alm.Shared.Remo	15284		2	70,808 K
Microsoft.VisualStudio.We	12324		2	90,512 K
ServiceHub.DataWarehouseH	5580		2	105,408 K
WebViewHost.exe	13668		2	55,956 K
conhost.exe	2940		2	11,576 K
msedgewebview2.exe	10308		2	103,552 K
msedgewebview2.exe	12492		2	8,880 K
msedgewebview2.exe	7936		2	45,892 K
msedgewebview2.exe	9396		2	29,668 K
msedgewebview2.exe	4420		2	19,160 K
msedgewebview2.exe	4344		2	68,864 K
msedgewebview2.exe	1088		2	52,956 K
ServiceHub.Host.CLR.x86.e	8560		2	86,092 K
node.exe	10212		2	43,752 K
node.exe	4576		2	44,092 K
conhost.exe	5488		2	11,744 K
conhost.exe	6980		2	11,752 K
node.exe	5384		2	38,636 K
VsDebugConsole.exe	8460		2	6,200 K
conhost.exe	11936		2	22,984 K

Figure 13: A snapshot of real-time identification of ”ReverseTcp.exe” by the forensic application on the VM.

Attackers employ diverse methodologies for delivering malicious payloads to victim machines. Given our primary emphasis on defensive measures, we have confined our analysis to two prominent attack vectors. The first scenario involves an attacker uploading a payload to a server hosting a service accessible via a specific port. Alternatively, attackers may disseminate the payload by incorporating it into a website or server. In such instances, unsuspecting users who attempt to access the compromised server or website inadvertently trigger the download of the payload. Additionally, attackers may resort to deception by disguising the payload as a Trojan horse. In response to any of these scenarios, the Computer Incident Response Team (CIRT) wields the capability to issue commands to the forensic application, thereby facilitating the retrieval of a comprehensive list of installed applications within the target virtual machine (VM). This scrutiny serves as a crucial component in the quest to identify any suspicious software installations or potential security breaches. Fig 13 provides an inclusive catalog encompassing active services and applications hosted within the VM. This exhaustive compilation of applications is readily accessible through the Computer Incident Response Team (CIRT) portal, thereby offering a streamlined reference point for cybersecurity professionals and facilitating in-depth investigative analysis. Of noteworthy significance within this list is the presence of the ”Reverse TCP Exe” application. This particular application bears significance as it was generated by the attacker with the explicit intent of establishing remote access to the virtual machine, thereby highlighting a critical security concern warranting further investigation and remediation.

Fig 14 provides a brief overview of the activity log within the CIRT portal. This log serves a dual purpose: first, it offers operational verification of both the Virtual Machine (VM) and the forensic applications throughout the simulation, ensuring their continuous activity and functionality. Secondly, it highlights the availability of the forensic analysis report on the portal, presenting a real-time and chronological account of activities and alerts generated during the simulation.

Figure 14: A snapshot of the activity log within the CIRT portal.

The HVMI tool demonstrated its capability to access and retrieve a comprehensive analysis report pertaining to the targeted virtual machine (VM). This encompassed a thorough examination of the VM's operational facets, encompassing active processes, running services, security logs, and the catalog of installed applications. Furthermore, the Cloud Service Provider (CSP) was equipped with the remote accessibility to monitor processes and network connections, allowing for the identification of the processes responsible for initiating and sustaining these connections. At a broader analytical spectrum, the HVMI tool extends two primary reporting avenues, specifically the Forensic Report and the Incident Report. The Forensic Report serves as a document that presents the results of forensic investigation. Fig 15 provides a snapshot of the interface, showcasing the automated forensic reports that have been systematically generated and made accessible through the portal.

Figure 15: A snapshot of available forensic reports on the portal.

Fig 16 provides a visual representation of the logs that have been systematically generated and recorded throughout the entire operational process. These logs are pivotal in facilitating comprehensive documentation of the actions, events, and activities undertaken within the cloud platform, rendering them invaluable for analytical and investigative purposes within the academic and technical domains.

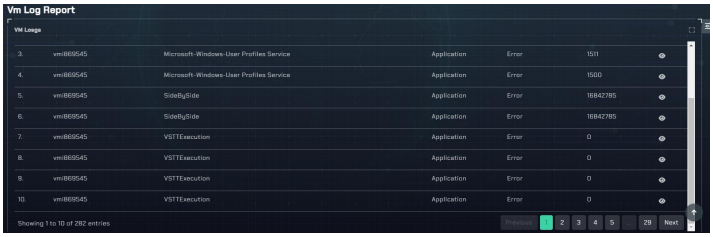


Figure 16: A visualization of VM logs generated during the entire process.

The Computer Incident Response Team (CIRT) portal offers the capability to export event logs in digital format, specifically in Comma-Separated Values (CSV) format for electronic use, or in hardcopy format for physical documentation purposes. This functionality allows for the efficient dissemination and preservation of critical event data for further analysis and reference in the context of incident response and cybersecurity management.

Within the context of cloud service providers (CSPs), the establishment of a robust cyber threat intelligence capability is of paramount importance. A pivotal facet of such a capability lies in the practice of sharing critical information with trusted partners, peers, and relevant stakeholders. This collaborative approach to cyber threat intelligence serves as a vital means of streamlining and comprehending the vast and very complicated landscape of cybersecurity data that CSPs encounter. By engaging in cyber threat intelligence and data sharing endeavors, organizations can effectively distill and prioritize the overwhelming volume of complicated cybersecurity data they encounter. This process involves gaining insight into the vulnerabilities and flaws within their information systems, as well as an understanding of the complicated tactics, techniques, and procedures (TTPs) employed by cyber adversaries. Additionally, it encompasses the identification and analysis of attack patterns. For the purpose of global accessibility and ease of comprehension, the results of cyber incidents, intelligence data, and attack patterns are systematically generated in the standardized Structured Threat Information Expression (STIX) format. This format facilitates the seamless exchange of cyber threat intelligence and enhances the collaborative efforts aimed at fortifying information systems against cyberattacks. Fig 17 provides a sample code snippet that serves as an illustrative for comprehending the visualization of Structured Threat Information Expression (STIX).

```
<stix:Threat_Actors>
    <stix:Threat_Actor id="trojan:threatactor-9b371afe-ddfd-4954-abaf-8abb357ac78e" xsi:type="ta:ThreatActorType" version="1.2">
        <threat-actor:Title>Trojan Horse backdoor</threat-actor:Title>
        <threat-actor:Type>
            <stixCommon:Value>APT</stixCommon:Value>
        </threat-actor:Type>
    </stix:Threat_Actor>
    <threat-actor:Observed_TTPs>
        <threat-actor:Observed_TTP>
            <stixCommon:Relationship>Use</stixCommon:Relationship>
            <stixCommon:TTP idref="trojan:ttp-649b70a0-0158-4cc6-a8d5-cf8a441dc290"/>
        </threat-actor:Observed_TTP>
    </threat-actor:Observed_TTPs>
</stix:Threat_Actors>
```

Figure 17: A sample code for visualization of patterns and syntax within STIX code.

STIX serves as a standardized language designed to convey comprehensive information concerning cyber threats. This encompassing data includes indicators of compromise (IOCs), which are distinctive attributes or artifacts linked to a cyberattack, such as specific IP addresses, domain names, file hashes, and other pertinent information instrumental in identifying or tracing an attack's origins. The strategic collection and analysis of IOCs empower organizations to pinpoint ongoing or past attacks, subsequently enabling them to initiate pertinent countermeasures and proactively deter future assaults. STIX plays a pivotal role in streamlining the aggregation, organization, and dissemination of IOCs, ensuring consistency and structure in the process. Fig 18 provides a visual representation of the relationship between the extracted IOCs during the attack simulation.

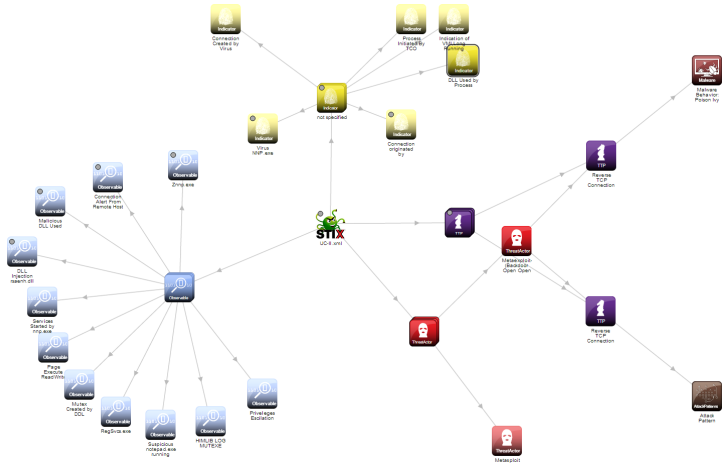


Figure 18: Provides a visual representation of the relation between the extracted IOCs.

In the cloud environment organizations frequently operate within shared infrastructure and resource environments, which inherently pose challenges in terms of tracking and detecting cyberattacks. The integration of STIX and IOCs into their cybersecurity framework equips organizations with a standardized mechanism to enhance their ability to recognize and respond to attacks more efficiently. This is achieved by facilitating the systematic collection and sharing of critical threat-related information and indicators of compromise, ultimately enhancing their cyber defense capabilities.

As an illustration, consider an organization that strategically incorporates STIX and IOCs into its cloud security framework. In this context, the organization leverages various sources of threat intelligence feeds or other pertinent information repositories to proactively identify potential indicators of compromise (IOCs). These IOCs might encompass entities like IP addresses or domains that are linked to previously documented malicious activities. Subsequently, the organization uses this invaluable information to configure its suite of security tools and policies. These configurations serve to either block or trigger alerts in response to these identified IOCs, thereby fortifying their capacity to avert or detect cyberattacks at an early stage, consequently reducing the potential for substantial damage.

Upon the completion of the comprehensive forensic analysis, the prevention of the cyber attack, and the documentation of attack patterns, the subsequent crucial step involves the formal reporting of the incident. An incident report serves as comprehensive documentation encompassing everything about the cyber attack incident that has occurred. These details encompass temporal information such as the precise timing of the incident, location, and a thorough narrative outlining the nature of the occurrence. Additionally, the report encapsulates actions taken in direct response to the incident.

Fig 19 provides a snapshot of the cyber incident report available on the portal.

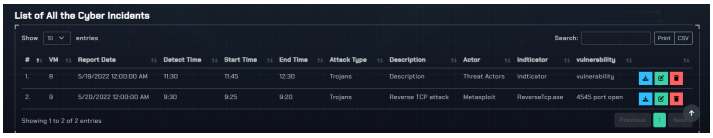


Figure 19: A snapshot of a comprehensive cyber incident report accessible via the portal.

The cyber incident report serves as a repository of critical insights into the security breach, exploring various facets and aspects of the incident from different angles. This comprehensive report encompasses pivotal details such as the attack vectors employed, the specific vulnerability that was exploited, the identity of the threat actor or actors involved, the classification of the attack type, and a detailed narrative about the attack's modus operandi. Moreover, the report incorporates a comprehensive analysis of attack pattern vectors and indicators, facilitating in understanding of the incident's intricacies. Additionally, it features a graphical STIX format, providing a visual overview of the incident's characteristics. Notably, the report also offers insights into recommended patches and fixes for the vulnerabilities that were exploited to compromise the system, thus contributing to the post-incident remediation process.

In summary, the outcomes of our simulation exercises demonstrate the efficacy of HVMI in detecting and analyzing cyberattacks within cloud infrastructures. HVMI emerges as a potent instrument, providing cloud service providers and cybersecurity practitioners with a robust means to safeguard cloud environments and mitigate the repercussions of security breaches. These results underscore the significance of HVMI as a proactive and potent tool in the arsenal against cloud-based cyber threats, offering valuable insights and capabilities to enhance the security posture of cloud systems.

5 Conclusion and Future Work

In conclusion, the ever-expanding realm of cloud computing necessitates a robust and adaptive security framework to combat the growing sophistication of cyber threats. The proposed HVMI solution, comprised of core components like the VMI-enabled forensic application, CFT portal, and data transmission web service, offers a comprehensive and promising strategy for detecting and analyzing attacks in real-time. Its commitment to iterative optimization and enhancement, coupled with proactive measures such as fortifying the forensic application, ensures that organizations can effectively safeguard their cloud infrastructure. By continually refining their security posture through HVMI, organizations can stay resilient against emerging threats and minimize the impact on the confidentiality, integrity, and availability of their cloud systems in this dynamic cybersecurity landscape. Nonetheless, here are several areas for future research that hold potential for enhancing HVMI's efficacy.

To ensure the effectiveness of HVMI in safeguarding large-scale cloud systems, optimizing its scalability is paramount. This endeavor entails devising techniques capable of efficiently processing substantial data volumes while fine-tuning HVMI's performance under heavy loads. The scalability enhancements will enable HVMI to protect expansive cloud infrastructures effectively, even as they evolve and expand. Future research can delve into advanced techniques for detecting and analyzing cyberattacks within cloud environments. Notably, machine learning-based approaches can be explored to elevate HVMI's capacity to identify and respond to increasingly sophisticated threats. By integrating machine learning models for anomaly detection, the framework can discern novel attack patterns and adapt dynamically to emerging threats.

Moreover, future work could explore intriguing intersections with emerging technologies such as blockchain and virtual reality. Integrating blockchain technology into HVMI has the potential to enhance the security and traceability of its forensic analyses. This innovation would instill greater confidence in the integrity of HVMI's findings, as they would be anchored in an immutable blockchain ledger. Meanwhile, integrating virtual reality could revolutionize forensic analysis, creating immersive and interactive experiences that empower investigators to explore attack scenarios in unprecedented depth. Lastly, synergizing HVMI with cloud tools represents a compelling roadmap for further research. This integration could automate security processes, providing rapid responses to emerging threats. Additionally, it would facilitate the swift deployment of security measures across cloud systems, bolstering their overall security posture. These directions mark a new exciting frontier in the ongoing effort to strengthen cloud security against ever-evolving and persistent cyber threats.

Resources For source code and further references, see the GitHub repository.

References

[1] F. Rehman, Z. Muhammad, S. Asif, H. Rahman, "The next generation of cloud security through hypervisor-based virtual machine introspection," in 2023 3rd International Conference on Artificial Intelligence (ICAI), 116–121, 2023, doi:10.1109/ICAI58407.2023.10136655.

[2] N. S. Shaikh, A. Yasin, R. Fatima, "Ontologies as Building Blocks of Cloud Security," International Journal of Information Technology and Computer Science (IJITCS), **14**(3), 52–61, 2022.

[3] J. Shahid, Z. Muhammad, Z. Iqbal, A. S. Almadhor, A. R. Javed, "Cellular automata trust-based energy drainage attack detection and prevention in Wireless Sensor Networks," Computer Communications, **191**, 360–367, 2022.

[4] M. Fatima, H. Abbas, T. Yaqoob, N. Shafqat, Z. Ahmad, R. Zeeshan, Z. Muhammad, T. Rana, S. Mussiraliyeva, "A survey on common criteria (CC) evaluating schemes for security assessment of IT products," PeerJ Computer Science, **7**, e701, 2021.

[5] S. Asif, M. Ambreen, Z. Muhammad, H. ur Rahman, S. Iqbal, "Cloud Computing in Healthcare-Investigation of Threats, Vulnerabilities, Future Challenges and Counter Measure," LC International Journal of STEM (ISSN: 2708-7123), **3**(1), 63–74, 2022.

[6] W. R. Simpson, K. E. Foltz, "Network Segmentation and Zero Trust Architectures," in Lecture Notes in Engineering and Computer Science, Proceedings of the World Congress on Engineering (WCE), 201–206, 2021.

[7] P. Purnaye, V. Kulkarni, "A comprehensive study of cloud forensics," Archives of Computational Methods in Engineering, **29**(1), 33–46, 2022.

[8] Z. Muhammad, F. Amjad, Z. Iqbal, A. R. Javed, T. R. Gadekallu, "Circumventing Google Play vetting policies: a stealthy cyberattack that uses incremental updates to breach privacy," Journal of Ambient Intelligence and Humanized Computing, 1–10, 2023.

[9] "Structured threat information expression (STIX™) 1.x archive website," .

[10] D. Barrett, G. Kipper, "2 - Server Virtualization," in D. Barrett, G. Kipper, editors, Virtualization and Forensics, 25–36, Syngress, Boston, 2010, doi: https://doi.org/10.1016/B978-1-59749-557-8.00002-3.

[11] Z. Aalam, V. Kumar, S. Gour, "A review paper on hypervisor and virtual machine security," in Journal of Physics: Conference Series, volume 1950, 012027, IOP Publishing, 2021.

[12] N. R. Nasab, "Security functions for virtual machines via introspection," 2012.

[13] B. Hay, K. Nance, "Forensics examination of volatile system data using virtual introspection," ACM SIGOPS Operating Systems Review, **42**(3), 74–82, 2008.

[14] B. D. Payne, D. d. A. Martim, W. Lee, "Secure and flexible monitoring of virtual machines," in Twenty-Third Annual Computer Security Applications Conference (ACSAC 2007), 385–397, IEEE, 2007.

[15] M. A. Kumara, C. Jaidhar, "Leveraging virtual machine introspection with memory forensics to detect and characterize unknown malware using machine learning techniques at hypervisor," Digital Investigation, **23**, 99–123, 2017.

[16] P. Mishra, E. S. Pilli, V. Varadharajan, U. Tupakula, "Intrusion detection techniques in cloud environment: A survey," Journal of Network and Computer Applications, **77**, 18–47, 2017.

[17] M. R. Watson, A. K. Marnerides, A. Mauthe, D. Hutchison, et al., "Malware detection in cloud computing infrastructures," IEEE Transactions on Dependable and Secure Computing, **13**(2), 192–205, 2015.

[18] A. K. Marnerides, P. Spachos, P. Chatzimisios, A. U. Mauthe, "Malware detection in the cloud under ensemble empirical mode decomposition," in 2015 international conference on computing, networking and communications (iCNC), 82–88, IEEE, 2015.

[19] H. Zhou, H. Ba, Y. Wang, T. Hong, "On the Detection of Malicious Behaviors against Introspection Using Hardware Architectural Events," IEICE TRANSACTIONS on Information and Systems, **103**(1), 177–180, 2020.

[20] H. Huseynov, K. Kourai, T. Saadawi, O. Igbe, "Virtual Machine Introspection for Anomaly-Based Keylogger Detection," in 2020 IEEE 21st International Conference on High Performance Switching and Routing (HPSR), 1–6, IEEE, 2020.

[21] B. Borisaniya, D. Patel, "Towards virtual machine introspection based security framework for cloud," Sādhanā, **44**(2), 1–15, 2019.

[22] S. Paakkola, "Assessing performance overhead of Virtual Machine Introspection and its suitability for malware analysis," 2020.

[23] Z. Muhammad, Z. Anwar, B. Saleem, J. Shahid, "Emerging Cybersecurity and Privacy Threats to Electric Vehicles and Their Impact on Human and Environmental Sustainability," Energies, **16**(3), 1113, 2023.

[24] S. Barnum, "Standardizing cyber threat intelligence information with the ... - stix," 2014.

[25] J. Shahid, Z. Muhammad, Z. Iqbal, M. S. Khan, Y. Amer, W. Si, "SAT: Integrated Multi-agent Blackbox Security Assessment Tool using Machine Learning," in 2022 2nd International Conference on Artificial Intelligence (ICAI), 105–111, IEEE, 2022.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 3, 108-117 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Accuracy Improvement-Based Wireless Sensor Estimation Technique with Machine Learning Algorithms for Volume Estimation on the Sealed Box

Kitipoth Wasayangkool, Kanabadee Srisomboon, Chatree Mahatthanajatuphat, Wilaiporn Lee*

Department of Electrical and Computer Engineering, Faculty of Engineering, King Mongkut's University of Technology North Bangkok, Thailand

ARTICLE INFO

Article history:
Received: 22 February, 2023
Accepted: 24 April, 2023
Online: 21 May, 2023

Keywords:
RSSI
Volume Estimation
Machine Learning
Neural Network

ABSTRACT

Currently, the quality and quantity of product must be inspected before transporting. Currently the popular unsealing box product inspecting is performed by weighing the box where the errors occur according to the tolerance of the weighting machine and tolerance weight of the product. On the other hand, the quantity of product can be inspected automatically using the image processing and recognition where the sealed box needs to be unpacked which is difficult to be implemented in practice. Then, the error in product transportation causes a loss profit of for the vendor and losing the reliability from customers. In this paper, we proposed a new volume estimation technique to estimate the product quantity in the sealed box using RSSI with machine learning for improving the monitoring accuracy. The proposed system includes one transmitter on the top and five receivers at bottom of the package. Based on practical environment, we align the product's pattern inside the boxed package into two cases including horizontal/vertical aligned pattern and random pattern. In the experiment, we compare the volume estimation accuracy of five machine learning techniques including linear regression, logistic regression, Naïve Bayes, KNN, and NN. From the results, the NN method provide the highest volume estimation accuracy among others and consumes the shortest estimation time. NN presents accuracy as 99.4% and consumes 6.51 milliseconds of estimation time. Moreover, for protecting the products from the delivery process, shockproof material must be put to cover the product in the box. Three shockproof types are considered in our experiment such as bubble wrap, paper, and airbag and our proposed system is considered all kinds of shockproof situations. The suggested method can estimate the volume of products without necessitating their opening or destruction. The suggested approach is also resistant to the impacts of packaging cushioning material.

1. Introduction

To the COVID-19 situation, the number of online purchases worldwide are grown up to 5.2 trillion US dollars and will reach 8.2 trillion US dollars in 2026. Figure 1 shows the rising trend of online marketing sales [1-3]. In Thailand, online shopping orders have increased by an average of 45% per year and they will rise with the rate steadily [4-5]. According to the new commodity trading behavior, customers order the products via online platform and the traders stag the products before sending to the customers. Once the number of orders is high, the error easily happens. Therefore, the bounce rate increased from approximately 8% in 2007 to 22% in 2017 and is anticipated to continue rising [6-8]. One of the reasons to return products is the incorrect number of product or a mismatch with the user order. Therefore, the monitoring quantity of the product inside the package before sending it to the customer is an important part of quality control and inspection to prevent mistakes and the associated expenses.

* Corresponding Author: Wilaiporn Lee, KMUTNB, wilaiporn.l@eng.kmutnb.ac.th

https://dx.doi.org/10.25046/aj080313

There is a requirement for varying quantities of merchandise since the behavioral characteristics of each transaction are distinct. Even though the contents of the parcel are the same, the number of objects in each parcel box will vary.

The entrepreneurs must verify their own cargo before delivering it to the carrier by implementing a quantity verification inspection procedure at the end. However, human error may occur if this strategy is used. In addition to losing energy, manufacturing delays may also result. Therefore, technology has been developed to assist with this verification.

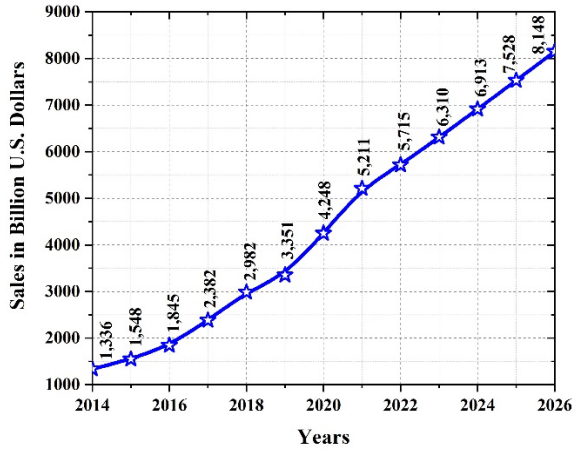


Figure 1: The ongoing expansion of internet shopping in 2014 – 2026.

Figure 2: RFID reader and RFID tag.

The number of products can be verified via many solutions such as directly counting by humans, image processing and recognition, RFID technology or calculate from the volume estimation methods. RFID scheme is one of the common methods used in worldwide. The RFID tag is attached to the products and used to count the number of products within the package. This approach is regarded as used with certain items; however, it cannot be used for some products such as water, seeds or many overlapping products because the RFID tag gets obscured, leading to measurement inaccuracies. In addition, this method is expensive due to the high cost of the RFID reader and the requirement to purchase RFID tags as the number of products. Figure 2 demonstrates an example of an RFID system [9-10]. Another technique is a weight meter. For the industry scale, the high-resolution weight meter is used for a quantity verification inspection procedure at the end that provides a high level of precision, but it is a premium cost that the high-resolution weight meter shows in Figure 3. For local commercial scales, the general weight meter is commonly used because this meter has low complexity and cost. However, this method has important restrictions the size of the items cannot exceed the scale's capacity, and it requires calibration and maintenance from the specialist inspector. Moreover, some products may have the same weights, but different volumes. Then, the only weight meter method may not be appropriate to check the number of products. On the other hand, if the volume of products can be calculated, the number of products can be exactly determined. Therefore, volume estimation should consider for estimating the number of products.

Figure 3: The inspector weighting machine.

Image-based volume estimation [11-12] is widely used in many researches. This method utilizes image processing techniques with image recognition to estimate the volume of an product. It requires to capture photographs from a variety of perspectives as well as adjust the lighting and clarity of the images. This method necessitates a visual inspection of the object to compute the volume, i.e., the package box must be opened. Consequently, image-based volume estimation method is unsuitable for determining the number of products already packed into a package.

Nowadays, the wireless sensors that operate on WI-FI technologies are unquestionably the most popular and frequently utilized today. It is a technology that is easily accessible and widespread. There are several uses of wireless sensor, such as applications for identifying the position of people or objects in indoor and outdoor environments [13-16], estimation of a person's posture inside a room [17-18], etc. There are several popular techniques for estimating the positions of objects, including the received signal strength indicator (RSSI), the time-of-arrival (TOA), the time difference of arrival (TDOA), and the angle of arrival (AOA) [19-24]. Each method has the unique advantages and disadvantages, including varying degrees of complexity, the precision accuracy, or the energy consumption. Due to its simplicity and low complexity, the RSSI method is the most utilized. The transmitter and receiver are the primary devices used to measure RSSI. The RSSI value delivered from the transmitter to the receiver will vary based on the environment, considering either the distance between the transmitter and receiver or the thickness of objects that obstruct the signal between the transmitter and receiver. These variations in RSSI readings can be utilized to estimate the position of a person or item [25-27]. From the RSSI measurement concept, we apply the RSSI value to estimate the volume of product in the sealed package for verified the quantities of merchandise.

In [28], we proposed the volume estimation method in the sealed package by applying the RSSI measurement. The proposed system installs 1 transmitter and 4 receivers to measure the RSSI value for improving the volume estimation accuracy by utilizing 4 machine learning algorithms including linear regression, logistic regression, Naïve Bayes, and K-nearest neighbors (KNN). From the experimental results, we found that the KNN method presented the highest volume estimation accuracy at 94%. However, regarding the arrangement of objects in the parcel box, only the horizontal arrangement was evaluated in this work. In the practical transit of packages, objects can be arranged in a variety of patterns. In addition, shockproof materials are normally put on the products for protecting it from damage in the delivery process. There are several types of cushioning materials; therefore, the performance and precision of the proposed volume estimation technique must consider the effect of the type of cushioning materials on the RSSI value.

In this paper, an RSSI measurement of wireless sensor and machine learning analysis-based system for estimating the product quantity under three types of sealed packages is presented. According to the RSSI characteristic, the signal is transmitted from the transmitter and the received signal is received by the receiver. Once the received signal strength is attenuated by the object which obstructs the transmission path, then we can utilize the received signal strength for estimating the amount of object in the sealed box without unpacking the box. We also evaluated the quantity of equipment, the number of products, and consider all of pattern of packed inside the package box in all instances. Four protection environments are considered including no cushioning material and the three commonly used cushioning materials: bubble wrap, paper, and airbags. For improving the accuracy, we apply five machine learning (ML) approaches, including linear regression, logistic regression, Naïve Bayes, KNN, and neural network (NN) for estimation of the volume of items in the material box. Based on the results of the experiment, the ML approach could be used to estimate the real volume of products with efficiency and precision. The provided system estimated the product with an accuracy of 99.40% using the NN method. However, the disadvantage is the long training time. On the other hand, the logistic regression is not appropriate to be implemented in volume estimation since it presents the lowest accuracy and consumes long estimation time. Regarding the optimal number of devices, it was determined that one transmitter and five receivers supported the greatest probability of placement. In calculating the number of items, the proposed system will be able to support the packaging of products in the absence of cushioning materials and with all three types of cushioning objects.

2. The Related Knowledge

In this section, we describe the RSSI calculation concept and five well-known ML methods. The considered ML methods are very popular on many applications and appropriate for the characteristic of RSSI data.

2.1. The Received Signal Strength Indicator (RSSI)

RSSI collecting techniques are utilized for a variety of different environments, including indoor or outdoor positioning systems, and determining distances based on RSSI measurements. In addition [29-32], the RSSI value has a low level of complexity and a short time to store the data, allowing low-cost and widely available storage devices to be utilized. Identifying the position of a person or object is an example of an application that uses the RSSI value to compute distance using the RSSI's distance measurement equation. The equation for computing the distance measurement from the RSSI level is shown in (1).

$$RSSI = 10log_{10}\frac{P}{1mW} \quad (1)$$

where P is the signal's strength (W) and $RSSI$ is the RSSI value (dBm). Typically, the RSSI will fluctuate as the distance between the transmitter and receiver changes or when barriers are present.

2.2. Machine learning technology

Machine learning is an intelligent system capable of analyzing data to propose or predict outputs to assist in decision-making. As the system accumulates more data, machine learning can help it make better decisions. This research utilizes linear regression, multinomial logistic regression, Naïve Bayes, KNN, and NN. Following is a detailed discussion of each technique.

2.2.1. Linear regression

Regression analysis is the application of machine learning where the prediction line of the learning model is generated by fitting the linear equation with the data. To estimate the output value, the new data is fed to the learning model which predict the output value from the generated prediction line. This technique is straightforward and efficient for studying linear data. It present a high degree of precision, then the linear regression equation is shown in (2).

$$y = \theta_0 + \theta_1 x_1 + \cdots + c \quad (2)$$

where y is the desired output, θ_i is the slope of the linear line of i data, x is the input feature, and c is the constant.

The mean square error (MSE) approach is applied as the cost function of the training process. From equation, the sum of the mean errors is shown in (3).

$$J(\theta) = \frac{1}{2m}\sum_{j=1}^{m}[y(x^{(j)} - y^{(j)})]^2 \quad (3)$$

where $x^{(j)}$ is the input feature, j is the point of the, and $y^{(j)}$ is the desired result.

2.2.2. Multinomial logistic regression

Logistic regression is a technique for predicting the binary output which is normally used in binary classification. In this study, we utilize data with the discrete outputs by implementing the multinomial logistic regression which can be considered as multi-class classification. This regression technique's core concept is to establish a decision boundary between the dataset to classify the type of data. If the data is evenly distributed, it will be difficult to classify it, and the error easily happens. When new data are obtained, the model decides based on previously created data lines. The multi-class logistic regression equation is illustrated in (4).

$$Y = \frac{1}{1+exp^{-(\alpha_0 x_0+\alpha_1 x_1+\cdots+c)}} \quad (4)$$

where Y is the probability that the input scenario occurs, α_i is the independent variable of i data, x is the clarify variable, and c is the constant.

The logistic regression equation is therefore modified by applying the cost function from equation (3). Logistic regression's cost function equation is presented as (5).

$$J(\theta) = -\frac{1}{m}[\sum_{j=1}^{m} y^{(j)} \log(y_\theta) + (1 - y^{(j)}) \log(1 - y_\theta)] \quad (5)$$

where y_θ is the preceding base for logistic regression, and $x^{(j)}$ is the considered data, and $y^{(j)}$ is the average result of the data considered.

2.2.3. Naïve Bayes

The Naïve Bayes method classifies the class of input data according to probability-based distribution of the data. In terms of the Naïve Bayes predictions, the conditional probability is determined based in the likelihood function to categorize the data class. The probability that the input data becomes a member of the considered class can be calculated as

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)} \quad (6)$$

where $P(y)$ is the probability of output class of the new data, $P(x)$ is the probability in each class of the dataset, $P(x|y)$ is the probability of the dataset correlating to new data and $P(y|x)$ is the probability of new data correlating to the dataset.

The solution in the Naïve Bayes method is typically presented as discrete data. According to the considered continuous data, the probability density function (pdf) distribution equation must be used, and it can be expressed as the following equation (7).

$$P(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) (7)$$

where μ is the mean value of the input data, and σ is the standard deviation of the input data.

2.2.4. K-Nearest Neighbor (KNN)

K-nearest neighbor compares the new data to the current data in the dataset to determine the similarity based on their distance. By predicting the result, KNN determines the class of new data from K nearest members of the data in the dataset. KNN measures the distance between the new data and the member in the dataset through the Euclidean distance (d) which is expressed in (5). In this approach, the processing time depends on the data volume of the dataset.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (8)$$

where d is the distance between the new data and the existing dataset, (x_1, y_1) is the point of the data in dataset, (x_2, y_2) is the point of the test data or new data.

2.2.5. Neural Network (NN)

A neural network is an artificial intelligence technique that trains a machine to process information in a manner base on the human brain. Deep learning is a method of machine learning that utilizes interconnected nodes or neurons, resembling the layered structure of the human brain. This produces an adaptive platform that allows computers to learn from their mistakes and continuously improve. Basically, NN consists of three layers: the input layer, the processing layer, and the output layer. The input layer mainly consists of one layer and delivers data to the hidden levels. The processing layer is a hidden layer that contains the number of neurons. Depending on a particular optimal, the number of neurons on hidden layers can be increased to improve precision. The output layer is the last layer responsible for weighting each neuron or set of outputs using the values from the hidden layer to achieve the most accurate results. The NN structure is seen in Figure 4.

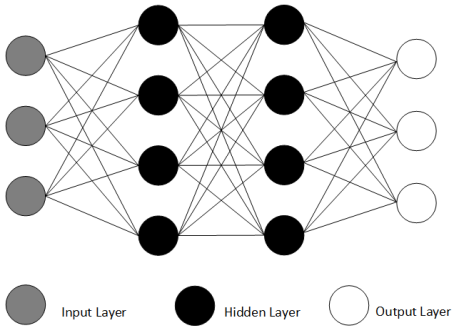


Figure 4: The structure of the neural network

The output of the previous step is fed to the system as the NN training process iterates over the data in the dataset to improve decision-making assume L is the output of the present hierarchy. In a process known as back propagation, the output of layer L is transmitted back to layer $L-1$. Stochastic gradient descent is used in this procedure to minimize the error which is illustrated in (9).

$$e = \frac{1}{2}\sum[t_k - a_k]^2 \quad (9)$$

where e is the error value, t_k is the average result for node k, and a_k is the error of the excitation effect from k nodes.

The equation for adjusting the decision weight is shown in (10).

$$\theta_{kj} = \theta_{kj} - \alpha\delta_j^L a_j^{L-1} \quad (10)$$

where θ_{kj} is an improvement in the hidden layer, and δ_j^L is the difference between the error value and the result.

When the sigmoid function is invoked, the equations for orienting the model configuration in the hidden layer appear in (11).

$$\delta_h^l = a_h^l(1 - a_h^l)\sum_j \delta_j^{l+1}\theta_{hj}^{l+1} \quad (11)$$

where δ_h^l is the result of considering all nodes in the hidden layer.

3. Methodology of the proposed system

In terms of measuring and storing RSSI values, the experiments were done in a noise-free chamber room, as seen in Figure 5. We use MCU ESP32 as a transmitter and the receivers. We install the transmitter on top of the station and the receivers are installed at the bottom of the station, and a test box is placed at the middle of the station. Then, the RSSI values are collected and transmitted to be further processed according to the procedure.

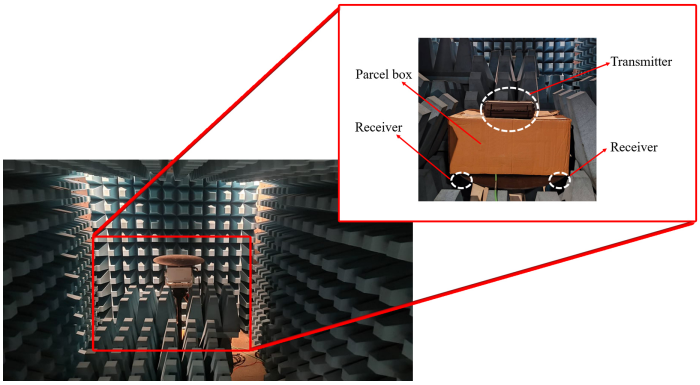


Figure 5: The proposed setup

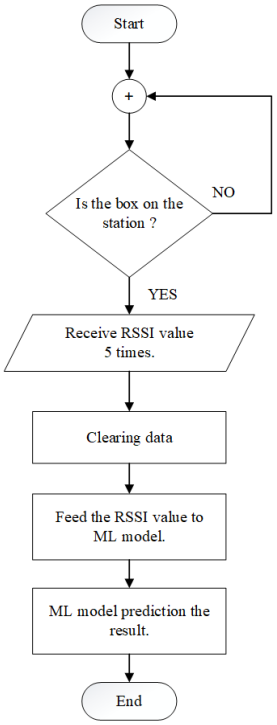


Figure 6: Flowchart of the proposed system

In this research, the RSSI value was collected in accordance with the flowchart depicted in Figure 6 , beginning with the determination of whether or not a package was present on the measurement station. When there is a package at the station, the RSSI value will be collected five times from each receiver. Then, the system will transmit it to the clearing data in order to to eliminate interference. The value will then be transmitted to a machine learning (ML) model for processing in order to forecast the result. The accuracy of estimation can be determined by

$$Acc = \frac{1}{m}\sum[y_{pred} - y_{actual}]^2 \quad (12)$$

where Acc is the accuracy rate, y_{pred} is the result of prediction, and y_{actual} is the actual value, m is amount of sample.

3.1. Experimental setup

In this section, we describe the experimental environment of the proposed system. In the experiment, the size of product is 8 cm. x 8cm. x 3 cm with the weight of 100 g. The size of box is 25 cm. x 35 cm. x 25 cm. To estimate the quantity of product in the sealed box, we collect the RSSI values from receivers under different volumes of product in the box where the determined volume are 0%, 20%, 40%, 60%, 80% and 100%. Then, we collect 2000 records of RSSI for each testing volume. Moreover, for weight measuring, we also collect 2000 records for each testing volume.

Firstly, we consider 5 cases of receiver's number that are 1 to 5 of receivers for finding the optimum RSSI value that cover all the object patterns. Secondly, we present the object patterns in which we arrange our object in two cases: pattern alignment and random pattern. Thirdly, we consider four our experimental environments including one non-protection product and three protection products. Additionally, we show the performance comparison of the proposed system under 5 ML methods including linear regression, logistic regression, Naïve Bayes, K-nearest neighbors, and neural network.

3.1.1. Receiver installation and the proportional number of receivers.

In the experiment, we consider the proportional number of receivers. We consider 5 cases by increasing the number of receivers from 1 to 5 receivers. For example, we placed a 4-receivers in each corner of the box, as depicted in Figure 7(a), and a 5-receiver by mounting an additional receiver in the middle of the box, as depicted in Figure 7(b), to examine the accuracy when the number of receivers varies.

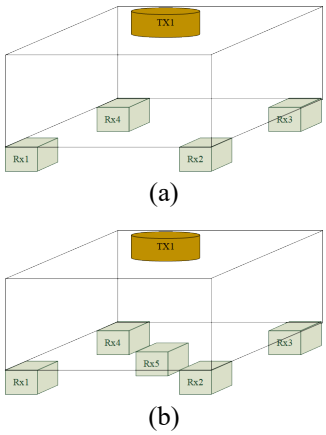


Figure 7: The example of 4 and 5 receivers in our experimental.

As illustrated in Figure 7, five receivers were recorded to establish the optimal number of sensors. For all considered ML methods, it was discovered that the efficiency of each method was better than other cases because a fifth receiver was added to the center of the bottom of the box. Then, we create a system with 1 transmitter and 5 receivers to accommodate all possible scenarios.

The performance comparison of volume estimation accuracy when varying the number of receivers depicted in Figure 8.

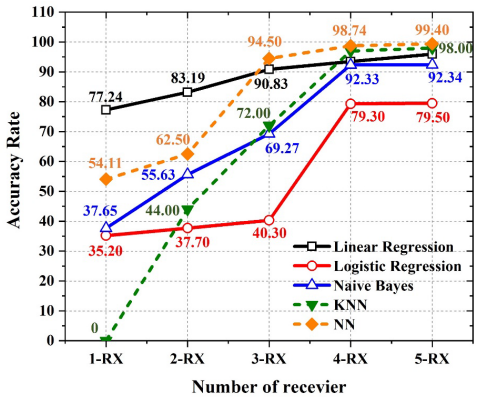


Figure 8: The performance comparison of volume estimation accuracy when varying the number of receivers.

3.1.2. The object alignment

In this section, the patterns of the object alignment is described which can be divided into two distinct categories, namely aligned pattern and random pattern. The item is arranged by increasing its volume from 0 to 100 percent in 10 percent increments. For instance, if the item occupies 10% of the box, we align the object in six situations, such as left, right, font, back, bottom, and random pattern, and then measure the RSSI between the transmitter and five receivers.

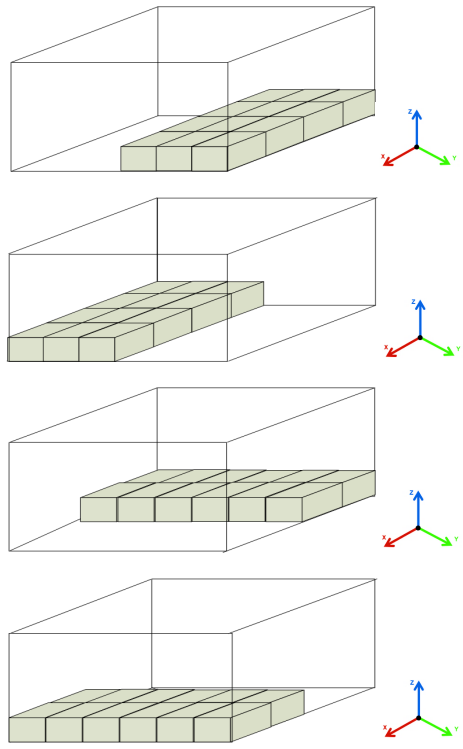


Figure 9: 10% of products' pattern that initial arrangement of items begins in the horizontal plane of the XY plane.

To examine the object layout within the box, we separated the arrangement of objects within the box into three distinct configurations. As illustrated in Figure 9, firstly, we present 10% of products' pattern that initial arrangement of items begins in the horizontal plane of the XY plane and subsequently increases along the Z axis to achieve 100 percent. The second pattern is to arrange the 10% of items on the YZ or XZ side planes and add up to 100 percent, as seen in Figure 10. In the final form, as seen in Figure 11, 10% of objects are randomly distributed in the box with no alignment between the objects. The organization of the items in accordance with this format is intended to facilitate their use in actual situations.

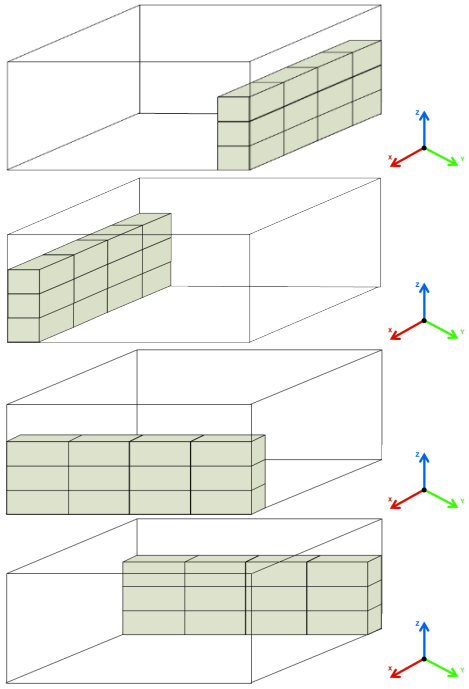


Figure 10: The second pattern is to arrange the 10% of items on the YZ or XZ side planes.

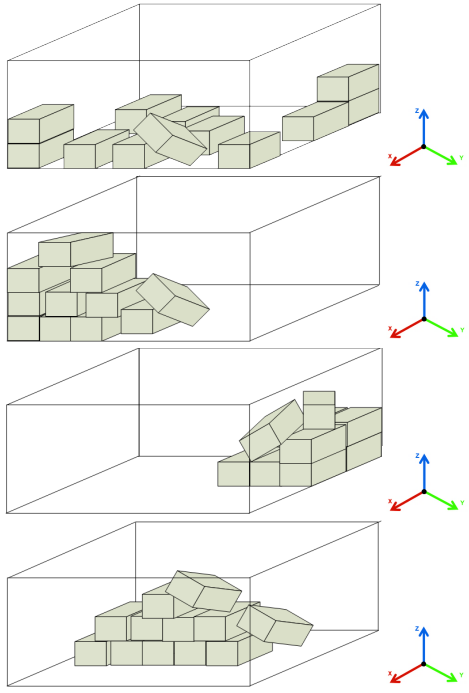


Figure 11: 10% of objects are randomly distributed in the box with no alignment between the objects

3.1.3. Data Cleansing

Before applying the ML techniques to estimate the volume of product, data cleansing is one important step to perform. When we obtain five RSSI values from 1 round of data accumulation, we evaluate the RSSI values, and we clean data by removing the RSSI value that differs from the measurement and replaced with the average from all other values before storing into the database. The clearing data algorithm is shown in Figure 12.

Table 1 shows the effectiveness comparison between the operating on data that was cleaned and not-cleaned. It was discovered that the volume estimation accuracy of the cleaned data is higher that the performance system of not-cleaned data. Moreover, from the results, we can conclude that cleansing data is an important procedure, especially for linear regression, KNN, and NN methods.

Table 1: Data outline clearing.

Techniques	The volume estimation accuracy of non-cleansing data (%)	The volume estimation accuracy of cleansing data (%)
Linear regression	85.70	95.93
Logistic regression	79.10	.7950
Naïve Bayes	82.46	.9240
KNN	94.80	98.00
NN	96.93	99.40

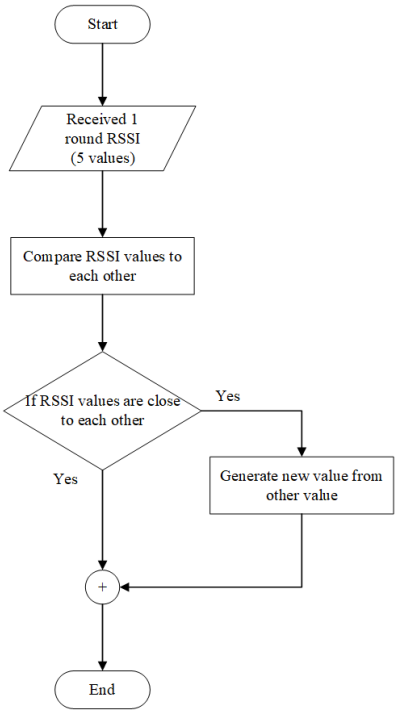


Figure 12: Flowchart of clearing data algorithm.

3.2. Comparison of methods efficiency

In our experiment, the volume estimation in the proposed system is considered under three object alignment including horizontally aligned, vertically aligned, and randomly patterns and we increase the number of items in the sealed package raising from 0% to 100% in 10% increments. The total records are 84,000 records that is separated to 70% for training dataset and 30% for testing dataset. The quality of the system was evaluated using the following five analytical methods: linear regression, logistic regression, Naïve Bayes, KNN, and NN.

The results of the experiments shown in Table 1 indicated that the efficiency of five receivers yielded the following results: the NN approach shows the highest accuracy, 99.40%, while KNN provides a value of 98.00%. The accuracy of linear regression and Naïve Bayes is 95.93% and 92.4%, respectively. The logistic regression also provides an accuracy of at least 79.50%. The investigation revealed that NN shows the highest accuracy because it learns the characteristic of RSSI information according to the object storing pattern and generates the weights through back propagation algorithm. It can be seen that characteristic of the RSSI values from each transmitter is not correlated to others then NN benefits from data from those transmitters. In the case of KNN, it shows a high accuracy since it estimates the output value by determining all data from the database. Consequently, the system achieves high precision.

In the case of linear regression, it generates a linear equation from an existing dataset and will use this equation to predict the output when the new data is coming into the system. In our experiment, this method gives a high precision of output estimation because the dataset of our system is a nearly linear pattern that proper to this method. Conversely, Naïve Bayes will do a probabilistic analysis of the new data. The new data will be verified that it has a probability to be similar to a group in the data in database and deciding to provide new information to that group based on that comparison, may improve the efficiency of data management. To categorize data types, the logistic regression method divides the data's boundaries using a line. If the data are highly similar, it will be unable to distinguish between the different types. Therefore, this procedure is less accurate than alternative ways.

Additionally, we determine the volume estimation time of each methods. Table 2 shows the time spent on each technique. It was discovered that the KNN approach did not require any training time, i.e. the KNN model measures simply the distance between old and new data to make decisions. On the other hand, it takes a long volume estimation time and will be longer if the database is huge. The Naïve Bayes consumes 14.72 milliseconds of training time since it only determines the average and standard deviation (SD) values to be used as the learning model. The linear regression approach and logistic regression consumes 19.10 milliseconds and 828.00 milliseconds, respectively. Finally, the NN consumes the longest training time, 365,060.00 milliseconds, since its training algorithm is highly complex.

Table 2: Time usage of 5 machine learning techniques.

Techniques	Time of training (ms)	Time of test (ms)
Linear regression	19.10	7.68
Logistic regression	828.00	7.24
Naïve Bayes	14.72	8.42
KNN	-	25.97
NN	365,060.00	6.51

From the aforementioned testing findings, it can be concluded that the approach with the best accuracy was the NN method and consumes the shortest estimation time of 6.51 milliseconds. Since the model will experiment by weighting each variable on the new

data set and generating predictions, it takes longer than other methods to predict. Next, the second most accurate method was KNN, which made predictions in 25.97 milliseconds. This approach used less time than the previous one since it calculated the distance between new and old data. However, if there is a huge quantity of historical data, this method will take a considerable amount of time to predict. The Naïve Bayes technique predicts in 8.42 milliseconds once it calculates the probability that the new data belong to the considered group. Linear regression and logistic regression predict in 7.68 and 7.24 milliseconds, respectively, using the new data mapping method to the linear line and data, respectively.

Based on the results of the experiments, it can be concluded that the NN methodology is the best method since it presents the most accurate and shortest volume estimation time, while logistic regression seems to be the worst. Therefore, the NN methodology can be considered as the most appropriate method to be implemented in the volume estimation application.

3.3. The effect of shockproof to RSSI

3.3.1. Type of shockproof

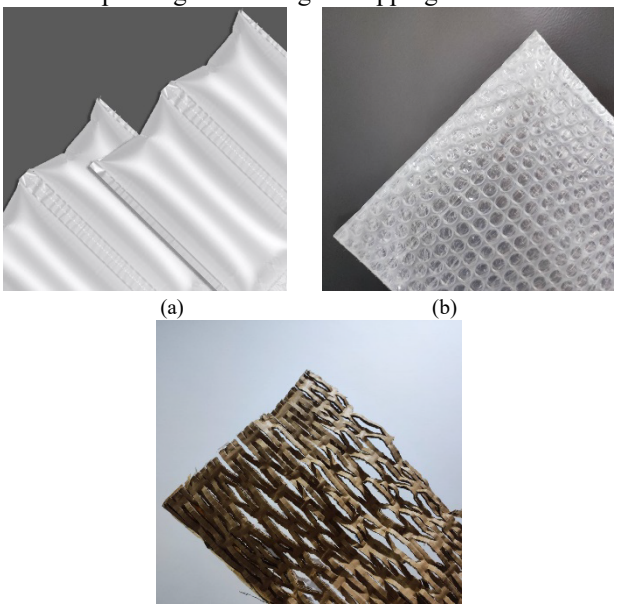
In addition, this research examines the effects created by shock-resistant materials utilized in transportation. As indicated in Table 3, the impacts of various shockproof materials such as bubble wrap, paper, and airbag are evaluated using RSSI values. In the case of bubble airbags filling the boxes to their maximum capacity, the RSSI values were observed to vary by 1 – 2 dBm relative to the empty boxes. The measured RSSI difference for the bubble wrap was between 5 and 7 dBm. In the latter instance, the difference was 7–10 dBm, as is the case with paper. The 100% of the capacity of the box has been padded with shock-absorbing material, according to all dimensions. Figure 13 depicts a shockproof material type.

Table 3: The differential of RSSI values with shockproof.

Type of Shockproof	RSSI value: Free space (dBm)	RSSI value: 100% of shockproof (dBm)	The differential RSSI values (dBm)
airbag	-26 to -27	-27 to -28	±1
Bubble wrap	-26 to -27	-31 to -33	±(7 – 5)
paper	-26 to -27	-33 to -36	±(10 – 7)

3.3.2. Accuracy rate with shockproof

Considering the real circumstances of transport that need the use of cushioning material to prevent potential damage to the object during transit, we pack the object with padding. Consequently, at this stage, measurements and processing were conducted to establish the impact that happens when the shockproof material fills the remaining space of the box. Airbag shockproof, bubble wrap and paper shockproof materials are popular to be used and we evaluate these materials in our experiment. The impact was precisely quantified and compared to a casing that lacked a shockproof coating. Table 4 demonstrates that the precision findings are like the previous results, i.e., in the case of the 5 receivers when the paper is used as a shockproof. The NN model has the greatest performance at 98.95%, followed by the KNN model at 97.25% and the linear regression model at 93.67%. The accuracy of Naïve Bayes and logistic regression was 90.22 and 79.03 percent, respectively. The precision of the other two materials, bubble wrap and airbag shockproof, is greater than that of paper. The airbag shockproof has the smallest effect on the RSSI quantification. The accuracy was determined to be (0-1) % different from that in the absence of cushioning material, and (1-2) % different from that in the presence of bubble wrap. In the last instance, the paper had an accuracy variation of (1-3) %. The conclusion is that RSSI measurement still produces accurate results when packing cushioning in shipping boxes.



(a) (b)

(c)

Figure 13: Types of shockproof.

In the previous section, the effectiveness of each cushioning material was evaluated. Different sorts of cushioning materials may be utilized in actual transportation environments. In this consideration, the next part of the experiment includes the situations of quantification without cushioning in comparison to all potential cases. There were 84,000 records of experimental data, which included non-shockproof, item with airbag shockproof, object with bubble warp, and paper shockproof. The test results are evaluated and shown in Table 4 by analyzing the RSSI data from all instances that were trained on different models. As a result, NN shows the highest performance for all shockproof environment while the logistic regression shows the worst. It can be mentioned that logistic regression and Naïve bayes are the learning algorithms for classification application therefore they show lower performance than others. Moreover, we compare the performance for all situation as depicted in figure 14. It was determined that the accuracy gained from the prediction of case 5 receivers using the NN approach was still the greatest at 99.07%, followed by KNN at 98% and linear regression at 95.10%. The results for Naïve Bayes and logistic regression were 92.20 percent and 79.28 percent.

Moreover, we compare the best and worst performance of the ML implemented techniques to the weight measuring technique which are shown in table 5. From the experimental results, weight

measuring technique presents 91.31% of accuracy where the tolerance value of the weight machine is 5%. As compared to the highest and lowest accuracy of volume estimation by ML techniques, the NN presents higher accuracy than weight measuring where logistic regression presents lower accuracy. It can be noticed that once NN learns the characteristic of RSSI values according to the product quantity, it present higher accuracy. On the other hand, the framework of logistic regression is not appropriate to predict the continuous value of quantity estimation, therefore it presents lower accuracy.

Table 4: The accuracy of volume estimation considering non-shockproof and three shockproof environments.

Techniques	Non-shockproof (%)	Airbag shockproof (%)	Bubble warp (%)	Paper shockproof (%)
Linear regression	95.93	95.45	93.82	93.67
Logistic regression	79.50	79.42	79.29	79.03
Naïve Bayes	92.34	92.14	90.95	90.22
KNN	98.00	98.00	97.55	97.25
NN	99.40	99.23	98.85	98.95

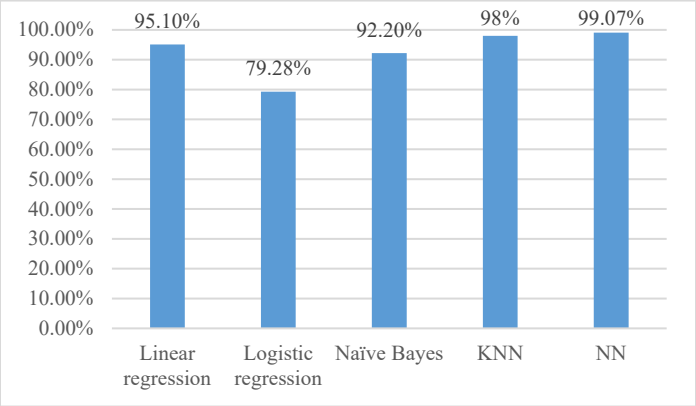


Figure 14: A performance comparison of overall performance

Table 5: Accuracy of unpack boxes volume estimation.

Accuracy rate of weighting machine (%)	Accuracy rate of logistic regression (%)	Accuracy rate of NN (%)	Accuracy rate of weighting machine (%)	Accuracy rate of logistic regression (%)
91.31	79.28	99.07	91.31	79.28

4. Conclusion and future work

In this study, an experiment to measure and estimate the quantity of products in non-open or destroy-sealed box containing different layouts of objects was presented. Using a receiver created from 5 nodes of MCU ESP32, a total of 84,000 records of RSSI are collected under noiseless environment and used to estimate the quantity of product using the ML techniques including linear regression, logistic regression, Naïve Bayes, K-nearest neighbors, and neural network. The impact of the packaging's cushioning material and various patterns of product in the sealed package are examined. Moreover, we compare the performance of volume estimation without the box unpacking technique which is done by weight measuring to the proposed method. From the experimental results, once the location of receivers are set appropriately where the coverage area of signal covers all area of the box, the measured RSSIs can be used to estimate the quantity of product in the sealed box. Among the determined ML techniques, NN shows the highest accuracy once it learns the characteristic of RSSIs under different product quantity to generate the weights of the algorithm. On the other hand, logistic regression shows the lowest accuracy since the learning characteristic is appropriate to the classification application where the output is discrete. By comparing to the weight measuring, it can be noticed that the proposed method is better than the weight measuring since the accuracy of the weight measuring depends on the device and the specific tolerance. Once the tolerance is high, the error also high. Moreover, the proposed method consumes short product quantity estimation time without unpacking or destructing the box.

In future work, improving the accuracy and adjusting the data processing algorithms to be more resistant to the arrangement of things within the package and to be utilized in a loud environment, as well as developing a system that is adaptable to changing parcel sizes, are the next steps.

Acknowledgment

This research was funded by King Mongkut's University of Technology North Bangkok. Contract no. KMUTNB-PHD-63-03.

References

[1] T. Anusha, "Trends and Analysis of E-Commerce Market: A Global Perspective," International Journal of Applied Marketing and Management, **6**, 11-22, 2021.

[2] M. Aghaei, A. Mohammad, R. Sahebi, "Investigating the Change in Customers, Sustainable Consumption Behaviour after the Outbreak of COVID-19," International Journal of Applied Marketing and Management, **6**, 34-49, 2021.

[3] B. Anam, H. Akram, H.M. Basit, A.U. Khan, "E-commerce trends during COVID-19 Pandemic," International Journal of Future Generation Communication and Networking, **13**, 2020.

[4] P. Goyal, K. Goyal, A. Chauhan, "Emerging Trends of E-Commerce perspective in Developing Countries," Journal of Positive School Psychology, **6**(4), 7090 – 7096, 2022.

[5] R. Somdech, T. Sriyakul, K. Jermsittiparsert, "The Era of e-Commerce & Online Marketing: Risks Associated with Online Shopping," International Journal of Innovation, Creativity and Change, **8**, 201-221, 2019.

[6] K. Chueamuangphan, P. Kashyap, C. Visvanathan, "Packaging Waste from E-Commerce: Consumers Awareness and Concern," in Ghosh S. (eds) Sustainable Waste Management: Policies and Case Studies Springer, **1**, 27-41, doi: 10.1007/978-981-13-7071-7.

[7] B. Stöcker, D. Baier, B.M. Brand, "New insights in online fashion retail returns from a customers' perspective and their dynamics," The Journal of Business Economics (JBE), **91**, 1149–1187, 2021.

[8] T. Rintamäki, M.T. Spence, H. Saarijärvi, J. Joensuu, M. Yrjölä, "Customers' perceptions of returning items purchased online: planned versus unplanned product returners," International Journal of Physical Distribution & Logistics Management, **51**(4), 403-422, doi: 10.1108/IJPDLM-10-2019-0302.

[9] M. Truijens, X. Wang, H. de Graaf, J.J. Liu, "Evaluating the Performance of Absolute RSSI Positioning Algorithm-Based Microzoning and RFID in Construction Materials Tracking," Mathematical Problems in Engineering, **2014**(3), 2014.

[10] K. Bouzaffour, B. Lescop, P. Talbot, F. Gallée, S. Rioual, "Development of an Embedded UHF-RFID Corrosion Sensor for Monitoring Corrosion of Steel in Concrete," IEEE Sensors Journal, **21**(10), 12306-12312, 2021.

[11] Ifmalinda, R.E. Putri, I. Rasinta, "Estimation of Size Volume and Weight of Oranges Using Digital Images Processing," in The 4th International Conference on Sustainability Agriculture and Biosystem, **100**(2), 315-321, 2010, doi: 10.1088/1755-1315/1059/1/012016.

[12] Y. Sun, Z.X. Liu, M. Li, Z.T. Zeng, Z.X. Zong, C.L. Ji, "An Object Recognition and Volume Calculation Method Based on Yolov3 and Depth Vision," Journal of Physics Conference Series, **1684**(1), 2020, doi:10.1088/1742-6596/1684/1/012009.

[13] A. Poulose, D.S. Han, "Performance Analysis of Fingerprint Matching Algorithms for Indoor Localization," in 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 661-665, 2020.

[14] M. Taha, R. Atallah, O. Dwiek, F. Bata, "Crowd Estimation Based on RSSI Measurements Using kNN Classification," in 2020 3rd International Conference on Intelligent Autonomous Systems (ICoIAS), 67-70, 2020.

[15] K. Coussement, S. Lessmann, G. Verstraeten, "A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry," Decision Support Systems, **95**, 27-36, 2017.

[16] R. D. Ainul, S. Wibowo, Djuwari, M. Siswanto, "An Improved Indoor RSSI Based Positioning System Using Kalman Filter and MultiQuad Algorithm," in 2021 International Electronics Symposium (IES), 558-564, 2021, doi:10.1109/IES53407.2021.9594009.

[17] P.D. Tinh, T.T.N. Mai, "Mobile Indoor Positioning System Utilizing WiFi RSSI and Motion Data," in 2020 4th International Conference on Recent Advances in Signal Processing Telecommunications & Computing (SigTelCom), 98-102, 2021, doi:10.1109/SigTelCom49868.2020.9199021.

[18] J. Ren, Y. Wang, W. Bai, C. Niu, S. Meng, "An improved indoor positioning algorithm based on RSSI filtering," in 2017 IEEE 17th International Conference on Communication Technology (ICCT), 1136-1139, 2017, doi:10.1109/ICCT.2017.8359812.

[19] W. Wang, Y. Zhang, L. Tian, "TOA-based NLOS error mitigation algorithm for 3D indoor localization," China Communications, **17**(1), 63-72, 2020.

[20] S. Jeong, T.K. Sung, K.E. Lee, J. Kang, "Joint TOA/AOA-based localization in wireless sensor networks," in 2014 8th International Conference on Signal Processing and Communication Systems (ICSPCS), 1-5, 2014.

[21] B. Al-Qudsi, M. El-Shennawy, Y. Wu, N. Joram, F. Ellinger, "A hybrid TDoA/RSSI model for mitigating NLOS errors in FMCW based indoor positioning systems," in 2015 11th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME), 93-96, 2015, doi:10.1109/PRIME.2015.7251342.

[22] F. B. Günay, T. Çavdar, "Mobile fleet localization model via RSSI, TOA and TDOA in wireless sensor networks," in 2014 22nd Signal Processing and Communications Applications Conference (SIU), 1431-1434, 2014, doi:10.1109/SIU.2014.6830508.

[23] M. I. Jais, P. Ehkan, R. B. Ahmad, I. Ismail, T. Sabapathy, M. Jusoh, "Review of angle of arrival (AOA) estimations through received signal strength indication (RSSI) for wireless sensors network (WSN)," in 2015 International Conference on Computer, Communications, and Control Technology (I4CT), 354-359, 2015, doi:10.1109/I4CT.2015.7219597.

[24] K.J. Baik, S. Lee, B.J. Jang, "Hybrid RSSI-AoA Positioning System with Single Time-Modulated Array Receiver for LoRa IoT," in 2018 48th European Microwave Conference (EuMC), 1133-1136, 2018, doi:10.23919/EuMC.2018.8541736.

[25] N. Raj, "Indoor RSSI Prediction using Machine Learning for Wireless Networks," in 2021 International Conference on COMmunication Systems & NETworkS (COMSNETS), 372-374, 2021, doi:10.1109/COMSNETS51098.2021.9352852.

[26] M. Schroth, T. Etkin, W. Stork, "A novel approach for human activity recognition using object interactions and machine learning," in 2021 IEEE Sensors Applications Symposium (SAS), 1-6, 2021, doi:10.1109/SAS51076.2021.9530029.

[27] B. Zheng, T. Masuda, T. Shibata, "An Indoor Positioning with a Neural Network Model of TensorFlow for Machine Learning," in 2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), 1-2, 2021, doi:10.1109/ISPACS51563.2021.9651131.

[28] K. Wasayangkool, K. Srisomboon, W. Lee, "The Volume Estimation Technique using RSSI with Machine Learning in Manufacturing Process," in 2022 37th International Technical Conference on Circuits/Systems Computers and Communications (ITC-CSCC), 1-4, 2022. doi: 10.1007/978-981-13-7071-7.

[29] A. Shojaifar, "Evaluation and Improvement of the RSSI-based Localization Algorithm: Received Signal Strength Indication (RSSI) Dissertation," Computing and Intelligent Systems, **2**, 10-12 2015.

[30] H. Ahmadi, R. Bouallegue, "Exploiting machine learning strategies and RSSI for localization in wireless sensor networks: A survey," in 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), 1150-1154, 2017, doi:10.1109/IWCMC.2017.7986447.

[31] P. Koovimol, S. Pattaramalai, "Experimental Machine Learning for RSSI fingerprint in Indoor WiFi Localization," in 2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 1018-1021, 2021, doi:10.1109/ECTI-CON51831.2021.9454865.

[32] D.O. Camacho, L.M. Proaño, "Increasing Accuracy in Positioning by RSSI: an Analysis with Machine Learning Algorithms," in 2019 International Conference on Information Systems and Computer Science (INCISCOS), 31-35, 2019, doi: 10.1109/INCISCOS49368.2019.00014.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 4, 111-121 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Colorized iVAT Images for Labeled Data

Elizabeth Dixon Hathaway[1,*], Richard Joseph Hathaway[2]

[1]*Department of Health & Human Performance, University of Tennessee at Chattanooga, Chattanooga, 37403, United States*

[2]*Department of Mathematical Sciences, Georgia Southern University, Statesboro, 30460, United States*

ARTICLE INFO

Article history:
Received:19 May, 2023
Accepted:19 July, 2023
Online: 26 August, 2023

Keywords:
Data visualization
Cluster heat map
iVAT image
Clustered tendency
Labeled data

ABSTRACT

A 2-dimensional numerical data set $X = \{x_1,\ldots,x_n\}$ with associated category labels $\{l_1,\ldots,l_n\}$ can be accurately represented in a 2-dimensional scatterplot where color is used to represent each datum's label. The colorized scatterplot indicates the presence or absence of spatial clusters in X and any special distribution of labels among those clusters. The same approach can be used for 3-dimensional data albeit with some additional difficulty, but it cannot be used for data sets of dimensions 4 or greater. For higher dimensional data, the improved Visual Assessment of cluster Tendency (iVAT) image can be used to indicate the presence or absence of cluster structure. In this paper we propose several new types of colorized iVAT images, which like the 2-dimensional colorized scatterplot, can be used to represent both spatial cluster structure and the distribution of labels among clusters.

1. Introduction

Let $O = \{o_1,\ldots,o_n\}$ be a set of n objects (e.g., people, trees). Suppose we have two types of information about the objects in O: (1) a set of corresponding *object data* $X = \{x_1,\ldots,x_n\} \subset \Re^s$, where for each k, datum x_k gives values for s different *features* (e.g., height, weight) of object o_k; and (2) a set of *labels* $L = \{l_1,\ldots,l_n\}$ indicating to which of several known *categories* each object belongs. The purpose of this paper is to introduce simple visual displays that allow us to compare the *cluster structure* in X with the *category structure* in L. Here we use the terms *cluster* or *spatial cluster* to refer to a group of close, similarly valued data points in X. Since we are dealing with visual displays, our eyes can be used to assess cluster structure. An interesting discussion of definitions of clusters from both the human and computer points of view is given in [1]. The potential uses of the new displays include: (1) checking to see if a cluster-based classification scheme (e.g., nearest centroid classification) might work well for labeling future objects; (2) providing a visual approach to feature selection for X by comparing displays using different subsets of features; (3) providing a simple visual approach for identifying clusters of interest, and (4) simply checking to see if the cluster structure of X aligns with the category structure of L.

In the case of s = 2 features we do not need a new approach as we can represent all the available information of object data X and labels L in a single, colorized scatterplot. We illustrate this below in Fig. 1 with a scatterplot of the hypothetical school data example from [2] involving 26 students (the objects) with their corresponding scaled SAT and high school GPA scores (the object data X) and their freshman math course outcomes of Pass or Fail (the labels or categories L). Note that the scatterplot reveals the clusters of the object data X along with the distribution of the labels among those clusters. Stated another way, we see that the colorized scatterplot allows us to check the *alignment* of the object data clusters with the label categories. Ignoring color, we can reasonably identify two spatial clusters: a lower left cluster (lower achievers) and an upper right cluster (higher achievers). Now also considering color, we see that the Fail category mostly aligns with the lower achievers while the Pass category mostly aligns with the upper achievers. We will refer to a datum x_k as a *minority point* if it is labeled differently from most of its nearest neighbors in X, and with this term we note that there are two minority points among lower achievers and one among higher achievers. The scatterplot makes it very clear which cluster should be targeted for additional learning support, which is the group of students who achieved less in high school.

A colorized scatterplot is all that is needed for labeled 2-dimensional data, but how do we display similar visual information for labeled data with s > 2? It is certainly possible to look at many colorized scatterplots of 2-dimensional slices of X but this can give misleading or at least incomplete information

* Corresponding Author: Elizabeth Dixon Hathaway, elizabeth-hathaway@utc.edu

about the actual situation when all s features are simultaneously considered. In 2 dimensions the simple (uncolored) scatterplot serves as the canvas to which we add color in order to display both object data and label information. What is the corresponding canvas for higher dimensional object data? The answer given here is the improved Visual Assessment of (cluster) Tendency (iVAT) image [3,4] or a similar image derived from it. In a single grayscale image iVAT is able to represent much of the cluster information for object data of any dimension, and so we choose it as a reasonable scatterplot substitute when s > 2. Color is added according to a variety of schemes given later. Presenting and testing these new Colorized iVAT (CiVAT) and related images is the purpose of this paper.

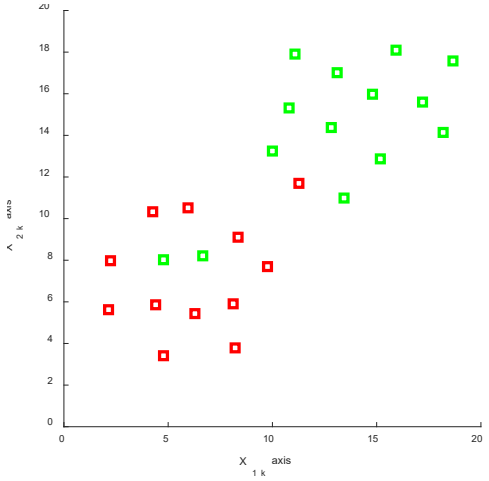


Figure 1. Green indicates Pass and Red indicates Fail.

We close this section by describing the organization of the remainder of this paper. Section 2 gives the necessary background regarding notation, terminology and iVAT images. The next section introduces several CiVAT and related approaches. Section 4 gives 10 examples using the new image-based techniques, seven of which use 2-dimensional object data. The point of using 2-dimensional examples is to provide training on the proper interpretation of the iVAT-based images by directly comparing the images with available colorized scatterplots; this direct comparison cannot be done when the object data dimension s ≥ 4. The final section gives some additional discussion and questions for future research.

2. Background

Our notation for objects O = {o_1,o_2,…,o_n}, object data X = {x_1,x_2,…,x_n} and category labels L = {l_1,l_2,…,l_n}were given in the last section. Additionally, the number of clusters is denoted by c and the number of label categories is denoted by d. In any particular example it is not necessarily true that the clusters of X align with the categories of L or even that c = d. An alternative to object data X that is very important in the following is called relational data, which is represented using an n × n matrix R, whose ijth entry R_{ij} = the relationship between objects o_i and o_j. With relational data, rather than describing objects directly (using the features of object data X), an object is described by saying how similar or dissimilar it is to each of the other objects. Relational data is natural in some cases such as when you try to describe the music of a particular musical band by saying how much it is like or unlike the music of other known bands.

In this paper the type of relationship described by R will always be dissimilarity between pairs of objects in O. Sometimes there may only be relational data R available and no object data X, but whenever there is object data, it is very easy to get corresponding relational data. In all of the following we will convert X to R using squared Euclidean distances:

$$R_{jk} = \text{dissimilarity}(o_j,o_k) = (x_{1j}-x_{1k})^2 + (x_{2j}-x_{2k})^2 + \ldots + (x_{sj}-x_{sk})^2, \text{ for } 1 \le j,k \le n \quad (1)$$

We mention that the much more difficult inverse problem of going from R to some object data set X is addressed using multidimensional scaling [5]. Note that the relational data matrices based on (1) will always satisfy for $1 \le j,k \le n$:

$$R_{jj} = 0 \quad (2a)$$

$$R_{jk} \ge 0 \quad (2b)$$

$$R_{jk} = R_{kj} \quad (2c)$$

Relational dissimilarity matrices are important to us because we can visually represent them using a single grayscale digital image, which leads us to a discussion of the original Visual Assessment of cluster Tendency (VAT) image from [6]. To represent a dissimilarity matrix R as a digital image, we interpret each matrix element R_{jk} as the gray level of the (j,k) pixel in the image. Sometimes we also refer to the actual pixel in row j column k of the image by R_{jk}. To make later colorization schemes simpler to describe we will assume that the dissimilarity matrix R has been scaled by dividing each element by the largest element of the original (unscaled) R, which means that the largest element in the new scaled R is 1 and the smallest is 0. In the representation of R as an image, the value 1 generates a white pixel and 0 generates a black one. Values in (0,1) generate various shades of gray. All images generated for this paper were done using the image display command **imagesc** in Matlab.

Does an image display of R in (1) give interpretable, visual information about the cluster structure in the object data X used to generate R? It turns out that the ordering of the data in X, equivalently the ordering of the rows and columns of R, is crucial to the answer to this question. The VAT procedure reorders the rows and columns of R to correspond to a reordering of the object data in X so that nearby data points are numbered consecutively (as nearly as possible). Precisely, we relabel the data according to the following scheme. The first reordered data point $x_{(1)}$ is taken to be a datum on the outer fringe of the data set X. Then the second (reordered) data point $x_{(2)}$ is chosen to be the remaining one closest (i.e., least dissimilar) to $x_{(1)}$. Then $x_{(3)}$ is chosen as the remaining point closest to any previously selected point (either $x_{(1)}$ or $x_{(2)}$). This is continued at each step by adding the next unchosen point that is closest to the set of previously chosen ones. In practice the reordering of the data of X is done virtually through the permutation of rows and columns of R, eventually producing the VAT reordered R*, as stated later in the VAT algorithm. The VAT procedure for reordering is related to Prim's Algorithm [7] for generating the minimal spanning tree of a graph, and this is discussed in [1].

We demonstrate the importance of ordering by revisiting the school data shown in Fig. 1 using figures taken from [2]. Fig. 2 gives a scatterplot of the data in its original, random ordering, along with the image for its corresponding relational data R, calculated using (1). In the scatterplot of Fig. 2(a), the first ordered point is represented by a red square and the rest of the ordering is represented using line segments connecting consecutively ordered data.

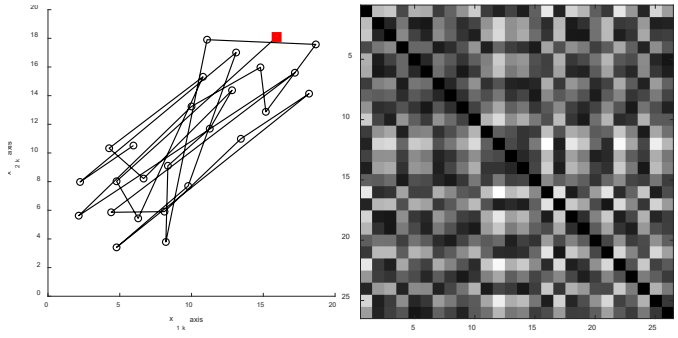


Figure 2: (a) Original X ordering starting with red square. (b) Image of R (uses the original X ordering).

Does the image in Fig. 2(b) usefully describe the cluster structure of the data? No. We cannot determine the number of clusters, the numbers of data in the clusters, the separation of the clusters, or even if there are any clusters. Now we will consider what happens if we reorder the data X (equivalently, the rows and columns of R) according to the VAT procedure. Fig. 3 gives the VAT reordered scatterplot for the data along with a display of the image corresponding to the VAT reordered R*. Notice in Fig. 3(a) that the first 13 data points are relatively close to each other as are the last 13 points. This closeness should generate darker pixels in the top left and bottom right 13 × 13 diagonal blocks of the image. Additionally, the relatively larger inter-cluster distances should produce lighter pixels in the other areas of the image. Fig. 3(b) confirms our expectations and gives representations of the two clusters as the two dark diagonal blocks. Appreciate that VAT produces this type of information directly from R, and there is no dependence on the actual dimensionality of the object data X.

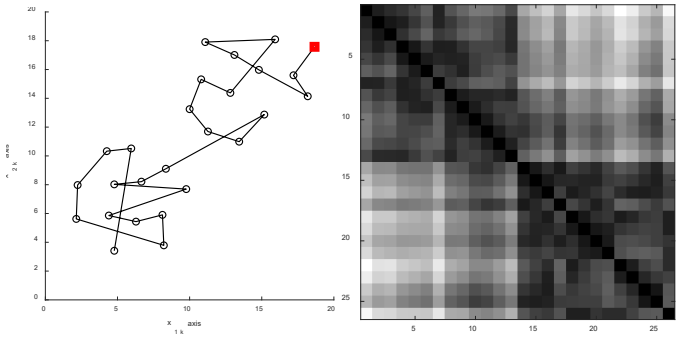


Figure 3: (a) VAT reordering starting with red square. (b) VAT image (uses the VAT reordered R*).

After seeing a few examples, it becomes easy to interpret VAT images in many cases. Blocks on the diagonal correspond to clusters. The number of clusters corresponds to the number of diagonal blocks. The numbers of objects in clusters corresponds to the sizes (i.e., the number of rows or columns of pixels) of the diagonal blocks. The degree of separation between clusters corresponds to the degree of grayscale contrast between diagonal blocks and off-diagonal block areas. When the clusters are compact and well separated, the VAT image shows a stark contrast of nearly black diagonal blocks with nearly white areas elsewhere. The VAT image in Fig. 2 is somewhat muddy in appearance because the clusters are not compact and well separated. The VAT algorithm follows.

Algorithm 1. VAT Reordering [6]

Input: R—n × n dissimilarity matrix

Step 1 Set J = {1, 2, …, n}, I = ∅; P(:) = (0, 0, …, 0) .

Step 2 Select $(i,j) \in \underset{p\in J,\ q\in J}{\arg\max} \{R_{pq}\}$.

Set P(1) = i; I = {i}; and replace J ← J – {i} .

Step 3 For r = 2, …, n:

Select $(i,j) \in \underset{p\in I,\ q\in J}{\arg\min} \{R_{pq}\}$.

Set P(r)=j; replace I←I∪{j} and J←J–{j}

next r.

Step 4 Obtain the VAT reordered dissimilarity matrix R^* using the ordering array P as $R^*_{ij} = R_{P(i)P(j)}$ for $1 \le i, j \le n$.

It is possible to use the VAT image as the basis for colorized versions but it turns out that there is an enhanced version, known as improved VAT (iVAT), which was first proposed in [3] and is based on replacing the original dissimilarities with the set of derived minimax distances. It is useful to understand how the minimax data are defined. Think of a complete graph with vertices $\{v_1,\ldots v_n\}$ and interpret R as giving edge lengths (i.e., distances) between each pair of vertices; that is R_{jk} is the direct distance, i.e., edge length, from v_j to v_k. What is the minimax distance from v_j to v_k? Let p be any path from v_j to v_k and let e_p be the maximum edge length (from R) used in path p. Then the minimax distance between vertices v_j to v_k is defined to be the minimum value of e_p over all possible paths p from v_j to v_k, and this value is taken to be the new minimax dissimilarity (or distance) between x_j and x_k.

There are predictable effects resulting from the switch to minimax data. First, all points within a cluster become nearer and more equidistant to each other, and this often provides more uniformity of tone in the diagonal blocks. Second, many or all points within one cluster share a common minimax distance to points in another cluster, which also gives off-diagonal block areas a more uniform appearance. Usually there is more observable contrast between diagonal blocks and off-diagonal areas using minimax data. Lastly, it is important to note that while both Euclidean and minimax distances work great for separated hyperspherical clusters, it is true that minimax distances can better handle cases of chainlike and stringy clusters by keeping intra-cluster pairwise distances more uniformly small. Fig. 4 gives the iVAT image corresponding to the school data.

Notice the increased uniformity of both diagonal and off-diagonal blocks. Note also the high degree of contrast. The **imagesc** command in Matlab does a virtual rescaling of the matrix range of values to [0,1] so that every VAT or iVAT image produced by it will include black pixels on the diagonal and at least one pure white pixel off the diagonal. Comparing Fig. 4 to Fig. 3(b) we see the usual situation; iVAT images provide cleaner and clearer descriptions of cluster structure than VAT images. For this reason, we choose to use iVAT as the basis for our higher dimensional substitute for scatterplots; it will represent the cluster

structure and added color will represent the label category information.

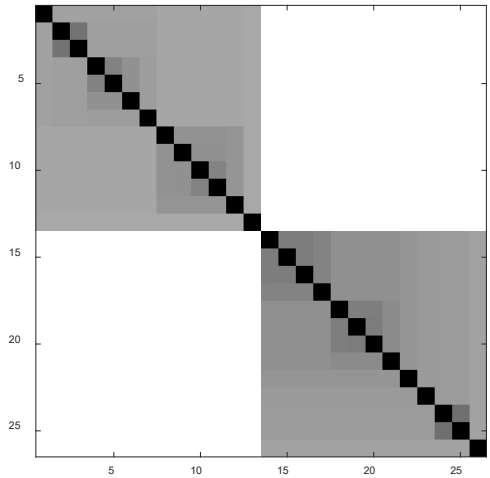


Figure 4: iVAT image (uses the VAT reordered minimax distances).

The originally proposed iVAT from [3] is done in a relatively inefficient manner. It is shown in [4] that greater efficiency (i.e., an order of magnitude of complexity) can be gained by first applying VAT to the original relational data R to get the VAT reordered R*, and then finding the minimax form of R* which we denote throughout the following by R'*. The special properties of a VAT reordered R* allow certain opportunities to gain efficiency in calculating the minimax distances. To be clear, the efficient technique in [4] produces an equivalent result to that obtained by finding the minimax version of R and then applying VAT. The efficient approach for minimax distance calculation is given next in Algorithm 2.

Algorithm 2. Efficient iVAT Calculation of Minimax Dissimilarities [4]

Input: R^*—VAT reordered n × n dissimilarity matrix
Step 1 Initialize minimax dissimilarity matrix $R'^* = [0]^{n\times n}$
Step 2 For r = 2, …, n

$$j = \underset{k=1,\ldots,r-1}{\arg\min} \{R^*_{rk}\}$$
$$R'^*_{rc} = R^*_{rc}\ , c = j$$
$$R'^*_{rc} = \max\ \{R^*_{rj}\ , R'^*_{jc}\ \}, c = 1, \ldots, r\text{-}1, c \neq j$$

next r
Step 3 R'^* is symmetric, thus $R'^*_{rc} = R'^*_{cr}$

To calculate an iVAT image for an original set of randomly ordered relational data R, we first apply Algorithm 1 to transform R to the VAT reordered R*, and then apply Algorithm 2 to R* to get the final iVAT matrix R'*. The display of R'* gives a nice image as in Fig. 4. A thorough and recent survey of VAT-based visualization schemes is given in [8], and discussion relating VAT to clustering methodology can be found in [1]. The VAT and iVAT images are part of a much larger set of visualization methodology known as cluster heat maps. A useful partial survey of heat maps through 2009 is found in [9]. In the next section we put color to the canvas.

3. Ordering and Colorization Schemes

In this section we introduce four schemes for adding color to the iVAT image R'*: two that use exactly the iVAT matrix R'* and two that use a label-reordered version of it. Note that the two using a label-reordered version still require application of iVAT in order to efficiently get the minimax distances which give us the clearest pictures.

We start by using an example to describe exactly what is meant by the Label Reordering (LR) of iVAT. In our example suppose that R'* is produced by iVAT and that the column and row orderings of R'* correspond to the object ordering

$$o_3, o_8, o_2, o_5, \mathbf{o_4}, o_1, o_7, o_6. \quad (3)$$

Also suppose that the corresponding label array for this iVAT ordering of objects is L* = [2, 1, 1, 2, **1**, 2, 1, 1]. We can readily see in this example that there are 8 objects and 2 categories and can easily determine the category of each object. For example, since o_4 is the fifth object in the iVAT ordering of (3) and L*(5) = 1, we know that o_4 is in category 1. What exactly is the LR for this example? It is the ordering obtained by first pulling out all the category 1 objects from (3), going from left to right, and then concatenating to it the list of all category 2 objects, going from left to right. Doing this for the example gives the LR order of objects

$$o_8, o_2, o_4, o_7, o_6, o_3, o_5, o_1. \quad (4)$$

In doing the LR reordering the objects are never actually referenced. Instead, L* is used to find the permutation that would take (3) to (4) and then the permutation is applied directly to L* and the rows and columns of R'*. Note that the LR permuted form of L* is simply [1, 1, 1, 1, 1, 2, 2, 2]. The process is continued in the obvious way if there are more than 2 categories.

So the two ordering schemes used later are iVAT and LR. Before stating the two colorization schemes we give a little background for manipulating digital images in Matlab. Suppose that R is a 2-dimensional matrix representing a monochrome image that we wish to colorize. We first need to scale R and then convert it to red, green, blue (RGB) format. This can be done in Matlab using the pair of statements **R = R ./ max(max(R))** and **R = cat(3,R,R,R)**. Execution of these two statements converts R to a 3-dimensional structure where $R_{i,j,1}$, $R_{i,j,2}$ and $R_{i,j,3}$ respectively give numerical values for the red, green and blue components of pixel (i,j). It is helpful to scale R immediately before conversion to RGB so that changes to the color components done by the colorization schemes will have consistent, predictable impact. In our implementations we use the following associations in Table 1 between category (or label) number, color and RGB values:

Table 1: Category-Color Associations

Category Number	Name of Associated Color	[R,G,B] Values
1	red	[1,0,0]
2	green	[0,1,0]
3	blue	[0,0,1]
4	cyan	[1,1,0]
5	magenta	[1,0,1]
6	yellow	[0,1,1]
7 or more	black	[0,0,0]

This color scheme is implemented and referenced in the algorithms using the label color matrix C given by

$$C = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}^T \quad (5)$$

Sometimes we will refer to R_{ij} as the matrix or structure entry and sometimes as the corresponding pixel. Also, the notation "**:**" should be understand (as in Matlab) to indicate free range of that parameter. In other words, C(2,:) refers to row 2 of C. As an example of using C, the color for category 3 has [R,G,B] values C(3, :) = [0, 0, 1] which is blue. More colors can easily be included by adding rows to C. The choice of C in (5) suffices for us as we do not have examples with more than 7 different label categories.

We now describe the two colorization schemes. The first type of colorized image is said to be Diagonally Colorized (DC), which when paired with the two ordering schemes gives the two images DCiVAT and DCLR. Here each diagonal pixel, is assigned the label color of its corresponding object. For an example, let R** denote the result of scaling and then converting to RGB format either the iVAT matrix R'* or the LR version of it. If the third row (and column) of the matrix R** corresponds to an object that is in category 2, then the (3,3) pixel is made (pure) green by assigning the RGB values of C(3,:) = [0, 1, 0] to $R^{**}_{3,3}$. This is a simple approach that gives clear information if the number of data is very small, but a single diagonal pixel may be very difficult or impossible to see if the number of data is large. For large data sets, we include the option of coloring *B* additional bands of elements to the immediate right and underneath the diagonal pixels; in the previous example that would mean to not just colorize $R^{**}_{3,3}$ but to also change RGB color values of $R^{**}_{3,4}$, $R^{**}_{3,5}$, …, $R^{**}_{3,\min(n,3+B)}$ and $R^{**}_{4,3}$, $R^{**}_{5,3}$, …, $R^{**}_{\min(n,3+B),3}$ to all match $R^{**}_{3,3}$. This thickened diagonal band will be visible for images of any size, as long as the number of additional bands B is approximately equal to n/50 or larger. Examples in the next section will give the reader a feel for how many additional bands are needed. Algorithm 3 which gives the first pair of our colorized images, DCiVAT and DCLR, is given next.

Algorithm 3. Diagonal Colorization: DCiVAT and DCLR

Input: R—n × n dissimilarity matrix
L—1 × n array where L(j) = category label of o_j
T—string giving order type to use: iVAT or LR
B—scalar in {0,1,2,…,~n/25} denoting the number of additional color bands
C—c × 3 matrix giving label color RGB values

Step 1 Apply Algorithm 1 (VAT) to R to get R* and 1 × n permutation array P

Step 2 Obtain the reordered label array L* using the permutation array P by L*(j) = L*(P(j)) for $1 \le j \le n$.

Step 3 Apply Algorithm 2 (iVAT) to R* to get R'*

Step 4 If T = iVAT, then do not alter R'*
If T = LR, then overwrite R'* by its LR version and L* by its LR version

Step 5 Scale R'* to have all elements in [0,1] and convert the scaled matrix R'* to RGB color format structure R**

Step 6 For r = 1, …, n
Change RGB color values of diagonal pixel R^{**}_{rr} to C(L*(r), :)
next r

Step 7 If B > 0, Change RGB color values of B pixels to right and below to match the new color of the diagonal pixel.
For r = 1,…,n
Change RGB color values of $R^{**}_{r,r+1}$, $R^{**}_{r,r+2}$, …, $R^{**}_{r,\min(n,r+B)}$ and $R^{**}_{r+1,r}$, $R^{**}_{r+2,r}$, …, $R^{**}_{\min(n,r+B),r}$ to all match C(L*(r), :)
next r

The second type of colorized image is said to be Block Colorized (BC), which when paired with the two ordering schemes gives BCiVAT and BCLR. In the block-colorization approach we color a pixel R^{**}_{rt} whenever it is intracategory, i.e., the row and column objects corresponding to that pixel share the same category. This procedure will colorize many pixels compared to the diagonal-colorization approach and if care is not taken then cluster information depicted by various shading patterns could be destroyed by overly intense off-diagonal colorization. Rather than make complete changes to the effected pixels, we redefine them as the average of the old RGB values and the new category RGB ones; i.e., we change the RGB color values of intra-category pixels R^{**}_{rt} to equal the average of (old) R^{**}_{rt} and C(L*(r), :). This has an analogous effect to using stain and not paint on a piece of wood; we get the advantage of color but can still see the grain underneath the color. Algorithm 4 is given next which gives the last two of our colorized images, BCiVAT and BCLR.

Algorithm 4. Block Colorization: BCiVAT and BCLR

Input: R—n × n dissimilarity matrix
L—1 × n array where L(j) = category label of o_j
T—string giving order type to use: iVAT or LR
C—c × 3 matrix giving label color RGB values

Step 1 Apply Algorithm 1 (VAT) to R to get R* and 1 × n permutation array P

Step 2 Obtain the reordered label array L* using the permutation array P by L*(j) = L(P(j)) for $1 \le j \le n$.

Step 3 Apply Algorithm 2 (iVAT) to R* to get R'*

Step 4 If T = iVAT, then do not alter R'*
If T = LR, then overwrite R'* by its LR version and L* by its LR version

Step 5 Scale R'* to have all elements in [0,1] and convert the scaled matrix R'* to RGB color format structure R**

Step 6 Change RGB color values of intra-category pixels
For all (r,t) ∈ {1,…,n} × {1,…,n} with L*(r) = L*(t)
Change the RGB color values of R^{**}_{rt} to equal the average of (old) R^{**}_{rt} and C(L*(r), :)

In the next section we report on various experiments involving the four approaches. Which type of colorization is better, DC or BC? Which type of ordering is better, iVAT or LR? Does any combination of colorization and ordering produce a useful higher dimensional analogue to the 2-dimensional colorized scatterplot?

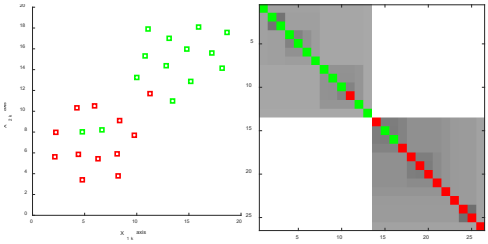
4. Experiments

In the first part of this section, we give all four new images for each in a set of seven 2-dimensional examples. The 2-dimensional examples allow us to compare the images directly to colorized scatterplots, and this will help us achieve two goals: (1) develop some skill in properly interpreting the new colorized images, and

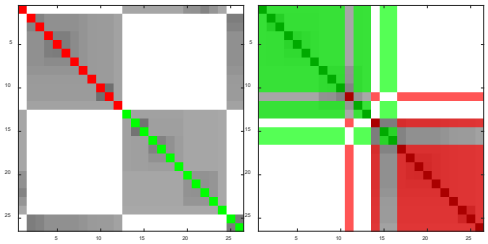
(2) determine if some of the new colorized approaches work better than others. Note that there is always color consistency between scatterplots and colorized images; i.e., scatterplot points of a certain color correspond to image pixels of the same color.

Example 1. Hypothetical School Data from Section 1.

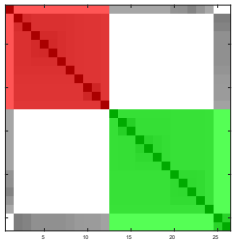
Fig. 5 repeats the colorized scatterplot for the School Data along with its four corresponding new images. Note that DCiVAT in Fig. 5(b) most clearly represents the situation of the single minority point in the green cluster and the two minority points in the red cluster. BCiVAT in 5(d) gives the same information but the colorized strips outside of the main diagonal blocks are a little distracting, at least initially. The DCLR and BCLR images in Fig. 5(c) and 5(e), respectively, disrupt the main diagonal block structure of the iVAT image by reordering according to category, but these images still convey the information of the scatterplot. Note how the off-diagonal-block patches of gray are essential for interpretation; for example, the vertical off-diagonal gray bars in Fig. 5(e) indicate that the last two green points are near (or in) the red cluster. No extra diagonal bands are needed (i.e., B = 0) for DCiVAT and DCLR because of the small sample size n = 26. With each example, notice which type of image most transparently and naturally conveys the essential cluster and category information in the scatterplot.



(a) Scatterplot (b) DCiVAT



(c) DCLR (d) BCiVAT

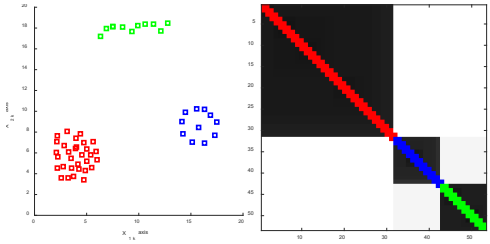


(e) BCLR

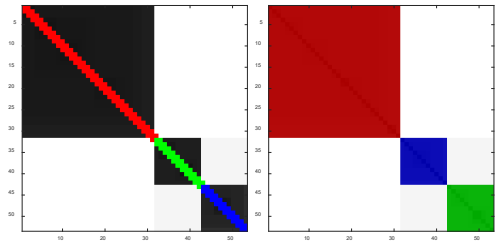
Figure 5: Colorized scatterplot and four images; sample size n = 26; extra diagonal bands B = 0 for DCiVAT and DCLR.

Example 2. Data with Perfect Alignment between Spatial Clusters and Label Categories.

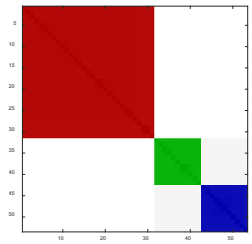
Fig. 6 gives the scatterplot and corresponding images. This example covers the ideal case where the clusters are well separated and perfectly aligned with the label categories. In this simple case each approach perfectly captures the essential information in the scatterplot. Each image shows the presence of 3 well-separated clusters, each filled with points of a single label category. Diagonal blocks, each of single color with light off-block areas, imply alignment between clusters and categories is total. Relative block sizes indicate the red cluster has about 3 times as many points as the blue and green clusters. The iVAT-produced minimax distances allow the stringy clusters to be handled just as naturally as the cloud clusters. The diagonals of DCiVAT and DCLR are slightly thickened for the sample size n = 53 using B = 1 extra bands of pixel coloration.



(a) Scatterplot (b) DCiVAT



(c) DCLR (d) BCiVAT



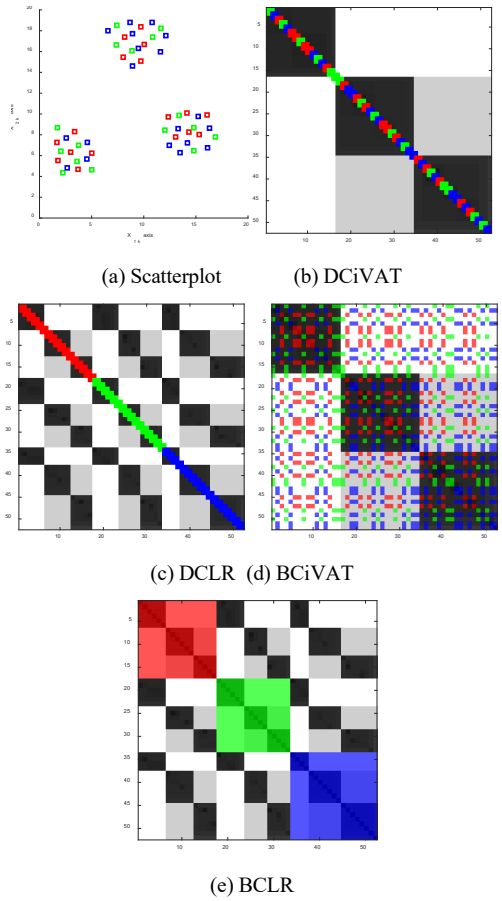
(e) BCLR

Figure 6: Colorized scatterplot and four images; sample size n = 53; extra diagonal bands B = 1 for DCiVAT and DCLR.

Example 3. Data with No Alignment between Spatial Clusters and Label Categories.

Fig. 7 gives the scatterplot and corresponding images. The scatterplot in Fig. 7(a) shows 3 well-separated spatial clusters having virtually no alignment between clusters and label categories. Once again, the DCiVAT image gives all the information in a clear fashion. It shows there are three spatial clusters and that each cluster has a mixture of points from each of the three label categories. The BCiVAT image is similarly informative but the colorful off-diagonal pixels are somewhat distracting. Note that the reordering of DCLR and BCLR has seriously disguised the spatial cluster information, and to reconstruct it you have to notice that the red objects appear in 3

small clusters, as do each of the green and blue objects. This could all be explained by having 3 mixed clusters as in the scatterplot but having to do this type of reconstruction is awkward at best.



(a) Scatterplot (b) DCiVAT

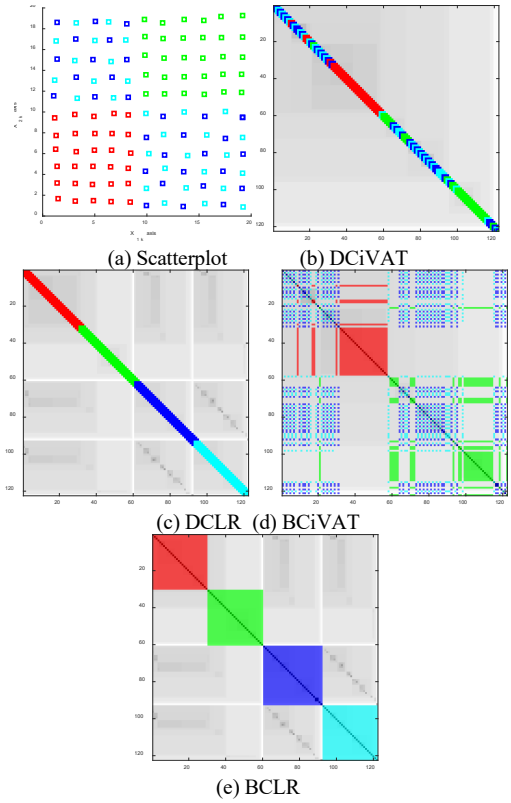
(c) DCLR (d) BCiVAT

(e) BCLR

Figure 7: Colorized scatterplot and four images; sample size n = 52; extra diagonal bands B = 1 for DCiVAT and DCLR.

Example 4. Data with No Spatial Clusters.

Fig. 8 gives the scatterplot and corresponding images. In this example we have a data set with virtually no cluster structure. In each of the images, a uniform gray tint and an absence of dark diagonal blocks indicate a lack of spatial clusters. The scatterplot contains additional information, though. There is some segregation of red and green points, while the blue and cyan ones are mixed together. Do any of the images in Fig. 8 represent this? The DCLR and BCLR do not give the color organization information in an easily interpreted form. No doubt the ghost diagonals in the bottom right-hand quarter of the DCLR and BCLR images contain some of this information but it is difficult to interpret. On the other hand, the iVAT ordering in DCiVAT and BCiVAT shows extensive intermingling of blue and cyan, along with some segregated groups of red and green. It is not a perfect representation of the scatterplot but it is consistent with what is shown there and it is instantly interpretable. Lastly, note the larger sample size n = 122 gives a need for B = 3 extra bands to widen the colorized diagonals in DCiVAT and DCLR.



(a) Scatterplot (b) DCiVAT

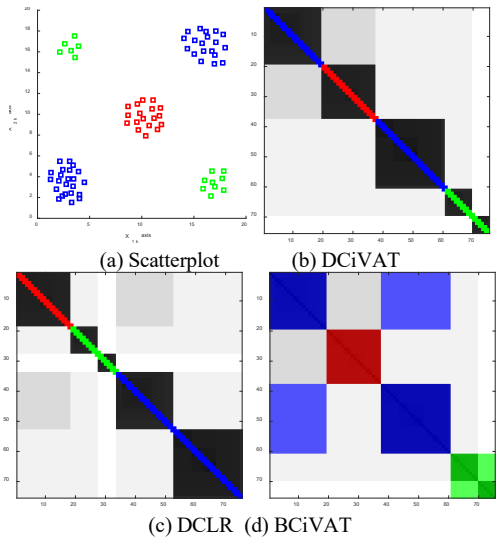
(c) DCLR (d) BCiVAT

(e) BCLR

Figure 8: Colorized scatterplot and four images; sample size n = 122; extra diagonal bands B = 3 for DCiVAT and DCLR.

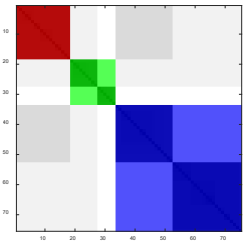
Example 5. Data with More Spatial Clusters than Label Categories.

Fig. 9 gives the scatterplot and corresponding images. This example is quite easy for all four approaches. The scatterplot shows 5 well-separated clusters: 1 red cluster, 2 blue clusters and 2 green clusters. This information is indicated most clearly by DCiVAT and DCLR. The BCiVAT and BCLR images are also consistent with the scatterplot and easily understood, but slightly more awkward because of the lighter blue and green off-diagonal blocks. Notice that the blue diagonal blocks in DCiVAT are not adjacent while the green ones are, and this shows that adjacency of blocks does not necessarily imply nearness of corresponding spatial clusters.



(a) Scatterplot (b) DCiVAT

(c) DCLR (d) BCiVAT

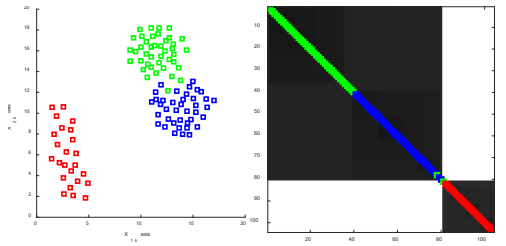


(e) BCLR

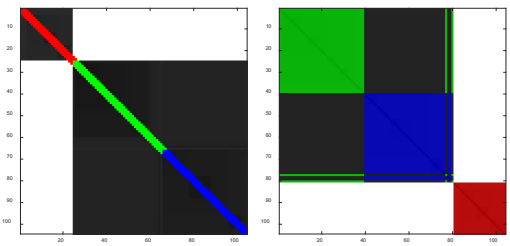
Figure 9: Colorized scatterplot and four images; sample size n = 75; extra diagonal bands B = 1 for DCiVAT and DCLR.

Example 6. Data with Touching Spatial Clusters #1.

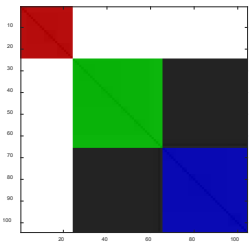
Fig. 10 gives the scatterplot and corresponding images. Several of the previous examples have dealt with extreme cases but we include one here with a more likely property: the existence of spatial clusters that are not well separated. The scatterplot in Fig. 10(a) shows a data set with a red, green and (mostly) blue cluster. The green and blue clusters are touching and could arguably be considered a single cluster. Notice what the images indicate for this example. At first glance, DCLR and BCLR seem to do a good job, indicating the presence of 3 clusters, each cluster populated by points of a single label category. The results for DCiVAT and BCiVAT are essentially the same but with the additional information that a couple of the green points may be invasive to the blue cluster. In each image the large dark block containing the blue and green areas indicate that those two clusters are very near each other and could arguably be considered a single cluster.



(a) Scatterplot (b) DCiVAT



(c) DCLR (d) BCiVAT

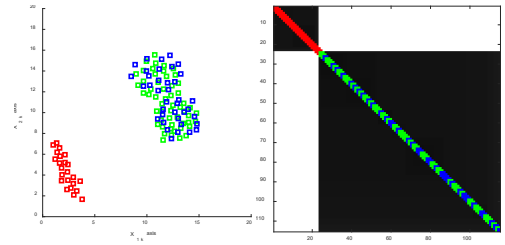


(e) BCLR

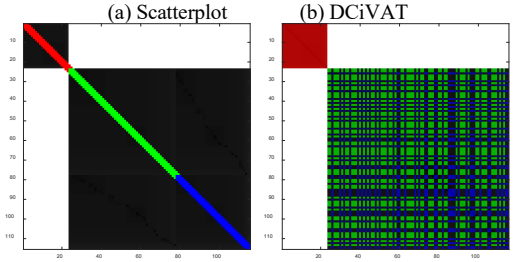
Figure 10: Colorized scatterplot and four images; sample size n = 104; extra diagonal bands B = 2 for DCiVAT and DCLR.

Example 7. Data with Touching Spatial Clusters #2.

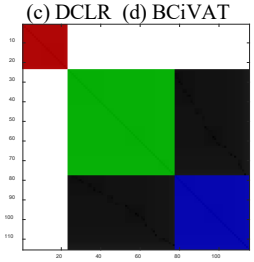
Fig. 11 gives the scatterplot and corresponding images. First note the spatial similarity of this and the previous example that is evident in a comparison of the Fig. 10(a) and Fig. 11(a) scatterplots. The real difference is all about the labeling of the points in the touching clusters. Notice how the DCiVAT and BCiVAT images in Fig. 11 perfectly represent the situation; there is a distant cluster of red points and a single cluster of points with commingled green and blue labels. On the other hand, DCLR and BCLR represent this and the data from Example 6 in an essentially identical manner, and we view this as a serious fault of the label reordering schemes.



(a) Scatterplot (b) DCiVAT



(c) DCLR (d) BCiVAT



(e) BCLR

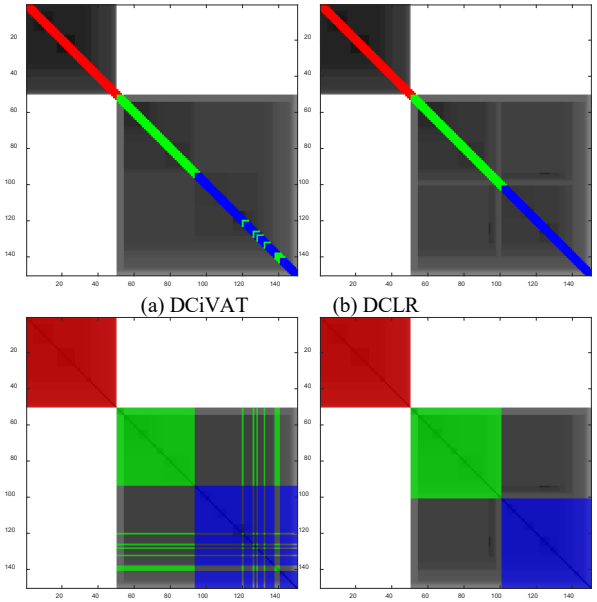
Figure 11: Colorized scatterplot and four images; sample size n = 115; extra diagonal bands B = 2 for DCiVAT and DCLR.

In the second group of examples we will apply all four approaches to higher-dimensional data sets taken from the UCI Machine Learning Repository at https://archive.ics.uci.edu/ml/index.php. The data sets used for the experiments below can be found by searching there for 'iris', 'divorce' and 'seeds'. For each data set we only give the four new images since no scatterplot is possible.

Example 8. Iris Data [10].

Fig. 12 gives the corresponding images. This data is considered to be real-valued continuous. Fisher's Iris Data [10] has probably been used more often in demonstrating clustering and classification schemes than any other data set. In this example we use the "bezdekIris.data" version of the data set from UCI. It consists of 150 4-variate object data describing the Setosa (red labels), Versicolor (green labels) and Virginica (blue labels) subspecies of iris. The data set contains 50 observations for each

of the 3 subspecies, with each observation giving a measurement of sepal length, sepal width, petal length and petal width. Various colorized scatterplots of 2-dimensional slices of the Iris Data are available in the literature (e.g., page 82 of [1]) and the images in Fig. 12 are consistent with what is known about it. The Setosa (red) cluster is well separated from the other two overlapping clusters. Note that the DCiVAT and BCiVAT images indicate the possible intrusion of the Virginica cluster by Versicolor data.



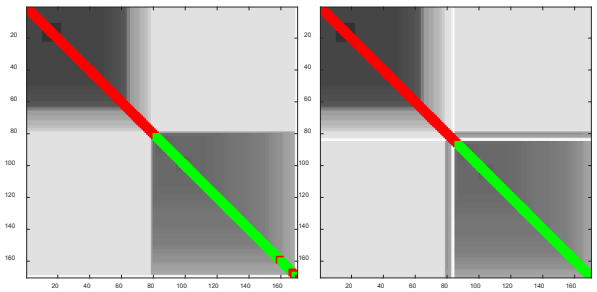
(a) DCiVAT (b) DCLR

(c) BCiVAT (d) BCLR

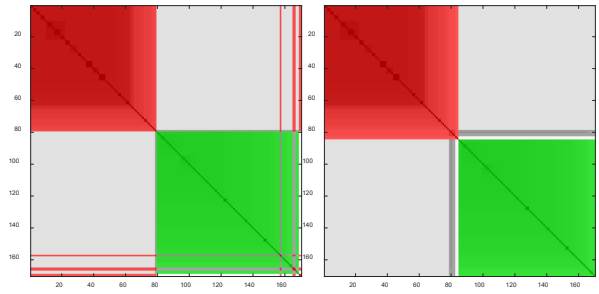
Figure 12: Four images for 4-variate Iris Data; sample size n = 150; extra diagonal bands B = 3 for DCiVAT and DCLR

Example 9. Divorce Data [11].

Fig. 13 gives the corresponding images. This data is considered to be integer valued and is from [11]. It consists of 170 54-variate data from 84 divorced (red label) and 86 married (green label) couples. Each person interviewed gave answers of the form 0 (never), 1 (seldom), 2 (average), 3 (frequently), 4 (always) to each of 54 questions such as "I feel right in our discussions" and "I wouldn't hesitate to tell her about my wife's inadequacy". All images indicate fairly separated clusters that are well aligned with the categories. There may be a bit of a trail of red transition points leading from the core of the red cluster over to the green cluster. The DCiVAT and BCiVAT images also indicate possible red intrusion into the green cluster. Appreciate that we have done something analogous to a colorized scatterplot even though we have 54-dimensional data!



(a) DCiVAT (b) DCLR

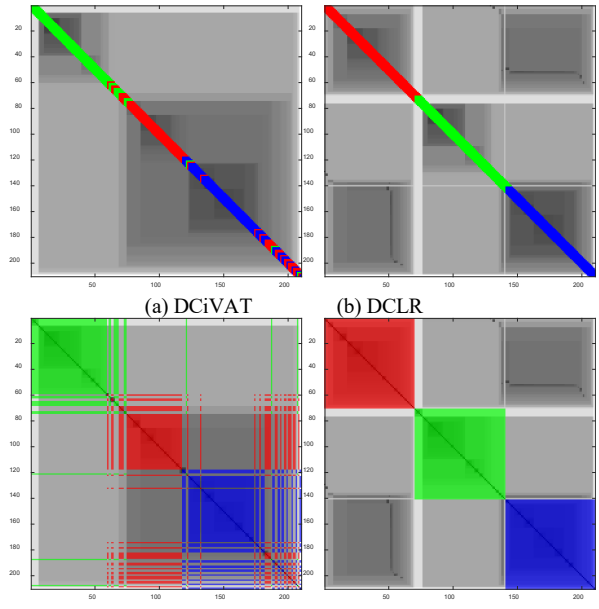


(c) BCiVAT (d) BCLR

Figure 13: Four images for Divorce Data; sample size n = 170; extra diagonal bands B = 4 for DCiVAT and DCLR.

Example 10. Seeds Data [12].

Fig. 14 gives the corresponding images for our final example. The real-valued, continuous Seeds Data is from [12] and consists of 210 7-variate observations of 3 varieties of wheat: Kama (red label), Rosa (green label), and Canadian (blue label). The data is divided equally among the 3 categories. Each datum consists of measurements of 7 geometric parameters of wheat kernels: area, perimeter, compactness, length, width, asymmetry coefficient and length of kernel groove. All of these measurements are obtained using a high-quality imaging of the kernel done by a soft X-ray technique. The DCiVAT and BCiVAT images indicate the presence of poorly separated clusters. There appears to be a green labeled cluster and a large red-blue one. The red-blue one has a red core part, a blue core part, and a mixed part. The DCLR and BCLR images may at first seem to indicate 3 cleanly aligned spatial clusters but the dark off-diagonal block is a reminder that there is considerable red and blue intermingling. This example is probably typical of many; it provides an imperfect but useful glimpse of some of the information that we seek.



(a) DCiVAT (b) DCLR

(c) BCiVAT (d) BCLR

Figure 14: Four images for Seeds Data; sample size n = 210; extra diagonal bands B = 5 for DCiVAT and DCLR.

5. Discussion

In this paper we presented four colorized images that simultaneously display cluster and category information for

labeled object data sets of any dimension. These four approaches are obtained by combining a colorizing scheme (Diagonalized or Block) with an ordering scheme (iVAT or LR). Having tested all approaches on the diverse set of examples in Section 4, we can now assess DCiVAT, DCLR, BCiVAT and BCLR. We believe that the superior method among the four is DCiVAT. In every 2-dimensional experiment it produced an accurate reflection of the corresponding scatterplot in a way that is most easily understood. We will say more about DCiVAT after we give specific criticisms of the other three approaches.

The label reordered approaches DCLR and BCLR may only give good results in the simplest possible case, when there are well-separated clusters with (nearly) perfect alignment between clusters and label categories. This situation is demonstrated in Example 2 where all colorization schemes did equally well. We found no situation where LR ordering works better than iVAT ordering and some where it absolutely fails, such as that seen by comparing Examples 6 and 7; in those examples DCLR and BCLR gave nearly identical results for two very different situations. In other examples LR ordering preserves, but scrambles, the cluster information so that visual interpretation of the images is [complicated; Example 3 gives a clear case of such cluster obfuscation. To summarize, LR is never better, sometimes flat wrong and at other times unnecessarily awkward. Additionally, the process of first doing iVAT and then undoing that ordering to get LR adds computational overhead above that of DCiVAT and BCiVAT.

In comparing DCiVAT to BCiVAT we know the ordering is the same and so it is only about the relative merits of the two different colorization schemes. Preference between these is admittedly subjective. There are cases where a BCiVAT image is very clear in its depiction of the data set, as in Example 7, and other cases where there is enough off-diagonal action to be distracting, as in Examples 3 and 5. It is possible that adjustments to BCiVAT that lighten the off-diagonal part of the image might very well improve its readability, but will it then produce something clearer than DCiVAT? We do not think so. Our conclusion is that BCiVAT is good, but DCiVAT is also good and never produces visually awkward results. DCiVAT is also the most straightforward of the four schemes to describe or program. Simply do iVAT and then colorize the diagonal pixel to match the category labels. Adding extra bands of colorization is easily done in the routine by adding one double loop. While our Matlab routines were written to allow the user to specify the number of additional colorization bands B, this choice could be automated in the DCiVAT program by using something like **B = floor(n/25)**. Python users can get the core VAT and iVAT routines written by Ismaïl Lachheb at https://pypi.org/project/pyclustertend/ .

This paper is concluded with a few ideas about possible future research involving DCiVAT or related colorization schemes. The first opportunity may be to colorize other VAT-type approaches. Many VAT type procedures have been devised since the original paper [6], and [8] gives a thorough survey of most of them with a total of 184 references. For example, there are various big data variants of VAT beginning with [13] and continuing through more recent work as in [14]. There are VAT-based procedures for rectangular (i.e., not square) relational data matrices [15,16] and even streaming data [17]. Are there opportunities in this vast set of methodology to add useful colorization, for depicting either label information or something else? Color is effectively used in [18] to draw attention to parts of 3-dimensional scatterplots. Are there more opportunities to usefully colorize? Might a block colorization scheme be more natural than diagonal colorization in some cases, such as for rectangular data matrices?

Another opportunity may be to refine our ability to interpret iVAT and colorized iVAT images. Did the examples in Section 4 teach us everything about DCiVAT interpretation or would a more extensive and systematic study yield more? Is it possible to use a colorized iVAT image to visually display the performance of a classifier, where correctly classified objects can be visually differentiated from incorrectly classified ones? Is it possible to tweak BCiVAT in a way that lowers its off-diagonal level of distraction? We should think more creatively about adding colorization to further increase the usefulness of VAT approaches.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] J.C. Bezdek, Elementary Cluster Analysis: Four Basic Methods that (Usually) Work. Gistrup, Denmark: River Publishers, 2022.

[2] E.D. Hathaway, R. Hathaway, Diagonally colorized iVAT images for labeled data, submitted to the IEEE International Conference on Data Mining, Orlando, FL, USA, (2022), doi: 10.1109/ICDMW58026.2022.00043.

[3] L. Wang, T. Nguyen, J. Bezdek, C. Leckie, K. Ramamohanarao, iVAT and aVAT: enhanced visual analysis for cluster tendency assessment, in: M.J. Zaki, J.X. Yu, B. Ravindran, V. Pudi (Eds.), Advances in Knowledge Discovery and Data Mining. PAKDD 2010. Lecture Notes in Computer Science, 6118. Springer, Berlin-Heidelberg, 16-27, 2010. doi: 10.1007/978-3-642-13657-3_5.

[4] T.C. Havens, J.C. Bezdek, An efficient formulation of the improved visual assessment of cluster tendency (iVAT) algorithm, IEEE TKDE. **23**, 568-584, 2011. doi: 10.1109/TKDE.2011.33.

[5] I. Borg, P.J.F. Groenen, Modern Multidimensional Scaling: Theory and Applications, 2nd Edition. Springer, New York, 200).

[6] J.C. Bezdek, R.J. Hathaway, VAT: A tool for visual assessment of (cluster) tendency, Proc. IJCNN, IEEE Press. (2002) 2225-2230, doi: 10.1109/IJCNN.2002.1007487.

[7] R.C. Prim, Shortest connection networks and some generalizations, Bell System Technical Journal. **36**(1959) 1389-1401, doi: 10.1002/j.1538-7305.1957.tb01515.x.

[8] D. Kumar, J.C. Bezdek, Visual approaches for exploratory data anlysis: A survey of the visual assessment of clustering tendency (VAT) family of algorithms, IEEE SMC Magazine. **6**, 10-48, 2020. doi: 10.1109/MSMC.2019.2961163.

[9] L. Wilkinson, M. Friendly, Michael, The history of the cluster heat map, The American Statistician. **63,** 179-184, 2009. doi: https://doi.org/10.1198/tas.2009.0033.

[10] R.A. Fisher, The use of multiple measurements in taxonomic problems, Annals of Genetics. **7**(1935) 179-188, doi: https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.

[11] M. Yöntem, K. Adem, T. İlhan, S. Kılıçarslan, Divorce prediction using correlation-based feature selection and artificial neural networks, Nevşehir Hacı Bektaş Veli University SBE Dergisi, **9**(1), 259-273, 2019. Retrieved from https://dergipark.org.tr/en/pub/nevsosbilen/issue/46568/549416.

[12] M. Charytanowicz, J. Niewczas, P. Kulczycki, P.A. Kowalski, S. Lukasik, S. Zak, A complete gradient clustering algorithm for features analysis of X-ray images, Information Technologies in Biomedicine, Ewa Pietka, Jacek Kawa (eds.), Springer-Verlag, Berlin-Heidelberg, 15-24, 2011. doi: https://doi.org/10.1007/978-3-642-13105-9_2.

[13] J.M. Huband, J.C. Bezdek, R.J. Hathaway, bigVAT: Visual assessment of cluster tendency for large data sets, Pattern Recognition. **38**, 1875-1886, 2005. doi: https://doi.org/10.1016/j.patcog.2005.03.018.

[14] L.H. Trang, P. Van Ngoan, N. Van Duc, A sample-based algorithm for visual assessment of cluster tendency (VAT) with large datasets, Future Data and Security Engineering (Lecture Notes in Computer Science, 11251) J. Kung, R. Wagner, N. Thoai, M. Takizawa (eds.), Springer-Verlag, Cham,

Switzerland, 145-157, 2018. doi: https://doi.org/10.1007/978-3-030-03192-3_11.
[15] J.C. Bezdek, R.J. Hathaway, J.M. Huband, Visual assessment of clustering tendency for rectangular dissimilarity matrices, IEEE Trans. Fuzzy Syst. 15, 890-903, 2008. doi: 10.1109/TFUZZ.2006.889956.
[16] T.C. Havens, J.C. Bezdek, A new formulation of the coVAT algorithm for visual assessment of clustering tendency in rectangular data, Int. J. of Intelligent Systems. **27**, 590-612, 2012. doi: https://doi.org/10.1002/int.21539.
[17] D. Kumar, J.C. Bezdek, S. Rajasegarar, M. Palaniswami, Adaptive cluster tendency visualization and anomaly detection for streaming data, ACM Trans. On Knowledge Discovery from Data. **11,** 1-40, 2016. doi: https://doi.org/10.1145/2997656.
[18] I.J. Sledge, J.M. Keller, Growing neural gas for temporal clustering, Proc. Int. Conf. Pattern Recognition, 1-4, 2008. doi: 10.1109/ICPR.2008.4761768.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 9, No. 1, 114-123 (2024)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Smart Agent-Based Direct Load Control of Air Conditioner Populations in Demand Side Management

Pegah Yazdkhasti*, Julian Luciano Cárdenas–Barrera, Chris Diduch

Department of Electrical & Computer Engineering, Smart Grid Research Center, University of New Brunswick, Fredericton, Canada

ARTICLE INFO

Article history:
Received: 10 November, 2023
Revised: 12 January, 2024
Accepted: 12 January, 2024
Online: 06 February, 2024

Keywords:
Demand-Side Management
Smart Grid
System Identification
Direct Load Control
Load Forecast
Rebound Effect

ABSTRACT

The integration of fluctuating renewable resources such as wind and solar into existing power systems poses challenges to grid reliability and the seamless incorporation of these resources. To address the inherent variability in renewable generation, direct load control emerges as a promising method for demand-side management. Thermostatically controlled appliances, like air conditioners, hold a significant role in this approach. However, effective direct load control necessitates accurate load magnitude estimation and the potential for load shifting. In this paper, we introduce a smart-agent architecture that employs a mathematical model to forecast aggregated power consumption behavior, even when changes are introduced by the controller. To assess system performance, a numerical simulator was developed, demonstrating the system's adaptability to changes, its self-retraining capability, and its continuous improvement in predicting aggregated power consumption.

1. Introduction

This paper builds upon the research initially introduced at ICCAD-2023 [1], where the authors presented a smart agent to make a mathematical model for a population of air conditioners and forecast their power consumption and control it. This paper expands the mathematical formulation for modeling and provides justification for the linear approximation. Additionally, it improves the forecasting model by introducing a new formulation for reserve capacity to reshape the aggregated power demands.

The utilization of clean resources, such as wind and solar power, for electricity generation has experienced substantial growth in the past decade [2]. However, the integration of renewable resources poses challenges to existing power grids, primarily due to the rapid fluctuations in generation, diminishing the reliability and quality of electric power systems [3]. While energy storage devices like batteries can mitigate the swift fluctuations of wind and solar power [4], the cost of deploying renewable energy increases with the use of energy storage. Direct Load Control (DLC) emerges as a cost-effective demand-side management (DSM) method to address the intermittency of generation [5]. Recent studies demonstrate that adjusting set points or operation times of thermostatically controlled loads (TCL), such as electric water heaters or air conditioners, effectively controls aggregated power consumption [6, 7, 8, 9].

The system operator plays a crucial role in balancing generation and demand sides, and an accurate forecast of the power consumption is essential for this task. In a DLC program, the system operator needs to understand the controller's capacity to increase or decrease the load, providing feasible instructions.. This research focuses on developing a methodology for fast and accurate load controllability forecasting.

In recent years, significant efforts have been dedicated to developing mathematical models and control strategies aimed at effectively managing the aggregate demand of air conditioners. One approach involves designing a control strategy to adjust the thermostat set point, thereby reducing peak demand and alleviating the impact of renewable energy variability [10, 11]. Alternatively, centralized or distributed/decentralized control strategies have been explored, focusing on directly managing the on/off states of AC compressors [11, 12, 13, 14, 15]. However, these methods fall short in providing practical means to estimate essential model parameters like thermal capacitance/resistance, or to dynamically adapt the model to changing conditions. Another challenge lies in the absence of information regarding the capacity to increase or decrease the total load within a future time window.

In response to these challenges, several studies have shifted focus towards forecasting the controllable load. Some utilize neural networks to predict air conditioning power consumption

*Corresponding Author: Department of Electrical & Computer Engineering, Smart Grid Research Center, University of New Brunswick, Fredericton, Canada, Email: pegah.ykh@unb.ca

based on temperature, humidity, and historical power consumption [16, 17, 18]. Others employ regression-based algorithms coupled with Kalman filters to forecast air-conditioning load using historical data [19]. Least square support vector machines have also been proposed for short-term load forecasts [20]. However, these methods may face impracticalities when integrated with a control system that alters load behavior. Furthermore, none of these methods offer insights into how much the load can be changed due to a control action or formulate the control capacity.

Introducing a dynamic process and a state space model, our proposed methodology provides an efficient and accurate forecast of the controllable load for the system operator. By constructing a model adaptable to changes in thermal parameters and load behavior, our method surmounts limitations associated with traditional approaches. Furthermore, it equips the system operator with the ability to estimate the controllable capacity of the load, a critical factor in load balancing tasks.

The effectiveness of our methodology was rigorously evaluated through simulations, comparing its performance with existing methods. The results demonstrate that our proposed method surpasses existing approaches in terms of both accuracy and speed. This underscores its potential to offer a more efficient and reliable solution for managing the demand side in response to the integration of renewable resources.

In conclusion, this paper introduces a novel methodology for forecasting the controllable load of air conditioners, capable of adapting to changes in thermal parameters and load behavior. The proposed methodology not only delivers an accurate and efficient load forecast but also empowers the system operator to estimate the controllable capacity of the load. With promising results in simulations, our method holds the potential for practical implementation, promising improved demand side management in response to the integration of renewable resources.

The remainder of this paper is structured as follows: Section 2 provides an overview of the system architecture, while Section 3 offers a concise introduction to the mathematical model of the system. The learning algorithm is detailed in Section 4, and Section 5 outlines the process of generating forecasts. The capacity to control the load is formulated in Section 6. The paper concludes with the presentation of simulation results in Section 7.

2. System Architecture

The objective of this study is to devise a smart agent for the efficient management of power consumption in thermostatically controlled appliances. This system operates within a hierarchical structure, collaborating with a system operator or a virtual power plant (VPP). Its role is to receive a target aggregated power consumption, interact with a load aggregator, and send control signals to adjust the load. The goal is to minimize the discrepancy between the measured and desired power consumption while maintaining customers' comfort. Figure 1 provides an overview of the system, which encompasses a load aggregator with three distinct modules:

1. Learning Model Module: This module is tasked with estimating a mathematical model that articulates the power consumption of the aggregated load based on input parameters.
2. Forecasting Module: Leveraging the estimated mathematical model and input forecasts, this module generates predictions for the baseline load and reserve capacity. Input forecasts encompass variables like weather conditions, occupancy patterns, and other factors influencing power consumption. The baseline load forecast represents the expected power consumption without any control action, while the reserve capacity forecast indicates the potential power reduction or increase through controller actions. These forecasts guide the controller in determining appropriate control signals to align the load with the desired power consumption.
3. Responsible for adhering to the desired aggregated load set by the system operator, this module controls the load. Further details on a simplified implementation of such a controller are presented in [21].

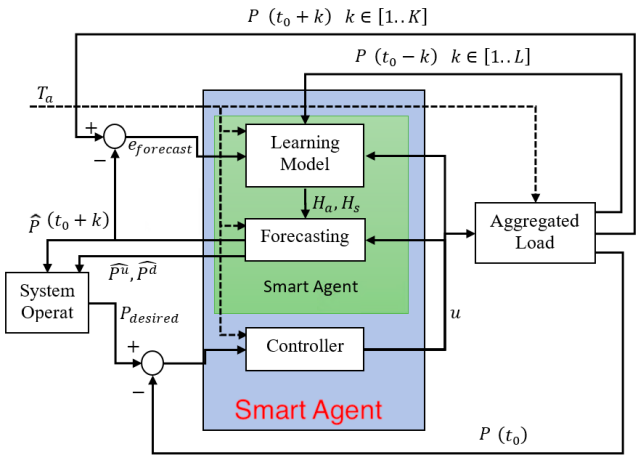


Figure 1: A block diagram for direct load control system [1]

In Figure 1, the power consumption of the load population (Aggregated Load) is input into the load aggregator at the current time denoted by t_0. Historical data from prior measurements ($P(t_0-k)$ for times preceding t_0) is utilized by the 'Learning Model' to establish a mathematical model for the system (H_a and H_s). Subsequently, the 'Forecasting' module utilizes this model to generate forecasts for the baseline load and control capacity. In the future, upon obtaining actual measurements of the aggregated load, the forecast error can be calculated as the difference between forecasted and actual power ($e_{forecast}$). This error serves as a trigger mechanism, initiating the training process.

3. Modeling the Aggregated Load

The controllable load in this research are a population of air conditioners. A Mathematical model can provide a powerful tool to analyze the behavior of such system. The following sections will present a commonly used model for single AC and a population of ACs.

3.1. Individual Load

In this context, we denote $m(k)$ as the on/off state of the air conditioner at time k, and T as the room temperature. The dynamics of room temperature changes are governed by Equation (1), a widely adopted formulation in the literature [22, 23, 12, 21]:

$$\dot{T}(t) = -\frac{1}{CR}(T(t) - T_a(t) + m(t)RP_{nom}\eta) \tag{1}$$

Here, R denotes the thermal resistance of the room, C is the thermal capacitance of the room, T_a represents the ambient temperature, P_{nom} corresponds to the nominal power rating of the air conditioner, and η stands for the coefficient of performance. The air conditioner switches on ($m(t) = 1$) when the room temperature surpasses the higher threshold setpoint, aiming to reduce the room temperature until it reaches the lower threshold, at which point the air conditioner turns off ($m(t) = 0$).

3.2. Aggregated Load

The power consumption of an air conditioner (AC) can be conceptualized as a system with two inputs: the thermostat setpoint (as the controllable input) and ambient temperature (as the disturbance signal). Other parameters, such as thermal resistance and capacitance, may be considered as external hidden parameters of the system.

Simulation results demonstrate that the aggregated power demand reaches a steady-state value when the inputs remain constant, as depicted in Figure 2. In general, the steady-state response of the aggregated power demand to the inputs is nonlinear. However, this surface can be approximated by a plane described by:

$$Pss(T_a, T_s) = Pss0 + \alpha T_a + \beta T_s, \quad T_a < 34^\circ C, T_s \leq T_a \tag{2}$$

Figure 2 illustrates the simulated steady-state aggregated power demand vs. the value obtained from (3). As expected, around the nonlinear zone of high ambient temperature, the plane exhibits larger differences than the simulated power demand. However, for the middle sections of the surface, it provides a good approximation for the steady-state aggregated power demand, with the error range over the middle part of the surface being less than 5%.

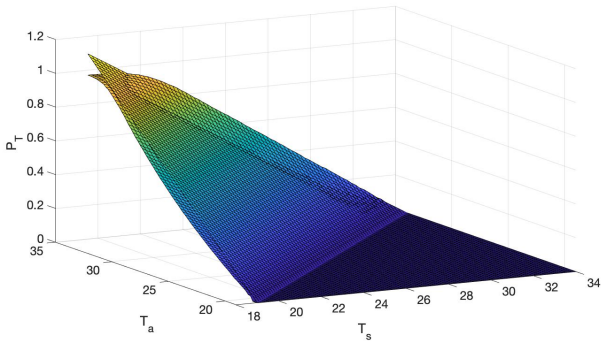


Figure 2: Comparison of the Approximated Plane Surface given by (2) with Simulated Aggregated Power Demand

Through numerical calculations, the coefficients for the best-fitting plane have been determined as:

$$P_{ss}(T_a, T_s) = 0.075T_a - 0.075T_s \tag{3}$$

With this observation, it becomes plausible to consider modeling this system as a Linear Time-Invariant (LTI) system, with the equilibrium point representing the DC gain of the model. Assuming the system is given by:

$$\delta P(s) = H_a(s)\delta T_a(s) + H_s(s)\delta T_s(s) \tag{4}$$

Here, δP denotes changes in the aggregated demand from the equilibrium point due to variations in ambient temperature and thermostat setpoints ($\delta T_a, \delta T_s$). H_a and H_s represent the LTI transfer functions for T_a and T_s respectively. The steady-state output is then given by:

$$\delta P_{ss} = H_a(0)\delta T_a + H_s(0)\delta T_s \tag{5}$$

This equation mirrors (2), where $H_a(0) = \alpha$ and $H_s(0) = \beta$.

Now, by approximating the aggregated demand with a LTI system, the dynamic response of the system can be captured by[24]:

$$P(k) = y_a(k) + y_s(k) + P_0 \tag{6}$$

Here, P_0 denotes the aggregated demand at the equilibrium point (or P_{ss}), while y_a and y_s represent the change in aggregated power consumption with respect to the equilibrium point, caused by variations in the ambient temperature and thermostat set point, respectively. These LTI system can be represented in state space model as:

$$H_a : \begin{cases} x_a(k+1) = A_a x_a(k) + B_a u_a(k), \\ y_a(k) = C_a x_a(k) \end{cases} \tag{7}$$

and

$$H_s : \begin{cases} x_s(k+1) = A_s x_s(k) + B_s u_s(k), \\ y_s(k) = C_s x_s(k) \end{cases} \tag{8}$$

where the matrices A_a, B_a, C_a, A_s, B_s and C_s should be identified through the learning process.

4. Learning Model

The purpose of this module is to formulate an approximation model for the population of air conditioners, eliminating the need for knowledge about any of the physical parameters introduced in Section 3.

The details of the model learning process is presented in [24]. As a short summary, the learning procedure initiates when the forecast error surpasses a predefined threshold level. At this juncture, T_a and T_s are determined, and the controller is temporarily disabled to prevent any artificial fluctuation in the aggregated power consumption.

The learning process encompasses two steps:

1. Determining H_a while holding T_s constant. In this scenario, changes in the output power are solely governed by variations in T_a.

2. Determining H_s when a small change is applied to T_s, and the effect of changes in T_a can be estimated and removed using H_a. The formulations for these steps are elaborated in detail in [24].

5. Load Forecast

To generate a load forecast, one can utilize the forecasted values of the model inputs in equations (7)-(8) to obtain a forecast for the output. However, this requires measurements of the states, which are not physical parameters that can be directly measured. As a result, a state observer is required to estimate these states.

5.1. State Estimation

The overall architecture of the state observer for this system is shown in Figure 3. In the figure, the changes in the output power are denoted by y, while the estimated impact of u_s and u_a are denoted by $\hat{y}_s$ and $\hat{y}_a$, respectively. The difference between the estimated changes and the actual change in the aggregated power is then fed to the state model through observer gains H_s and H_a to improve state estimation.

$$\begin{cases} \hat{x}_a(k+1) = A_a\hat{x}_a(k) + B_a u_a(k) + H_a C_a(x_a(k) - \hat{x}_a(k)) \\ \hat{y}_a(k) = C_a\hat{x}_a(k) \end{cases} \tag{9}$$

and similarly for the x_s state:

$$\begin{cases} \hat{x}_s(k+1) = A_s\hat{x}_s(k) + B_s u_s(k) + H_s C_s(x_s(k) - \hat{x}_s(k)) \\ \hat{y}_s(k) = C_s\hat{x}_s(k) \end{cases} \tag{10}$$

where $\hat{x}_a$ and $\hat{x}_s$ are the estimated states.

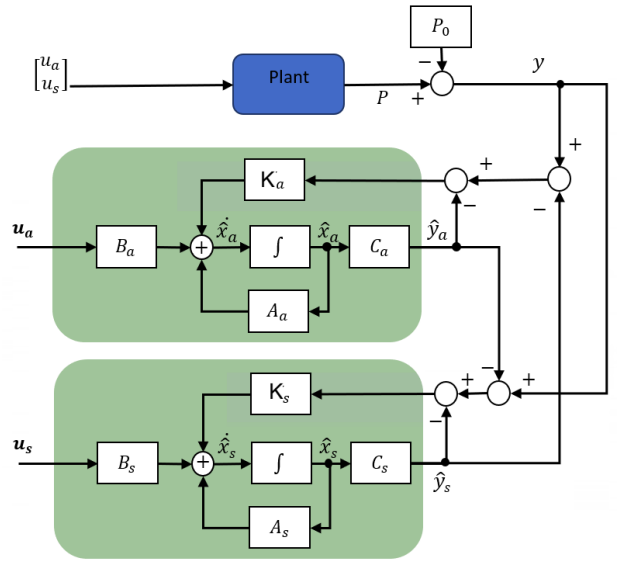


Figure 3: The state observers for each input [1].

These equations can be rewritten in a matrix format as:

$$\begin{cases} \hat{x}(k+1) & = \mathcal{A}\hat{x}(k) + \mathcal{B}\begin{bmatrix} u_a \\ u_s \end{bmatrix} + \\ & \quad \begin{bmatrix} K_a \\ K_s \end{bmatrix}(y(k) - C\hat{x}(k)) \\ \hat{y}(k) & = C\hat{x}(k) \end{cases}$$

where

$$\mathcal{A} = \begin{bmatrix} A_a & 0 \\ 0 & A_s \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} B_a & 0 \\ 0 & B_s \end{bmatrix} \tag{11}$$

and

$$C = \begin{bmatrix} C_a & C_s \end{bmatrix} \tag{12}$$

By considering the state error: $\tilde{x}(k) = x(k) - \hat{x}(k)$, we will have:

$$\tilde{x}(k+1) = (\mathcal{A} - KC)\tilde{x}(k) \tag{13}$$

And K can easily be determined so that the dynamics of the state error would be faster than the y system through functions such as pole placement technique in MATLAB.

5.2. Forecasting Baseline Load

When forecasting the baseline load, we operate under the assumption that the controller does not influence the load. Let's consider a forecasting horizon of L samples, assuming that the forecast is generated at time index k_0. Within this interval, we have:

$$\hat{u}_a(k) = \hat{T}_a(k) - \hat{T}_{a0} \qquad k \in [k_0 + 1, k_0 + L] \tag{14}$$

$$\hat{u}_s(k) = 0 \qquad k \in [k_0 + 1, k_0 + L] \tag{15}$$

Here, $\hat{T}_a(k)$ is the forecasted ambient temperature at time k. Then the forecast of the baseline aggregated power will be:

$$\hat{P}_b(k) = P_0 + \hat{y}_a(k) \qquad k \in [k_0 + 1, k_0 + L] \tag{16}$$

where P_0 is the power consumption at equilibrium point and

$$\begin{cases} \hat{x}_a(k+1) = A_a\hat{x}_a(k) + B_a\hat{u}_a(k) \\ \hat{y}_a(k) = C_a\hat{x}_a(k) \end{cases} \tag{17}$$

where the initial state $\hat{x}_a(k_0)$ can be obtained from the state observer that was presented earlier.

Equation (16) operates under the assumption that no previous control actions have occurred, implying no fluctuations in the aggregated power resulting from prior thermostat changes. Nevertheless, past control actions may continue to influence future power consumption. Given that the forecasting module receives information about each control signal to the thermostat setpoints, it can consistently revise the state space variables of the system. Consequently, the forthcoming changes in $\hat{y}_s$ can be determined by:

$$\begin{cases} \hat{x}_s(k+1) = A_s\hat{x}_s(k) \\ \hat{y}_s(k) = C_s\hat{x}_s(k) \end{cases} \tag{18}$$

and the baseline load will be:

$$\hat{P}_b(k) = \hat{P}_0 + \hat{y}_a(k) + \hat{y}_s(k) \qquad k \in [k_0 + 1, k_0 + L] \tag{19}$$

The payback effect, also known as rebound, is a well-established phenomenon in DSM. It characterizes the transient surge in aggregated power that occurs once control over the aggregated load is relinquished, allowing the power to revert to its baseline. This transient, initiated when the system shifts from controlled to uncontrolled operation, has the potential to give rise to unexpected secondary demand peaks.

The controller, tasked with executing demand-side management measures, holds the key to influencing the rebound effect by sustaining control over the devices even beyond the system operator's requested interval. However, this strategy may not always be practical or desirable, particularly if associated with high costs or energy consumption. Hence, in this section, we assume that the controller will exclusively operate during the specified interval, refraining from additional adjustments thereafter.

To equip the system operator with valuable insights into the repercussions of control actions, we introduce a method for estimating the payback of any desired demand even before initiating control. This approach aids the system operator in evaluating the acceptability of control actions and preemptively avoiding an escalation in future peak demands. By furnishing a dependable estimate of the payback effect, this method empowers the system operator to optimize the control strategy and mitigate potential risks to power system operation.

5.3. Approximating Control Signal

Let us consider a scenario where the system operator or VPP issues a dispatch, denoted by $\Delta P_d(k)$, for a control interval of C samples. In this scenario, the desired load can be considered as a change to the baseline load, which is given by:

$$\Delta P_d(k) = P_d(k) - \hat{P}_b(k) \tag{20}$$

Here, $\Delta P_d(k)$ represents the amount of power that needs to be added to or reduced from the baseline load. To effect this change, the thermostat set points, $\hat{u}_s$, need to be adjusted so that they create appropriate $\hat{y}_s$ values.

Now, the controller needs to generate a $\hat{u}_s(k)$ that results in the output $\hat{y}_s(k)$, matching $\Delta P_d(k)$. This assumes that the linear model remains valid when y_s changes by ΔP_d.

$$\begin{cases} \hat{x}_s(k+1) = A_s\hat{x}_s(k) + B_s\hat{u}_s(k) \\ \hat{y}_s(k) = \Delta P_d(k) = C_s\hat{x}_s(k) \end{cases} \tag{21}$$

This equation can be rewritten as:

$$\Delta P_d(k+1) = C_s[A_s\hat{x}_s(k) + B_s\hat{u}_s(k)] \tag{22}$$

And the control signal. $\hat{u}_s(k)$ will be:

$$C_sB_s\hat{u}_s(k) = \Delta P_d(k+1) - C_sA_s\hat{x}_s(k) \tag{23}$$

If the order of the system model is n, then the matrices C_s and B_s will have dimensions $1 \times n$ and $n \times 1$, respectively. The product C_sB_s will then result in a scalar value, and:

$$\hat{u}_s(k) = (C_sB_s)^{-1}(\Delta P_d(k+1) - C_sA_s\hat{x}_s(k)) \tag{24}$$

subject to:

$$-\alpha \le u_s(k) \le \alpha \tag{25}$$

for customers' comfort (will be described more in detail in Section 6). Then, the thermostat set points will be:

$$\hat{T}_s(k) = T_{s0} + \hat{u}_s(k) \tag{26}$$

for k in the control interval. It should be noted that initial state variable, $\hat{x}_s(k_0)$, is estimated using the state observer described in Section 5.1.

Figure 4 illustrates an example of an arbitrary desired aggregated load, represented by the top curve. The baseline load forecast, $\hat{P}_b$, is also shown in this figure, where the forecast interval is $6:00-10:00$ am. The desired aggregated load for the interval of $7:00-8:30$ am (control interval) is denoted by P_d. Equations (24)-(26) were utilized to determine how the thermostat set points should be adjusted in order to achieve the desired load. The resulting adjustments are illustrated by the bottom curve of Figure 4.

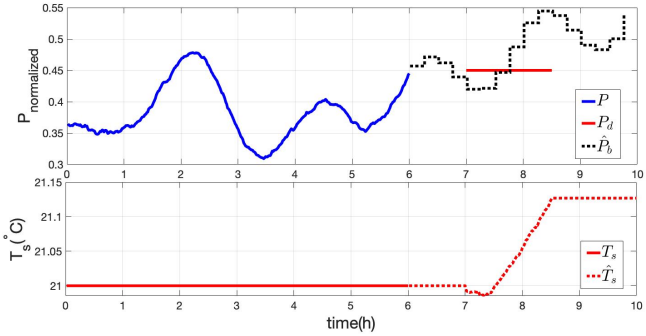


Figure 4: Example of desired load vs. calculated thermostat set point adjustments to follow the load.

5.4. Forecasting Rebound

Once the system operator changes the shape of the power demand over an interval, the profile of the demand will also change in the subsequent intervals, which is referred to as the payback or rebound effect of the control action. If an optimal controller adjusts the thermostat set points to follow the desired load demand and stops controlling it after that interval, the forecasting module can predict the shape of the demand after the control interval even before it starts. This feature can be beneficial for the system operator to estimate the shape of the load after a proposed desired aggregated load and determine if the consequences of the control actions are acceptable (e.g., not producing a larger peak in the future). However, it should be noted that various control strategies may affect this forecast.

Let us assume that the system operator issues a desired aggregated power demand, denoted by $P_d(k)$. This desired load can be considered as a change to the baseline load:

$$P_d(k) = \hat{P}_b(k) + \Delta P_d(k), \tag{27}$$

where $\Delta P_d(k)$ is the amount of power that should be added to or reduced from the baseline load. This change should be achieved by adjusting the thermostat set points ($\hat{u}_s$) to create an appropriate $\hat{y}_s$. Thus, the controller should create a $\hat{u}_s(k)$ to match $\hat{y}_s(k)$ with $\Delta P_d(k)$, so an estimation of the control signal will be:

$$\hat{u}_s(k) = (C_sB_s)^{-1}(\Delta P_d(k+1) - C_sA_s\hat{x}_s(k)) \tag{28}$$

Using (28), an estimation of the thermostat control signal can be calculated with the given desired aggregated load. With this

estimated u_s for the control interval and the forecast of the ambient temperature, a new forecast can be generated to approximate the payback of a control action.

$$\begin{cases} \begin{cases} \hat{x}_s(k+1) = A_s\hat{x}_s(k) + B_s\hat{u}_s(k) & \text{for } k \text{ in control interval} \\ \hat{x}_s(k+1) = A_s\hat{x}_s(k) & \text{otherwise} \end{cases} \\ \hat{y}_s(k) = C_s\hat{x}_s(k) \end{cases} \tag{29}$$

The aggregated power demand will be:

$$\hat{P}_{payback}(k) = \hat{P}_0 + \hat{y}_a(k) + \hat{y}_s(k) \tag{30}$$

6. Feasible Desired Demand

One of the key parameters in DLC is ensuring that the controller does not interfere with customers' comfort. For air conditioners, the controller can adjust the thermostat set point to change the aggregated power consumption to follow a desired load. However, if the controller sets the thermostat to a very high or very low value, it may cause discomfort for the end-users and prompt them to opt-out of the DLC program. Therefore, the controller cannot change the thermostat set point to any arbitrary value. The set point must be changed within a certain range to ensure that customers remain comfortable.

This means that the system operator must send a desired demand that can be followed without disturbing the customers. Such a demand is called a feasible desired demand. To ensure that only feasible desired demands are sent, a method will be presented in the next subsections that defines more precisely the customers' comfort and the criteria for feasible desired demands.

6.1. Customers' Comfort

One method of ensuring customers' comfort in DLC is to limit the range of change for the thermostat set point. The preferred thermostat set point for a group of loads can be denoted by T_s^0. The controller can then adjust the T_s within the range of $[T_s^0 - \alpha, T_s^0 + \alpha]$, where α determines the comfort zone. Once the T_s reaches these limits, the controller cannot adjust it beyond this range, thus losing control over the aggregated power. This signifies the point when there is no more reserve capacity available.

It is crucial to note that this method of limiting the range of T_s affects the controller's ability to adjust the aggregated power consumption to follow a desired load. Infeasible desired loads may arise due to the limitations imposed by the comfort zone. Thus, the system operator should only send feasible desired loads that can be followed without disturbing the customers' comfort. In the following subsections, the concept of customers' comfort and feasible desired loads will be defined more precisely, and a method for informing the system operator about the criteria for feasible desired loads will be presented.

6.2. Formal Formulation

A feasible desired dispatch is a power profile that satisfies three conditions. First, it should be within the range, i.e. $0 \le P_d(k) \le P_{max}$. Second, the controller should be able to produce appropriate changes to the thermostat set points, so that the aggregated demand of the loads would follow the desired load. Third, the changes on the thermostat set points should not exceed the customers' comfort level, as described previously in Section 6.1.

This means that the desired load should be within the range of feasible set point changes. If the desired load is outside of this range, the controller will not be able to follow it without causing discomfort to customers, and therefore, it is not a feasible desired demand.

Assuming that the aggregated power can be modeled with a second-order system, the transfer function representation is given by:

$$\frac{Y_s(z)}{U_s(z)} = \frac{b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} \tag{31}$$

Here, $Y(z)$ and $U(z)$ represent the z-transform of the $y_s(k)$ and $u_s(k)$ signals, respectively. This equation in time domain will be:

$$b_1 u_s(k-1) + b_2 u_s(k-2) = y_s(k) + a_1 y_s(k-1) + a_2 y_s(k-2) \tag{32}$$

In this equation, $y_s(k)$ represents the required changes that should be compensated by changing the thermostat set points, and $u_s(k)$ is the control signal which is the change in thermostat set point, issued by the controller to follow $y_s(k)$. It is assumed that the initial conditions are zero, i.e.:

$$u_s(k) = 0 \qquad \text{for } k < 0 \tag{33}$$

and

$$\Delta P_d(k) = 0 \qquad \text{for } k \le 0 \tag{34}$$

If we require $y_s(k)$ to track $\Delta P_d(k)$ over the control interval, then it can be shown that the solution for $u_s(k)$ is:

$$\begin{aligned} u_s(k) = \quad & \rho\Delta P_d(k+1) + \sigma\Delta P_d(k) + \tau\Delta P_d(k-1) \\ & + \textstyle\sum_{i=1}^{k-2} \left((-b_2/b_1)^{k-1-i}\tau\Delta P_d(i)\right) \end{aligned} \tag{35}$$

where

$$\begin{cases} \rho = \frac{1}{b_1} \\ \sigma = \frac{a_1 b_1 - b_2}{b_1^2} \\ \tau = \frac{a_2 b_1^2 - a_1 b_1 b_2 + b_2^2}{b_1^3} \end{cases} \tag{36}$$

As discussed earlier, a given desired demand is feasible if, for all time steps k, it satisfies the following two conditions:

$$T_s^0 - \alpha \le u_s(k) \le T_s^0 + \alpha \tag{37}$$

or:

$$\begin{aligned} T_s^0 - \alpha \le \quad & \rho\Delta P_d(k+1) + \sigma\Delta P_d(k) + \tau\Delta P_d(k-1) + \\ & \textstyle\sum_{i=1}^{k-2} \left((-\frac{b_2}{b_1})^{k-1-i}\tau\Delta P_d(i)\right) \le T_s^0 + \alpha \end{aligned} \tag{38}$$

To ensure the feasibility of a desired demand profile, it is essential to verify that it satisfies Equation (38) for all k samples of the profile. However, in order to obtain a general form for feasibility, it may be useful to study the feasibility of specific class of desired demand profiles.

An example class is the trapezoidal shape, where y_s starts from 0 with a constant slope, reaches a maximum value of P_d^{max} and remains constant for k_c samples, then goes back down to 0 with a constant slope.

One approach to studying feasibility is to determine the length of time that $u_s(k)$ can satisfy (37) when ΔP_d is a ramp. This evaluation can be repeated for different values of the ramp rate.

Figure 5 displays two ramp-shaped desired demand profiles and the corresponding control signal u_s. The u_s curve for each profile is computed using Equation (35) until the customers' comfort level is reached at $T_s^0 \pm \alpha$, where $\alpha = 0.5°C$. The plot illustrates that higher ramp slopes require more changes in T_s, causing the system to reach its limit faster. Consequently, the feasibility of profiles with steeper slopes is lower than those with slower rates of change. Figure 10 compares the feasibility intervals of ramp-shaped desired demand profiles with various slopes.

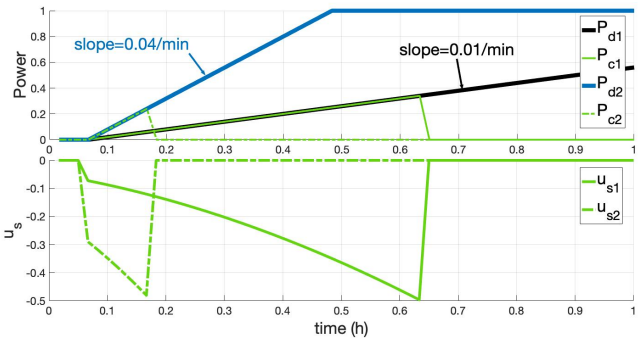


Figure 5: Feasibility intervals for two ramp shape desired demand profiles

It should also be mentioned that the ramp-rate is given for normalized power, which means that the actual rate for power will be obtained by multiplying this rate with the number of individual loads N and nominal power P_{nom}. For instance, a rate of $0.02/min$ for a group of $1,000$ air conditioners with a nominal rating of $2kW$ would correspond to a ramp rate of $40kW/min$.

Feasibility studies can also be conducted for the class of trapezoidal-shaped desired demand profiles with varying slope and P_d^{max}. Figure 6 provides an example of such a profile, along with the calculated T_s needed to track the desired demand when α is assumed to be $0.5°C$. As the slope and P_d^{max} values change, so will the corresponding feasibility interval. Figure 12 illustrates how the feasibility interval is affected by these parameters.

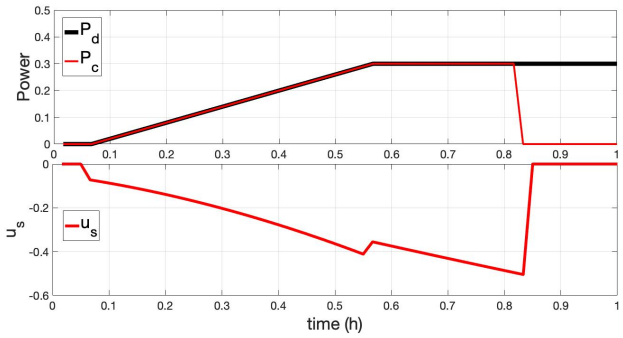


Figure 6: Feasibility interval for a trapezoidal-shaped desired demand

7. Simulation Results

To assess the effectiveness of the proposed method, a numerical simulator was developed in MATLAB to simulate a population of $N = 10,000$ individual air conditioners. As explained in Section 3, we assume that all ACs have the same nominal power, and their thermal capacities follow a log-normal distribution. The simulation employs values listed in Table 1, which have also been used in several other studies.

Table 1: Simulation Parameters [13]

Parameter	Value	Description
P_{nom}	$2kW$	Nominal power rating of the AC
η	3	Coefficient of performance
R	$2°C/kW$	Thermal resistance
μ_C	$3.6kWh/°C$	Mean value of the thermal capacitance
σ_{rel}	0.2	Standard deviation of log-normal distribution for C

7.1. System Modeling

For the given population of ACs in this simulation, the mathematical system that was identified is formulated as:

$$A_a = \begin{bmatrix} 1.975 & -.977 \\ 1 & 0 \end{bmatrix}, B_a = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, C_a = \begin{bmatrix} 0.0027 & -.0025 \end{bmatrix}$$

$$A_s = \begin{bmatrix} 1.898 & -.903 \\ 1 & 0 \end{bmatrix}], B_s = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, C_s = [-.1381, .1376]$$

7.2. Baseline Forecast

Figure 7 shows an example of a forecast profile for the next four hours, generated by the forecasting module at 6:00am on July 3, 2020, using the ambient temperature profile in Barbados. The dashed curve represents the forecasted normalized aggregated power calculated using Equation (16), while the solid curve represents the actual normalized aggregated power calculated using a numerical simulator. The figure illustrates that the forecast module is capable of predicting future power consumption with high accuracy, demonstrating the effectiveness of the mathematical model.

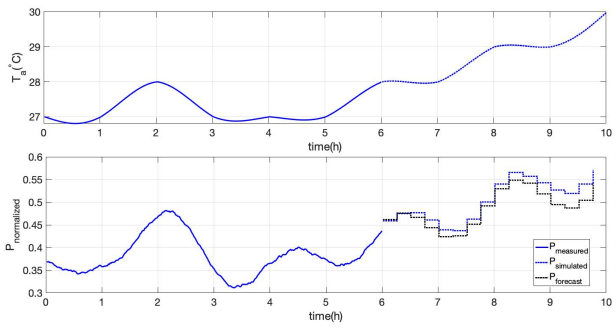


Figure 7: A forecast of the aggregated power consumption, vs. simulated aggregated power consumption in 15 minutes intervals for 4 hours.

Comparison metrics such as the mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE) are commonly used by researchers to assess the accuracy of forecast models. The proposed method's performance is evaluated using these metrics and compared with other methods in Tables **??**. The results show that the proposed method has slightly improved the performance of the baseline forecast. However, the main strength of this method is its ability to predict the load in the presence of control actions, as discussed in the next section. It should also be noted that the MAE and RMSE are calculated based on the normalized power, which is the total power demand divided by the number of loads and the nominal power of each device.

7.3. Forecasting Rebound Effect

Figure 8 illustrates the impact of a desired aggregated load on the original baseline forecast ($\hat{P}_b$) and the resulting baseline with payback ($\hat{P}_{payback}$). The blue curve represents the original forecast for the baseline load, while the green curve depicts the desired aggregated load. The orange curve shows the resulting forecast after applying the desired load, which includes the payback effect. The payback effect causes the load profile to be higher after the control interval, compared to the profile if no control action was taken. This figure demonstrates how the desired aggregated load can be used to adjust the power consumption profile over a specified interval.

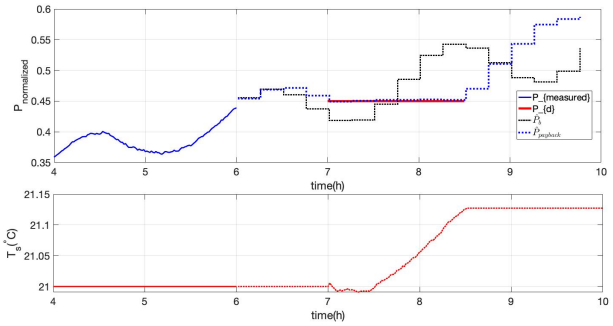


Figure 8: Baseline forecast ($\hat{P}_b$) and forecast with payback effect($\hat{P}_{payback}$).

7.4. Forecasting Control Capacity

Figure 9 illustrates an example of using a high-rate slope as a desired demand (the red curve). Applying the mathematical model, the required temperature set point is depicted in green in this figure. Applying these values for the thermostat set points on a simulated population of ACs will result in the aggregated power demand that is shown by the solid blue curve. If we apply a controller to follow this desired demand, the thermostat set point and the aggregated power look like the dotted blue curves.

In Figure 10, the feasibility intervals of desired demand profiles with different slopes are compared. The graph indicates that as the ramp speed increases, the time during which the load can be controlled to align with the desired power consumption decreases. This phenomenon is attributed to the necessity of preserving customers' comfort as discussed earlier.

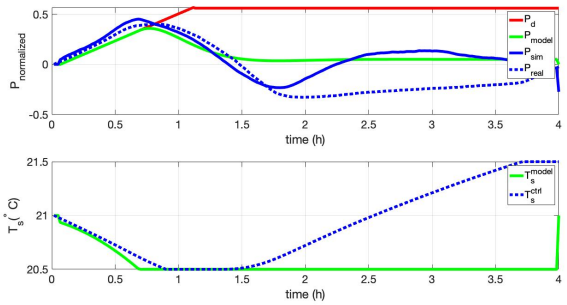


Figure 9: Feasibility of a high-speed ramp desired demand obtained from the model and simulation

Figure 11 illustrates an example of using a trapezoidal-shape as a desired demand (the red curve). Applying the mathematical model, the required temperature set point is depicted in green in this figure. Applying these values for the thermostat set points on a simulated population of ACs will result in the aggregated power demand that is shown by the solid blue curve. If we apply a controller to follow this desired demand, the thermostat set point and the aggregated power look like the dotted blue curves.

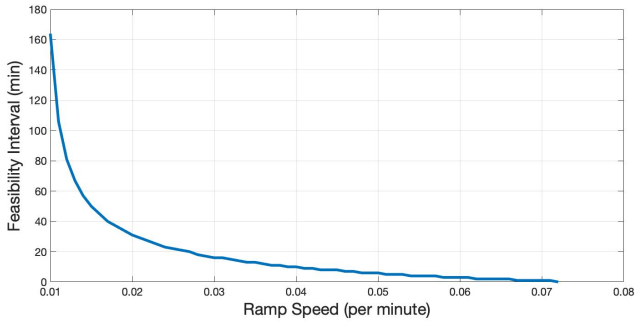


Figure 10: Feasibility intervals for ramp shape desired demand profiles with varying slopes

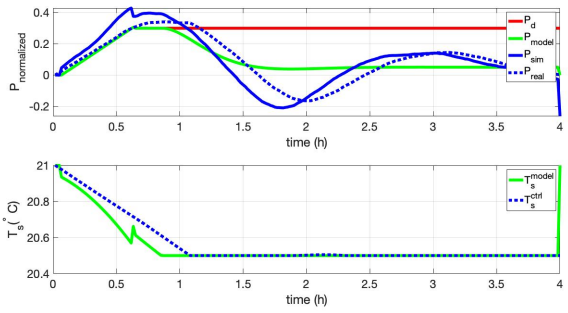


Figure 11: Feasibility of trapezoidal desired demand obtained from the model and simulation

An important aspect of the forecast module is to communicate this information to the system operator. With this knowledge, the operator can make informed decisions on how to generate the desired demand, taking into account the capacity of the controller to track the load.

Figure 12 visually demonstrates the impact on the feasibility interval when altering the slope and P_d^{max} values of the trapezoidal shape. Once more, the observation is evident: a higher slope (faster rate) and a larger maximum value lead to reduced controllability intervals in order to maintain customers' comfort.

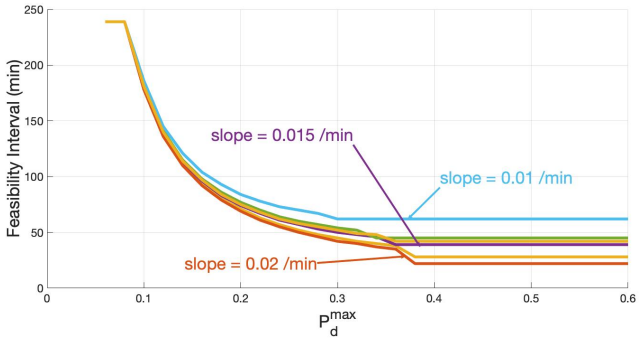


Figure 12: Feasibility intervals for trapezoidal shaped desired demand profiles with varying slopes and P_d^{max}

Conclusions

Thermostatically controlled devices have the potential to be utilized as controllable loads for direct load control, presenting a practical solution to address rapid fluctuations in power generation. The effective management of such loads requires precise predictions of future power consumption and the ability to adjust capacity relative to the baseline.

This study introduces an intelligent agent specifically designed to learn the behavior of a population of air conditioners. It provides a mathematical model for forecasting parameters such as baseline load, controllable load, and the impact of past and present control actions on the forecast. Moreover, the agent assesses potential alterations in the load. The proposed learning mechanism offers two significant advantages: 1) it eliminates the need for knowledge about physical parameters like thermal capacitance or resistance, and 2) it demonstrates adaptability to changing parameters over time. This makes it a practical and versatile plug-and-play load aggregator suitable for deployment in various locations and environments.

References

[1] P. Yazdkhastil, C. P. Diduch, "Smart Agent-Based Direct Load Control of Air Conditioner Populations in Demand Side Management," in 2023 International Conference on Control, Automation and Diagnosis (ICCAD), 1–7, 2023, doi:10.1109/ICCAD57653.2023.10152358.

[2] M. Fan, K. Sun, D. Lane, W. Gu, Z. Li, F. Zhang, "A Novel Generation Rescheduling Algorithm to Improve Power System Reliability with High Renewable Energy Penetration," IEEE Transactions on Power Systems, **33**(3), 3349–3357, 2018, doi:10.1109/TPWRS.2018.2810642.

[3] A. Halder, X. Geng, P. R. Kumar, L. Xie, "Architecture and algorithms for privacy preserving thermal inertial load management by a load serving entity," in 2017 IEEE Power Energy Society General Meeting, 1–1, 2017, doi:10.1109/PESGM.2017.8274310.

[4] Y. Cheng, "Methods for mitigating the effects of intermittent energy production of photovoltaic sources," in 2011 International Conference on Power Engineering, Energy and Electrical Drives, 1–6, 2011, doi:10.1109/PowerEng.2011.6036564.

[5] L. Hucheng, W. Xinwei, Y. Yubo, S. Wei, G. Yanfeng, "Simulation and analysis of electric water heater load regulation model based on direct load control," in 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2), 1–5, 2017, doi:10.1109/EI2.2017.8245591.

[6] M. Shaad, C. P. Diduch, M. E. Kaye, L. Chang, "A basic load following control strategy in a direct load control program," in 2015 IEEE Electrical Power and Energy Conference (EPEC), 339–343, 2015, doi:10.1109/EPEC.2015.7379973.

[7] T. Mega, S. Kitagami, S. Kawawaki, N. Kushiro, "Experimental Evaluation of a Fast Demand Response System for Small/Medium-Scale Office Buildings," in 2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA), 291–295, 2017, doi:10.1109/WAINA.2017.84.

[8] M. Shad, A. Momeni, R. Errouissi, C. P. Diduch, M. E. Kaye, L. Chang, "Identification and Estimation for Electric Water Heaters in Direct Load Control Programs," IEEE Transactions on Smart Grid, **8**(2), 947–955, 2017, doi:10.1109/TSG.2015.2492950.

[9] P. Yazdkhasti, C. P. Diduch, "A Mathematical Model for the Aggregated Power Consumptions of Air Conditioners," in 2018 the 6th international conference on smart energy grid engineering, 1 – 8, 2018.

[10] S. Kundu, N. Sinitsyn, S. Backhaus, I. Hiskens, Modeling and control of thermostatically controlled loads, Power Systems Computation Conference, 2011.

[11] W. Zhang, J. Lian, C. Y. Chang, K. Kalsi, "Aggregated Modeling and Control of Air Conditioning Loads for Demand Response," IEEE Transactions on Power Systems, **28**(4), 4655–4664, 2013, doi:10.1109/TPWRS.2013.2266121.

[12] J. Mathieu, S. Koch, D. Callaway, "State estimation and control of electric loads to manage real-time energy imbalance," IEEE Transactions on Power Systems, **28**(1), 430–440, 2013, doi:10.1109/TPWRS.2012.2204074.

[13] N. Mahdavi, J. H. Braslavsky, M. M. Seron, S. R. West, "Model Predictive Control of Distributed Air-Conditioning Loads to Compensate Fluctuations in Solar Power," IEEE Transactions on Smart Grid, **8**(6), 3055–3065, 2017, doi:10.1109/TSG.2017.2717447.

[14] S. H. Tindemans, V. Trovato, G. Strbac, "Decentralized Control of Thermostatic Loads for Flexible Demand Response," IEEE Transactions on Control Systems Technology, **23**(5), 1685–1700, 2015, doi:10.1109/TCST.2014.2381163.

[15] K. Meng, D. Wang, Z. Y. Dong, X. Gao, Y. Zheng, K. P. Wong, "Distributed control of thermostatically controlled loads in distribution network with high penetration of solar PV," CSEE Journal of Power and Energy Systems, **3**(1), 53–62, 2017, doi:10.17775/CSEEJPES.2017.0008.

[16] C.-R. Chen, S.-C. Shih, S.-C. Hu, "Short-term Electricity Forecasting of Air-conditioners of Hospital Using Artificial Neural Networks," in 2005 IEEE/PES Transmission Distribution Conference Exposition: Asia and Pacific, 1–5, 2005, doi:10.1109/TDC.2005.1547136.

[17] D.-X. Niu, Q. Wanq, J.-C. Li, "Short term load forecasting model using support vector machine based on artificial neural network," in 2005 International Conference on Machine Learning and Cybernetics, volume 7, 4260–4265 Vol. 7, 2005, doi:10.1109/ICMLC.2005.1527685.

[18] S. Singh, S. Hussain, M. A. Bazaz, "Short term load forecasting using artificial neural network," in 2017 Fourth International Conference on Image Information Processing (ICIIP), 1–5, 2017, doi:10.1109/ICIIP.2017.8313703.

[19] C. Perfumo, J. H. Braslavsky, J. K. Ward, "Model-Based Estimation of Energy Savings in Load Control Events for Thermostatically Controlled Loads," IEEE Transactions on Smart Grid, **5**(3), 1410–1420, 2014, doi:10.1109/TSG.2014.2298840.

[20] R. Gao, , X. Liu, "Short-term load forecasting based on least square support vector machine combined with fuzzy control," in Proceedings of the 10th World Congress on Intelligent Control and Automation, 1048–1051, 2012, doi:10.1109/WCICA.2012.6358034.

[21] P. Yazdkhasti, C. P. Diduch, "A PID controller for Direct Load Control of Thermostatically Controlled Appliances," in 2019 IEEE 58th Conference on Decision and Control (CDC), 1913–1918, 2019.

[22] N. Mahdavi, J. H. Braslavsky, C. Perfumo, "Mapping the Effect of Ambient Temperature on the Power Demand of Populations of Air Conditioners," IEEE Transactions on Smart Grid, **9**(3), 1540–1550, 2018, doi: 10.1109/TSG.2016.2592522.

[23] C. Perfumo, E. Kofman, J. Braslavsky, J. Ward, "Load Management: Model-Based Control of Aggregate Power for Populations of Thermostatically Controlled Loads," Energy Conversion and Management, **55**, 36–48, 2012.

[24] P. Yazdkhasti, C. P. Diduch, "A Model-Based Short-term Load Forecast Methodology for Aggregated Power Consumption of Thermostatically Controlled Appliances in DSM," in 2021 IEEE Energy Conversion Congress and Exposition (ECCE), 727–732, 2021, doi:10.1109/ECCE47101.2021.9595608.

[25] L. Von Krannichfeldt, Y. Wang, G. Hug, "Online Ensemble Learning for Load Forecasting," IEEE Transactions on Power Systems, **36**(1), 545–548, 2021, doi:10.1109/TPWRS.2020.3036230.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 6, 120-143 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Dual Mode Control of an Inverted Pendulum: Design, Analysis and Experimental Evaluation

Laura Álvarez-Hidalgo, Ian S. Howard*

School of Engineering, Computing & Mathematics, University of Plymouth, Plymouth, PL4 8AA UK

ARTICLE INFO

Article history:
Received: 14 November, 2023
Accepted: 13 December, 2023
Online: 30 December, 2023

Keywords:
Inverted pendulum
State space control
Luenberger observer
3D printing
LQR

ABSTRACT

We present an inverted pendulum design using readily available V-slot rail components and 3D printing to construct custom parts. To enable the examination of different pendulum characteristics, we constructed three pendulum poles of different lengths. We implemented a brake mechanism to modify sliding friction resistance and built a paddle that can be attached to the ends of the pendulum poles. A testing rig was also developed to consistently apply disturbances by tapping the pendulum pole, characterizing balancing performance. We perform a comprehensive analysis of the behavior and control of the pendulum. This begins by considering its dynamics, including the nonlinear differential equation that describes the system, its linearization, and its representation in the s-domain. The primary focus of this work is the development of two distinct control modes for the pendulum: a velocity control mode, designed to balance the pendulum while the cart is in motion, and a position control mode, aimed at maintaining the pendulum cart at a specific location. For this, we derived two different state space models: one for implementing the velocity control mode and another for the position control mode. In the position control mode, integral action applied to the cart position ensures that the inverted pendulum remains balanced and maintains its desired position on the rail. For both models, linear observer-based state feedback controllers were implemented. The control laws are designed as linear quadratic regulators (LQR), and the systems are simulated in MATLAB. To actuate the physical pendulum system, a stepper motor was used, and its controller was assembled in a DIN rail panel to simplify the integration of all necessary components. We examined how the optimized performance, achieved with the medium-length pendulum pole, translates to poles of other lengths. Our findings reveal distinct behavioral differences between the control modes.

1. Introduction

1.1. Declaration

This paper represents a substantial extension of work originally presented at the 2022 International Conference on System Science and Engineering (ICSSE) [1]. The differences with the previous publication are that here we now include the ability to change the physical pendulum's mechanical characteristics, provide a more rigorous system analysis, build a custom testing rig, and improve upon previous system identification and testing procedures.

1.2. Overview

An inverted pendulum is a mechanical system comprising a rigid pole, with a pivot at one end that is located on a mobile cart. The challenge of building an inverted pendulum is that, in its inverted upright configuration, it represents a marginally an unstable system. The task is to maintain the system in this inherently unstable upright position, even in the presence of minor disturbances such as a gentle tap. Achieving this balance requires the implementation of a control strategy. The control mechanism must continuously measure the angular displacement of the pendulum from the vertical and correspondingly manipulate the cart's position to counteract any deviations.

1.3. Previous work

The inverted pendulum is widely recognized as one of several classical problems in the field of control engineering that is enlightening to study [2]. It has been used as a benchmark in robotics and control theory for almost the last 100 years and is often chosen to test and evaluate new control methods [3]. This preference arises because pendulum balancing represents behavior

*Corresponding Author: Ian S. Howard, ian.howard@plymouth.ac.uk

relevant to a wide range of theoretical challenges and practical applications [4,5].

One prominent example of these behaviors is the inverted pendulum's manifestation as a simple non-linear dynamic system, characterized by stable and unstable equilibrium points. During the dynamic transition from a downward to an upright position, the pendulum exhibits nonlinearity, primarily arising from how the pendulum angle affects the applied torque due to the pendulum's weight [6]. In addition, it represents a system for which it is crucial to achieve stability at the upright position in the presence of disturbances, which can only be achieved by moving the cart appropriately along its rail [7].

James Kerr Roberge was one of the first researchers to describe an inverted pendulum [6]. Since then, many inverted pendulum designs and their variants have been constructed and studied, including ones aimed at teaching and research [8,9], mobile designs [10,11], pendulums mounted on drones [12], as well as rotary designs [13,14].

Derivations of the dynamics of inverted pendulums and their simulation have been carried out by many researchers, for example [15]. Many different approaches to control have also been investigated. These include PID and classical approaches to control in the s-domain [16,17], state feedback control [1,18], and Lyapunov-based controller design [19], and comparisons have been made between different controllers [20]. Machine Learning (ML) approaches are becoming an increasingly successful way to deal with hard control problems. They mark a change from designing controllers based on explicit mathematical models derived from physics to more empirical methodologies that are essentially data driven [21]. ML approaches make use of reinforcement learning (RL) based on Q-Learning [22], Policy Iteration [23], and Deep Q-Networks [24]. The PILCO RL algorithm is especially data efficient, since it builds and makes use of a probabilistic model of the task dynamics as it learns to balance the pendulum [25]. Further ML approaches involve neural networks [26,27] and genetic algorithms [28,29]. Hybrid methods using control engineering approaches and neural network techniques have also been investigated [30]. Researchers often compare different control approaches [31]. More complex two-link inverted pendulums have also been studied by several researchers and implemented using a range of control techniques [32]. Other researchers have even investigated the use of reinforcement learning to control a pendulum with three links [33].

The inverted pendulum also forms a basis for understanding simple balancing robots [34–36]. There is also increasing interest in the construction of legged and humanoid robots[37,38], in which control of balance plays an important role [39–41].

Falls in the elderly are a common health issue worldwide and consequently understanding the mechanisms of how humans maintain balance whilst standing is an area of much research [42–45]. The inverted pendulum has also been used to model and understand this process [44–52]. More recent work has also included the use of experiments with robotic manipulanda to investigate how humans can balance items with their hands [53–56].

Given the significance of the inverted pendulum in the field of control engineering, inverted pendulum theory finds extensive applications across diverse fields, including robotics, aerospace systems, marine systems, flexible systems, mobile systems, and locomotive systems [57–59]. The characteristics of the inverted pendulum make it well-suited for modelling a multitude of practical scenarios, highlighting its significance and profound influence across various industries.

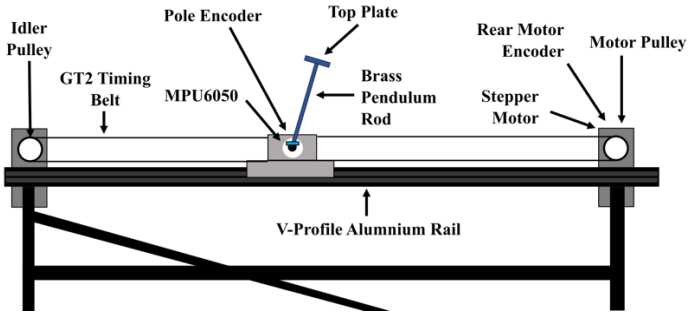


Figure 1: Schematic diagram showing main pendulum components.

2. Mechanical design of the Inverted Pendulum

2.1. Extension of previous work

The current inverted pendulum design expands upon our previous publication [1], in several important ways. We made modifications to the mechanical components of the pendulum system, enabling us to alter its physical characteristics and now utilize a range of pendulum configurations, with changes in pendulum length (635mm, 335mm, 233mm), viscous damping (adding a paddle to the pole), and friction (compression of a brake on the pendulum pole).

2.2. Inverted pendulum components

Here we build an inverted pendulum system consisting of a pole pivoted at one end on a cart that moves on a linear track actuated by a stepper motor, which can be balanced in its inverted position by observing the pole angle and controlling cart movement. Our design comprises several distinct component parts, illustrated in Figure 1.

The pendulum assembly is composed of several distinct components. These include a v-groove aluminum profile track and a cart unit that supports the inverted pendulum, moving along the track on a wheeled cart. Additionally, a stepper motor unit propels the cart via a timing belt from one end of the track, while a passive idler pulley supports the belt at the opposite end. The motor and pulley mechanisms are securely affixed to the aluminum profile structure using T-nuts. This design not only offers the flexibility to easily remove and replace these parts with similar components but also simplifies the process of tensioning the drive belt.

2.3. Track

The mechanical design integrated a V-slot profile to realize the pendulum track. This streamlined construction since it enabled the use of readily available accessories. These included a stepper motor mounting plate, an idler pulley mechanism, as well as gantry plates.

2.4. Cart

A 4-wheeled gantry plate formed the basis of the cart mechanism. During balancing operation, it traversed the V-slot aluminum track along its sides as needed to balance the pendulum pole. Wheel clearances were adjusted so the cart remained securely on the track without excessive wobbling, while also avoiding undue friction.

A thick flat PLA+ 3D-printed rectangular sheet with mounting holes was affixed to the gantry plate. This served as the support for both a ball bearing race and an encoder unit, which held and facilitated the rotation of a shaft. This shaft, in turn, held the pendulum pole, allowing it to swing freely. The incremental encoder measured the pole's angular deviation from the vertical position, as depicted in Figure 2.

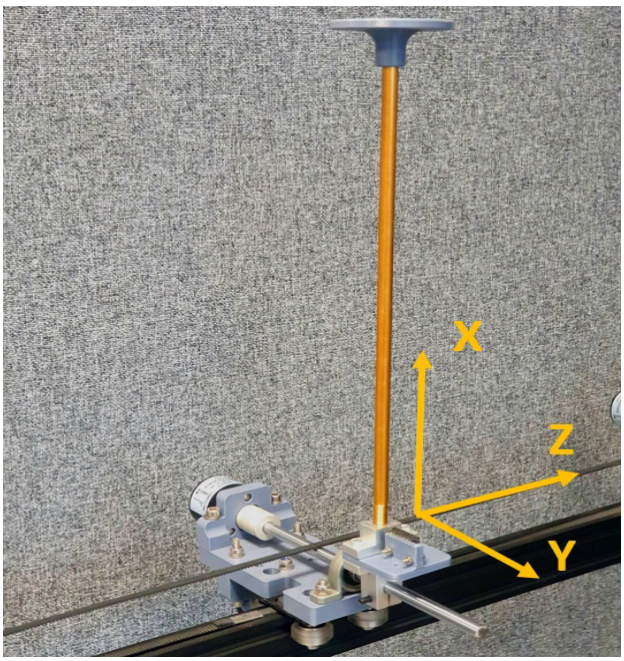


Figure 2: 3D-printed pendulum cart incorporating an encoder and an inertial measurement unit (IMU). A flat-topped table can be attached to the end of the pendulum to support objects on its surface.

2.5. Shaft friction adjustment

Normally, the rotation of the pendulum shaft is resisted by a low level of friction, arising from the bearing and the incremental encoder. To increase the amount of friction and examine its effect on pendulum behavior, a spring-loaded braking assembly was constructed (Figure 3). This consisted of a semicircular brake pad section made of PLA+ that could be pressed against the pendulum rotary shaft using a compression spring, thereby hindering its rotation. By fully withdrawing the brake, it was also possible to remove its effect completely.

2.6. Standard pendulum pole

The standard pendulum pole consists of a pole crafted from a brass segment, selected for its easy machinability and high density. One end of the pole was threaded to securely screw into an attachment component connected to the main shaft, ensuring a sturdy attachment as depicted in Figure 2. This led to an overall pendulum length of 335mm. While a relatively short pole increases the balancing challenge, necessitating quicker cart reactions due to the system's elevated natural frequency, it yields several benefits. A compact pole is not only more manageable but also ensures increased safety by minimizing accidental impact risks. Moreover, it provides characteristics that better match other systems, like smaller balancing robots [60].

2.7. Additional pendulum poles

An easy way to alter the fundamental characteristics of the pendulum is to change the length of the pole. To do so, two additional pendulum poles were built (Figure 4). These poles consisted of 8mm diameter stainless steel poles, leading to pendulum lengths of 222mm and 635mm. Since they were only required for intermittent use, no screw attachment was used, thereby facilitating construction. Instead, they were simply clamped at their endpoint into another attachment component connected to the main shaft.

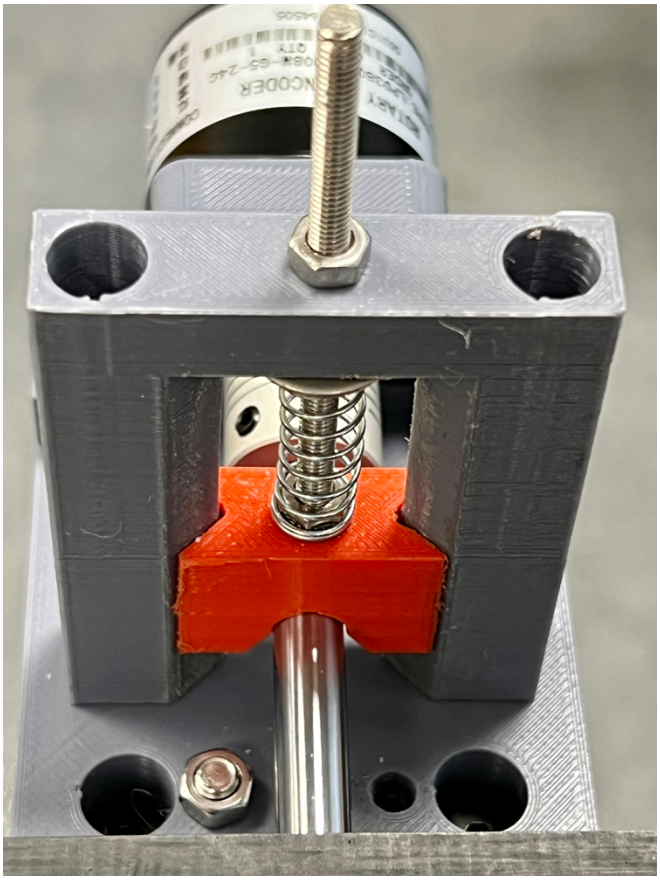


Figure 3: Shaft friction adjustment mechanism for the pendulum cart, designed to alter the sliding friction around the pendulum's rotational axis.

2.8. Pendulum pole end attachments

To provide a platform for placing objects, and to shield its endpoint for safety reasons, a round disc was 3D printed from PLA+ and slid onto the end of the pendulum pole, where it was held in place by friction.

To offer a means to change the viscous air resistance experienced by the pendulum pole as it swung, the round disc at the endpoint of the pole could be replaced with a paddle (Figure 5). The paddle consisted of a 5mm thick square measuring 100mm

by 100mm and was 3D printed from PLA+. It slid onto the end of the pole via a central mounting hole and was again held in place by friction. By rotating the paddle 90°, it was possible to adjust the amount of viscous resistance the pendulum pole experienced from a low to a high value.

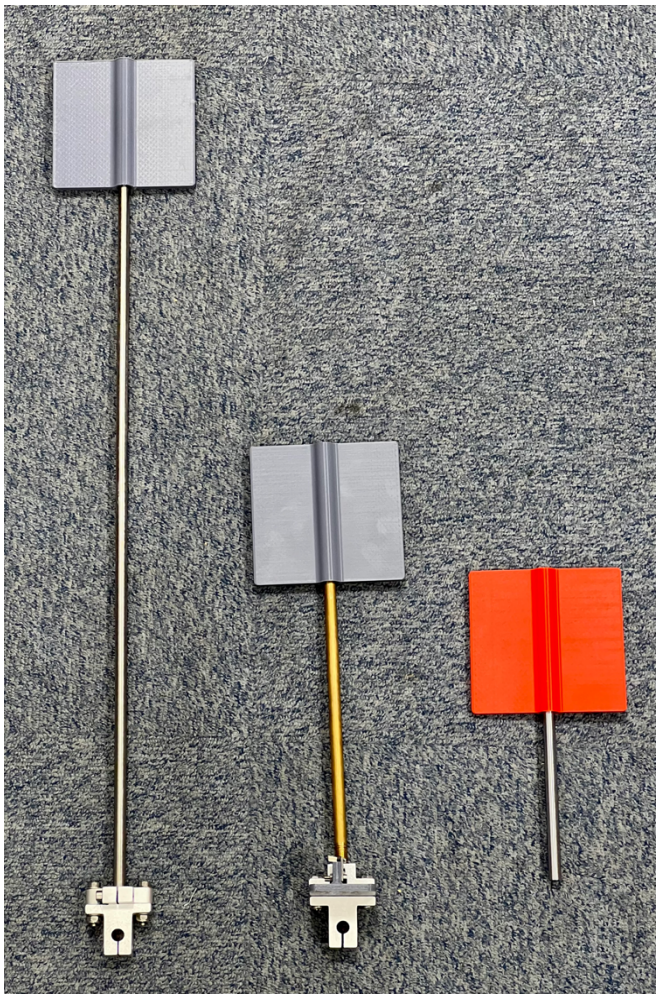


Figure 4: Three different pendulum poles were utilized to evaluate the controller's sensitivity to changes in pendulum length.

2.9. Stepper motor actuation

The motor drive assembly includes an aluminum plate, situated on the profile rail, which serves as firm support for a NEMA23 stepper motor. The motor is securely affixed to the plate using bolts. The motor is connected to a drive pulley at its front, and an encoder is mounted on the rear end of its shaft. This enables accurate measurement of the cart's position, although it is only needed to analyze the pendulum's behavior and is not involved in the balancing process (see Figure 6).

To operate the stepper motor, an A4988 stepper controller is employed, driven from an Arduino Mega 2560 R3 Microcontroller. The latter is programmed in C++ and provides precise control of the pendulum cart along the linear rail.

2.10. Belt

A pulley and belt mechanism are used to convert the motor's rotary motion into linear movement, thereby appropriately driving the cart along its rail. The cart traverses its designated rails using a GT2 timing belt, which is typically used in 3D printers. The belt, affixed to the cart using steel clamps, spans almost the entire length of the track. The stepper motor, located at one end of the track, has a 60-tooth GT2 motor pulley secured to its shaft. A passive idler pulley is situated at the opposite end of the track. Ball bearings, integrated into the idler pulley, minimize frictional resistance, ensuring smooth operation even under the stress of high belt tension. Adjusting the precise location of the idler provides an easy means to modify belt tension.

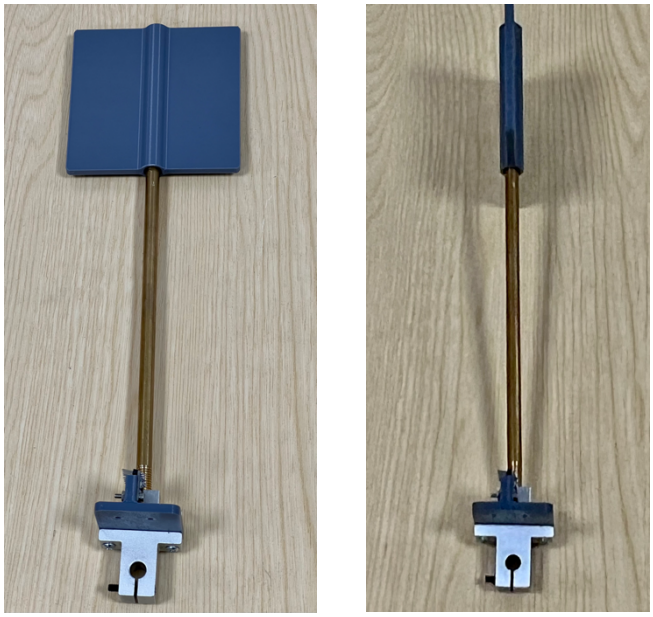


Figure 5: The pendulum paddle can be rotated, thus adjusting the viscous drag due to air resistance from low to high values.

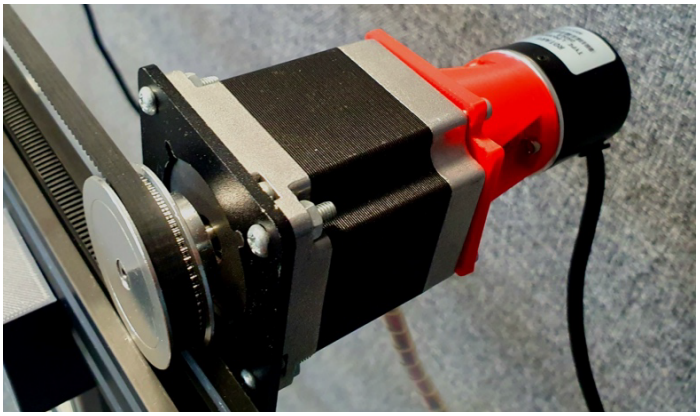


Figure 6: Stepper motor actuation, showing the drive pulley and a custom-made 3D-printed encoder mount at the motor's rear.

2.11. Inertial Measurement Unit

To support future developments of the pendulum system, a cost-effective 6-DOF accelerometer/gyro (MPU-6050) was strategically mounted to a 3D-printed support on the pendulum pole, aligning it with the pendulum shaft's rotational axis. This configuration presents an alternative method to measure the pole's angular displacement. The pendulum shaft's rotation revolves around the MPU-6050's y-axis. When in the inverted

configuration, its x-axis points downward, and the pole points upwards along its negative x-axis, with its z-axis horizontal (Figure 2).

2.12. Modular adjustable pendulum design

The utilization of modular construction within the system ensures that the parameters governing the behavior of the inverted pendulum can easily be adjusted or reconfigured. Such adaptability can prove useful in educational contexts. With minimal adjustments to the apparatus, diverse tasks can be allocated to distinct student cohorts, each tackling a specific control problem. For instance, the pole's length, an essential aspect of the system's dynamics, can be altered by substituting the pole with another of a different length. Adaptability extends further to the motor unit. For example, stepper motor drive could be exchanged with actuation employing a BLDC motor, a modification that would support force control, as opposed to velocity control, of the pendulum system.

2.13. 3D printing

The components for the pendulum cart were designed using AutoCAD Fusion 360. This software also facilitated the conversion of the designs into STL format files, which is a critical step for additive manufacturing. The mechanical parts were then fabricated using PLA+ material on a Creality 6SE 3D printer. It is noteworthy to mention that although tougher plastics could further enhance the durability of the design, PLA+ was chosen for its ease of printing and cost-effectiveness.

2.14. Pendulum Stand

Operating the pendulum necessitates mounting the track at an elevation that ensures unobstructed swinging of the pole. We designed a custom-engineered support stand using aluminum profiles to secure the pendulum system (refer to Figure 7). This stand offers a robust yet lightweight construction that facilitates easy transportation.

The support stand comprises two support pillars, fabricated from aluminum profile. These pillars are anchored at their base with additional lengths of aluminum profile, and 3D printed feet are used at each end to provide stable support. The top of each pillar is fitted with a 3D-printed bracket, tailor-made to accommodate the aluminum v-rail. To enhance the rigidity of the structure and to increase its resistance to mechanical stress, diagonal aluminum profile sections are incorporated, to brace the assembly. This results in a rigid structure, minimizing potential vibrations or displacements that could affect the system's performance.

3. Mechanical tapper for performance evaluation

3.1. Testing balancing systems

Monteleone and his team [61] presented a methodology to evaluate the balance resilience of robots, utilizing unique performance indicators and a custom-made testbed. Through extensive testing on a humanoid robot, their study demonstrated the method's effectiveness in designing more robust robotic systems.

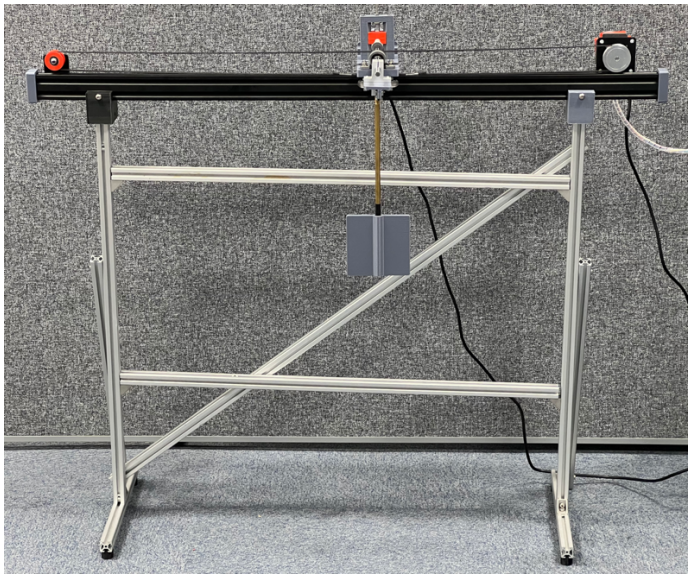


Figure 7: Pendulum mounted on its stand: The structure uses diagonal bracing to increase its rigidity.

3.2. Tapper components

In the same vein, to conduct tests across various pendulum conditions, including different pendulum lengths, friction, and damping levels, as well as different control laws, and to compare the results, it was necessary to disturb the pendulum pole consistently. To achieve this, a tapping mechanism was constructed (Figure 8).

We built and used a testing rig to deliver repeatable disturbances to the pendulum pole whilst balancing, to examine the recovery and robustness of control. This could be carried out during balancing whilst the cart was either static or moving.

3.3. Finger-spring mechanism

The primary component of this tapping mechanism is the 'tapper finger,' which is mounted onto a baseplate. This mounting baseplate for the tapping mechanism is attached to a cart that can be maneuvered up and down a V-groove rail track by means of a stepper motor.

The finger is composed of a stainless-steel pole inserted in a PLA+ holder, which pivots around a rotary axis located 3 cm from its lower end. Two sets of springs are connected at the endpoint of the holder and at the base on either side, pulling in opposite directions. When the finger is in its undisturbed equilibrium position, these springs ensure that it maintains a 0° orientation.

This finger-spring assembly forms an underdamped second-order system. Its behavior, particularly the overshoot following appropriate excitation, serves as an effective method to strike the pendulum pole. To generate a movement suitable for producing a tap, it is necessary to displace the lower end of the finger from its equilibrium position around its pivot and then release it suddenly. This action results in a rapid movement: the finger travels back through its equilibrium position and out the other side, which enables it to impact the pendulum pole and then quickly withdraw.

Using this tapping mechanism, it is important to note that the pendulum pole must be positioned at an appropriate distance from

the tapping finger before operation commences, to prevent multiple impacts. At the tap, energy is transferred from the finger to the pendulum pole, and if the distance to the target pendulum pole is set correctly, it ensures that the finger only strikes once and then retreats without making further contact.

3.4. Finger actuation

Although it would be possible to manually displace the lower end of the tapping pole and release it by hand, we incorporated an RC servomechanism into the design to achieve this action automatically and more consistently. The RC servo first displaces the finger from its equilibrium position. Then, owing to the cam mechanism's design, it releases the finger suddenly as it passes the end point. This action consequently results in a rapid underdamped second-order trajectory of the end of the tapper finger, ideal for exciting the pendulum pole.

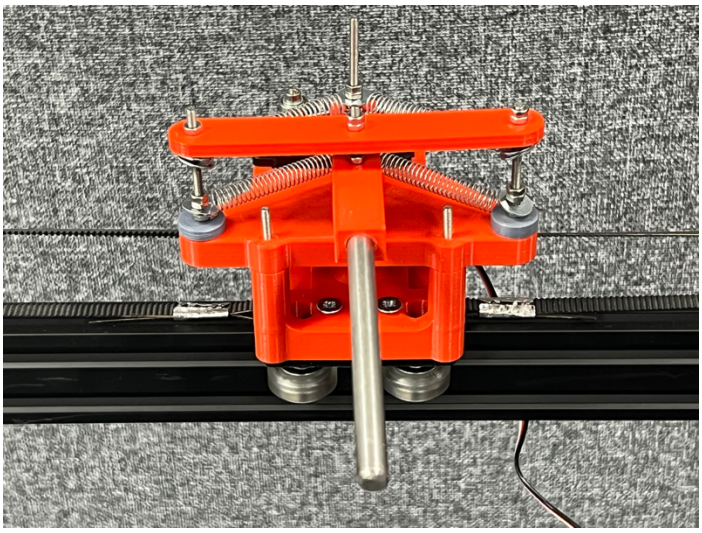


Figure 8: 3D-printed tapping mechanism. Two sets of springs are configured to pull the bottom of the tapping pole to the left and right, thereby establishing a neutral equilibrium position at 0° as depicted. An RC servo is positioned to travel 180°, engaging and then releasing the rear of the tapping mechanism. This action, assisted by the tension of simultaneously contracted springs, causes the pole to swing in an under-damped motion, with overshoot delivering an appropriate tap to the pendulum pole. All custom parts were designed using AutoCAD Fusion 360 and printed with PLA+ material.

To achieve a consistent tap, it is essential to maintain a constant distance between the tapping pole and the pendulum pole and to ensure that the tap occurs at the same location during each trial. This consistency is achieved by visually aligning the tapping finger with the pendulum pole before a tap is initiated.

3.5. Tapper cart

In the inverted pendulum position control mode, the pendulum cart remains stationary, simplifying the process of tapping its pole. However, in the velocity control mode, the cart moves along the rail while balancing. The goal was to create a tapping mechanism suitable for both velocity and position modes, necessitating the ability of the tapping mechanism to track the pendulum cart's movement by employing an additional cart. This capability ensures taps can be delivered effectively, even while the pendulum cart is in motion. To achieve necessary synchronization, the tapping mechanism's cart is propelled along a separate V-groove aluminum profile track, using a stepper motor that receives the same control signal as the pendulum cart's stepper motor.

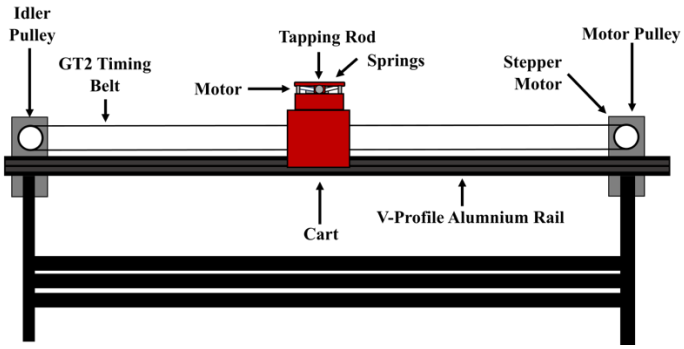


Figure 9: Schematic of the tapping mechanism mounted on its adjustable stand, showing all its main components.

3.6. Adjustable height tapper stand

The upper track of the tapper mechanism was mounted to the side of the support pillars, allowing for adjustable height of the tapper, as illustrated in Figures 9 and 10.

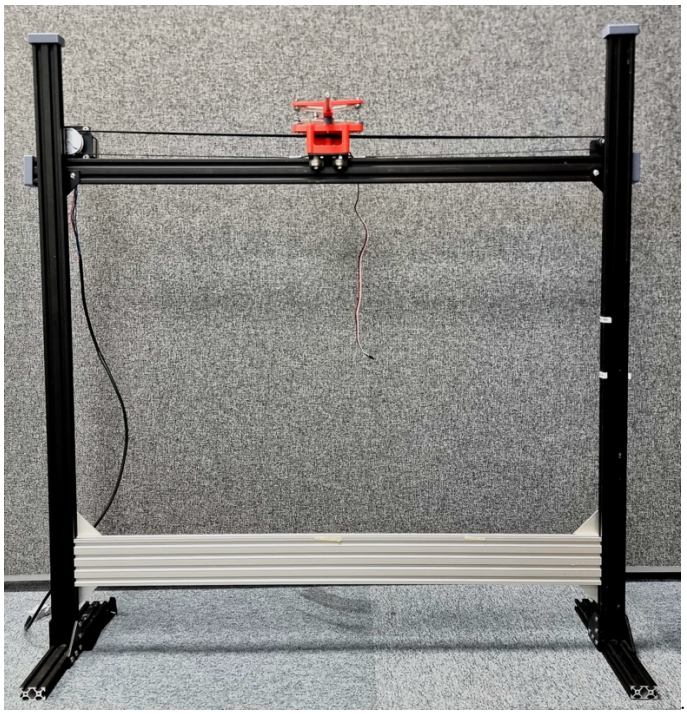


Figure 10: Side view of the adjustable mechanical tapper assembly. The cart supporting the tapping mechanism can be driven left and right to synchronize with the pendulum cart. The cart is mounted on a rail that can be slid up and down the outer stand legs, and then fastened firmly in place with screws, enabling adjustment of the height at which tapping occurs.

4. Analysis of pendulum dynamics

4.1. Mathematical analysis

We performed a comprehensive analysis of the pendulum's dynamics, including the nonlinear differential equation, its linearization, and s-domain representation. This provided a theoretical foundation for our practical implementation.

4.2. Equilibrium positions

A pendulum exhibits two distinct equilibrium points. In the stable equilibrium state, the pendulum hangs downward, functioning like a traditional pendulum, similar to those in pendulum clocks. At this stable equilibrium point, if the pendulum is slightly displaced, it will begin to oscillate back and forth with a characteristic frequency determined by its dynamic properties. Factors such as damping in the joints and air resistance lead to a gradual decrease in oscillation amplitude over time. Eventually, the pendulum will stop moving and return to a stationary state at its equilibrium position. In contrast, an unstable equilibrium occurs when the pendulum is delicately balanced upright on its pivot point.

4.3. Analysis of non-linear pendulum dynamics

If sliding friction is neglected, the kinematics of an inverted pendulum can be characterized by the following non-linear differential equation:

$$(I + ml^2)\frac{d^2\theta}{dt^2} + \mu\frac{d\theta}{dt} = mgl\sin\theta + ml\frac{d^2x_P}{dt^2}\cos\theta \quad (1)$$

Here, the terms represent the following:

θ: Angle of the pendulum pole to the vertical axis

μ: Coefficient of viscosity

m: Mass of the pendulum

I: The Moment of Inertia (MoI) of the pendulum pole about its center of mass

l: Distance from the pivot point to the pendulum pole's center of mass (typically half the length of the pole)

x_p: Displacement of the pivot

Although exponential decay due to viscous resistance is often assumed to be the primary cause of oscillatory decay in second-order systems like the pendulum, it is known that sliding friction leads to a linear decay of oscillatory amplitude [62–65]. To account for sliding friction, we can also write

$$(I + ml^2)\frac{d^2\theta}{dt^2} + \mu\frac{d\theta}{dt} + fsign\left(\frac{d\theta}{dt}\right) = mgl\sin\theta + ml\frac{d^2x_P}{dt^2}\cos\theta \quad (2)$$

In this context, an additional friction term exists, scaled by the coefficient f, which is dependent on the sign of the angular velocity. The presence of this sign term complicates formal analysis; therefore, we initially disregard the effects of friction.

We observe that the provided kinematic description suffices for deriving control, assuming reliance solely on the cart's velocity as the control input. Additionally, it's worth noting that force control, a common approach in numerous inverted pendulum implementations [66], would necessitate an extra equation to accurately capture the dynamics of the cart's force.

Refactoring Eqn. (1) with the highest-order differential term on the left-hand side, yields:

$$\frac{d^2\theta}{dt^2} = -\frac{\mu}{(I+ml^2)}\frac{d\theta}{dt} + \frac{mgl}{(I+ml^2)}\sin\theta + \frac{ml}{(I+ml^2)}\frac{d^2x_P}{dt^2}\cos\theta \quad (3)$$

we now write:

$$\frac{d^2x_P}{dt^2} = \frac{dv_c}{dt} \quad (4)$$

We will now represent the constant terms using coefficients as follows:

$$a_1 = \frac{\mu}{(I+ml^2)} \quad (5)$$

$$a_2 = \frac{-mgl}{(I+ml^2)} \quad (6)$$

$$b_0 = \frac{ml}{(I+ml^2)} \quad (7)$$

This leads to the equation for dynamics:

$$\frac{d^2\theta}{dt^2} = -a_1\frac{d\theta}{dt} - a_2\sin\theta + b_0\frac{dv_c}{dt}\cos\theta \quad (8)$$

To express the system in state space form as two first-order differential equations, selecting the first state x_1 is straightforward since it represents the pendulum angle, denoted as θ:

$$x_1 = \theta \quad (9)$$

$$\Rightarrow \dot{x}_1 = \frac{d\theta}{dt} \quad (10)$$

We now write the second state variable x_2 as:

$$x_2 = \frac{d\theta}{dt} - b_0 v_c\cos\theta \quad (11)$$

Re-arranging Eqn. (11) gives:

$$\Rightarrow \frac{d\theta}{dt} = x_2 + b_0 v_c\cos\theta \quad (12)$$

$$\Rightarrow \dot{x}_1 = x_2 + b_0 v_c\cos x_1 \quad (13)$$

Differentiating Eqn. (12) with respect to time and using the product rule to the right-hand side terms

$$\Rightarrow \frac{d^2\theta}{dt^2} = \dot{x}_2 + b_0\frac{dv_c}{dt}\cos\theta - b_0 v_c\sin\theta \quad (14)$$

$$\Rightarrow \frac{d^2\theta}{dt^2} = \dot{x}_2 + b_0\frac{dv_c}{dt}\cos x_1 - b_0 v_c\sin x_1 \quad (15)$$

Substituting the Equations (12, 15) into Eqn. (8) and replacing angle terms with state variables

$$\Rightarrow \dot{x}_2 + b_0\frac{dv_c}{dt}\cos x_1 - b_0 v_c\sin x_1 = -a_1(x_2 + b_0 v_c\cos x_1) - a_2\sin x_1 + b_0\frac{dv_c}{dt}\cos x_1 \quad (16)$$

$$\Rightarrow \dot{x}_2 - b_0 v_c\sin x_1 = -a_1(x_2 + b_0 v_c\cos x_1) - a_2\sin x_1 \quad (17)$$

$$\Rightarrow \dot{x}_2 = -a_1(x_2 + b_0 v_c\cos x_1) - a_2\sin x_1 + b_0 v_c\sin x_1 \quad (18)$$

This leads to the following expression:

$$\Rightarrow \dot{x}_2 = -a_1 x_2 - a_2 \sin x_1 + (b_0 \sin x_1 - a_1 b_0 \cos x_1) v_c \quad (19)$$

The two equations (13) and (19) can be used in a non-linear simulation of the pendulum system.

5. Linearizing the non-linear system

5.1. Equilibrium

We now extend the mathematical analysis to derive the linearized state space model by calculating and evaluating the system's Jacobian around equilibrium positions. To linearize the nonlinear differential equation description of the pendulum around its equilibrium points, we first need to identify their locations. Equilibrium occurs when the control input is zero, and the state derivatives are also zero. That is

$$\dot{x}_1 = x_2 + b_0 v_c \cos x_1 = 0 \quad (20)$$

$$\dot{x}_2 = -a_1 x_2 - a_2 \sin x_1 + (b_0 \sin x_1 - a_1 b_0 \cos x_1) v_c = 0 \quad (21)$$

Since control velocity is zero at the equilibrium points, we see that $x_2 = 0$ and:

$$\dot{x}_2 = -a_1 x_2 - a_2 \sin x_1 = 0 \quad (22)$$

From Eqn. (22) we see that:

$$-a_2 \sin x_1 = 0 \quad (23)$$

$$\Rightarrow x_1 = \{0, \pi\} \quad (24)$$

Thus, the system has an equilibrium in an inverted configuration at 0 radians and a non-inverted configuration at π radians. To linearize the system at these equilibrium points, we need to calculate the Jacobian of the system, denoted as J_A, with respect to the system state, and evaluate it at those points. We first express the two state equations as functions:

$$f_1 = x_2 + b_0 v_c \cos x_1 \quad (25)$$

$$f_2 = -a_1 x_2 - a_2 \sin x_1 + (b_0 \sin x_1 - a_1 b_0 \cos x_1) v_c \quad (26)$$

We then calculate the partial derivatives of these two functions with respect to the state variables:

$$\frac{\partial f_1}{\partial x_1} = -b_0 v_c \sin x_1 \quad (27)$$

$$\frac{\partial f_1}{\partial x_2} = 1 \quad (28)$$

$$\frac{\partial f_2}{\partial x_1} = -a_2 \cos x_1 + (b_0 \cos x_1 + a_1 b_0 \sin x_1) v_c \quad (29)$$

$$\frac{\partial f_2}{\partial x_2} = -a_1 + (b_0 \cos x_1 + a_1 b_0 \sin x_1) v_c \quad (30)$$

5.2. Jacobian in matrix form

This leads to the Jacobian matrix:

$$J_A = \begin{bmatrix} -b_0 v_c \sin x_1 & 1 \\ -a_2 \cos x_1 + (b_0 \cos x_1 + a_1 b_0 \sin x_1) v_c & -a_1 + (b_0 \cos x_1 + a_1 b_0 \sin x_1) v_c \end{bmatrix} \quad (31)$$

We now evaluate this matrix at the equilibria points when control is zero. For the inverted configuration equilibrium at 0 radians, we have

$$J_A[x_1 = 0] = \begin{bmatrix} 0 & 1 \\ -a_2 & -a_1 \end{bmatrix} \quad (32)$$

For the non-inverted configuration equilibrium at π radians we have

$$J_A[x_1 = \pi] = \begin{bmatrix} 0 & 1 \\ a_2 & -a_1 \end{bmatrix} \quad (33)$$

We now need to linearize the control of the system. To do so, we calculate the Jacobian of the control, denoted as J_B, with respect to the control input. This involves calculating the partial derivatives of the two system functions with respect to the control input.

$$\frac{\partial f_1}{\partial v_c} = b_0 \cos x_1 \quad (34)$$

$$\frac{\partial f_2}{\partial v_c} = b_0 \sin x_1 - a_1 b_0 \cos x_1 \quad (35)$$

$$\Rightarrow J_B = \begin{bmatrix} b_0 \cos x_1 \\ b_0 \sin x_1 - a_1 b_0 \cos x_1 \end{bmatrix} \quad (36)$$

Evaluating this matrix for the equilibrium at 0 and π radians we have

$$J_B[x_1 = 0] = \begin{bmatrix} b_0 \\ -a_1 b_0 \end{bmatrix} \quad (37)$$

$$J_B[x_1 = \pi] = \begin{bmatrix} -b_0 \\ a_1 b_0 \end{bmatrix} \quad (38)$$

The linearized system in state space notation takes the form:

$$\dot{X} = AX + BU \quad (39)$$

$$Y = CX + DU \quad (40)$$

From Equations (33, 37), we can write the linearized system for the inverted configuration at 0 radians in matrix form as:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -a_2 & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_0 \\ -a_1 b_0 \end{bmatrix} v_c \quad (41)$$

Since we require the state space model to generate an output corresponding to the pendulum angle θ, its output equation is therefore:

$$Y = [1 \quad 0]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{42}$$

5.3. Linearized ODE

Multiplying out the matrix Equations (41, 42), we see that we have two linear equations. For the first state, we have:

$$\dot{x}_1 = x_2 + b_0 v_c \tag{43}$$

$$\Longrightarrow x_2 = \dot{x}_1 - b_0 v_c \tag{44}$$

$$\Longrightarrow \dot{x_2} = \ddot{x}_1 - b_0 \frac{dv_c}{dt} \tag{45}$$

For the second state, we have:

$$\dot{x}_2 = -a_2 x_1 - a_1 x_2 - a_1 b_0 v_c \tag{46}$$

$$\Longrightarrow \ddot{x}_1 - b_0 \frac{dv_c}{dt} = -a_2 x_1 - a_1 x_2 - a_1 b_0 v_c \tag{47}$$

Substituting back in x_1, $\dot{x}_1$, $\ddot{x}_1$ and x_2 from Eqn. (12)

$$\Longrightarrow \frac{d^2\theta}{dt^2} - b_0 \frac{dv_c}{dt} = -a_2\theta - a_1\left(\frac{d\theta}{dt} - b_0 v_c\right) - a_1 b_0 v_c \tag{48}$$

$$\Longrightarrow \frac{d^2\theta}{dt^2} = -a_2\theta - a_1 \frac{d\theta}{dt} + b_0 \frac{dv_c}{dt} \tag{49}$$

5.4. System eigenvalues and stability

The eigenvalues (λ) of the system matrix A represent the behavior of the pendulum system. These eigenvalues are related to the poles in the transfer function. The eigenvalues, denoted as λ, of matrix A can be determined by solving the following matrix equation, involving the calculation of the determinant, where I is the identity matrix:

$$|(A - \lambda I)| = 0 \tag{50}$$

If all eigenvalues have a negative real part, the system will be stable. Conversely, if any eigenvalue has a positive real part, the system will be unstable. It is also important to note that if the real part of an eigenvalue is zero, then the system is marginally stable, existing on the boundary of stability, neither conclusively stable nor unstable. Complex eigenvalues typically lead to oscillatory behavior, especially if they have a non-zero real part.

To incorporate feedback control into this system, the system must use a feedback mechanism. One approach involves utilizing full state feedback, as depicted in Figure 11. In this figure, the matrix K represents the gain of the state feedback, while R(t) denotes a reference input. If the reference input is zero, the state feedback can be described by the following expression:

$$U = -KX \tag{51}$$

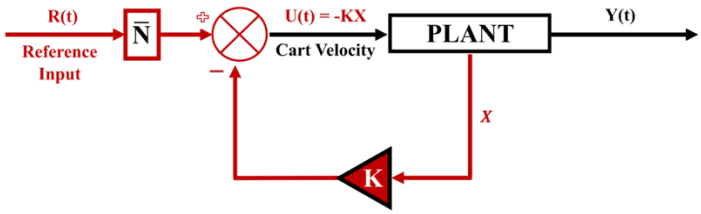


Figure 11: Signal flow graph of a plant under direct full-state feedback control: The red line delineates the feedback path, which includes multiplication by the feedback gain, denoted by K. Additionally, a feedforward gain term, represented by $\bar{N}$, is introduced to improve tracking of the reference input.

Substituting the state space system equations (39) and (40) into this expression leads to the modified state space equations, which represent the system dynamics under the influence of the feedback mechanism.

$$\dot{X} = AX + BU = (A - BK)X \tag{52}$$

$$Y = CX + DU = (C - DK)X \tag{53}$$

Implementing state feedback alters the system dynamics leading to a new expression for the state derivative. This alteration involves not just multiplying the state by matrix A, but rather by (A−BK). Consequently, the eigenvalues (λ) of the full state feedback system can be determined by solving the updated characteristic equation:

$$|(A - BK - \lambda I)| = 0 \tag{54}$$

Consequently, by modifying the gain matrix K, we can manipulate the location of the system's eigenvalues. The method for calculating K is discussed in Section 7.

5.5. Using a Luenberger observer

Many procedures in control design assume that the full state vector is available. However, this is often not the case, as in our pendulum design. In such circumstances, we can use an observer to estimate the full state using a linear plant model. The Luenberger observer computes the state estimate according to the differential equation:

$$\dot{\hat{X}} = A\hat{X} + BU + L(Y - C\hat{X}) \tag{55}$$

The observer uses the state space matrices A and B to provide a linear model of the plant. In our case, we determine the observer gains, denoted as L, using MATLAB. Similar to the state feedback gain, the Luenberger observer gain vector L must be chosen such that all the eigenvalues of the observer system, as solutions to the characteristic equation, possess appropriate negative real values. The signal flow graph for the Luenberger observer is shown in Figure 12. The system's eigenvalues satisfy the following characteristic equation:

$$|(A - LC - \lambda I)| = 0 \tag{56}$$

The calculation of L is discussed in Section 7.

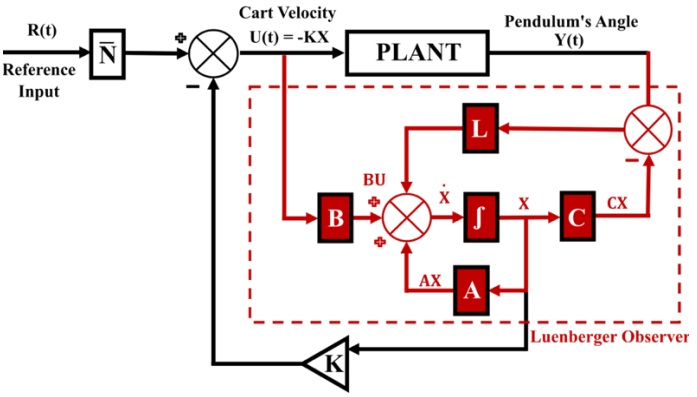


Figure 12: Estimation of plant state using a Luenberger observer. The observer, which simulates the real dynamics of the inverted pendulum, generates an estimated state, denoted as $\bar{X}$. This estimated state serves as a proxy for the actual system state, labeled X, some of which may be unobservable. The signal Y(t) is employed to correct the state estimate.

6. Augmenting the state space model

6.1. Adding cart positional state

To enable control of both the cart's position and the balancing of the pole, we introduce an additional state variable x_3 to explicitly represent the cart's position. We can relate the cart position to the control velocity input, since:

$$\dot{x}_3 = v_c \tag{57}$$

The linearized system dynamics are then represented by a 3x3 matrix, that includes the new positional state variable. The updated matrix equation is:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -a_2 & -a_1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_0 \\ -a_1 b_0 \\ 1 \end{bmatrix} v_c \tag{58}$$

The output equation remains similar to before, but with an appended coefficient of zero in the C matrix:

$$\Rightarrow Y = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{59}$$

We can compute the numeric values of the matrices using MATLAB. See Figure 13 for the udpated state feedback controller signal flow graph schematic.

6.2. Adding integral action

We can further improve cart position performance and reduce its steady-state error by adding integral action on the cart position (see Figure 14). To incorporate integral action, a state is devised within the controller to compute the integral of the positional error signal. This is then used as a feedback term, as denoted by the red path on the schematic. Therefore, to achieve integral feedback, we simply augment a state-space system by adding another state Z, whereby the state Z is the integral of the error between the desired output ref_p (representing a reference input for cart position) and actual output Y. Thus, the standard state-space equation:

$$[\dot{X}] = [AX + BU] \tag{60}$$

Becomes:

$$\begin{bmatrix} \dot{X} \\ \dot{Z} \end{bmatrix} = \begin{bmatrix} AX + BU \\ Y - \text{refp} \end{bmatrix} \tag{61}$$

Where the output is given by:

$$\text{Y} = \text{CX} + \text{DU} \tag{62}$$

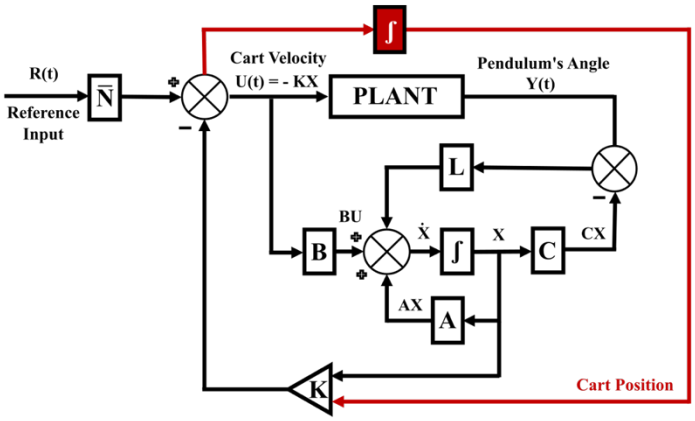


Figure 13: Adding a state for cart position provides a means to control cart position.

State feedback control is now generated from the state X and also from the state Z. That is:

$$\text{U} = -\text{KX} - \text{K}_Z\text{Z} \tag{63}$$

Thus, to add integral action to the state-space model of the cart position-augmented pendulum, and use the cart position to generate error integrated over time, we further augment the system matrices given in Eq. (58). We add a fourth state, x_4, to represent the integrated cart position error.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -a_2 & -a_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b_0 \\ -a_1 b_0 \\ 1 \\ 0 \end{bmatrix} v_c \tag{64}$$

$$\text{y} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \tag{65}$$

Notice that here we update the integral state by selecting the position state x_3 to generate the $(Y - ref_p)$ term used for integral action. For this model, we assume the reference position, ref_p, is zero.

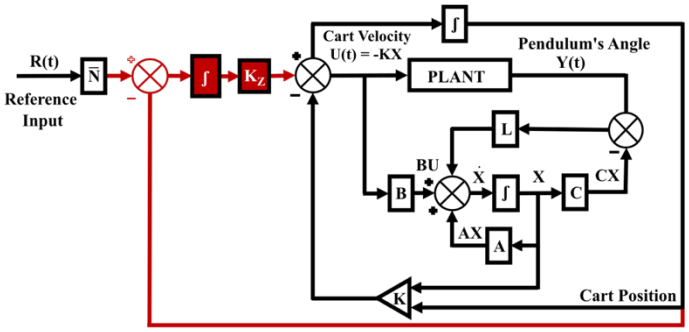


Figure 14: Illustration of integral error feedback within the system. This mechanism reduces steady-state positional error of the cart by comparing the reference setpoint with the estimated cart position, integrating the positional error, and using it in the feedback path.

7. Designing state feedback controllers

7.1. Determining feedback gain

In our pendulum design, a linear full state feedback controller is employed to balance the inverted pendulum. This method enables the maintenance of balance, even in the presence of noise and disturbances. Implementing this controller requires obtaining K, the feedback gain vector.

Various strategies can be used to find K. One method is to use pole placement, whereby we calculate K in order to achieve what we consider to be a good choice of poles for the system when it is operating under full state feedback control. Alternatively, the gain K can be found by formulating gain calculation as an optimization problem, where we specify an objective function indicative of what we consider desirable performance should be. In this work, we adopted the latter optimal control approach. Specifically, we find the gain K utilizing the MATLAB lqr command (which designs a linear quadratic regulator).

7.2. Velocity control mode

To design an optimal controller to balance the inverted pendulum using velocity as the control input, we need to consider the linear 2 state model given by the equations (41) and (42). To build an appropriate cost function for the optimization, suitable values were implemented along the leading diagonal of the 2x2 Q matrix to penalize non-zero system states. In addition, a suitable value is used in the 1x1 R matrix to penalize the control input.

Penalization of the state serves a crucial function: It ensures that the system approaches its target value. Within the scope of this design, it assists in keeping the pendulum's angle close to zero, facilitating effective balancing. In contrast, penalizing control with R serves to reduce the speed of the cart. The penalization values were determined through experimentation.

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{66}$$

$$R = 1 \tag{67}$$

7.3. Position control mode

To implement the control of the cart position as well as balancing the pendulum pole, we make use of the 4-state system that incorporates integral action. The linear state space system is described by equations (64) and (65). The diagonal entries of the 4x4 Q matrix, along with the single scalar value in the R matrix, were defined to aptly penalize state and control. The penalization values in Q and R were found by trial and error.

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 20 \end{bmatrix} \tag{68}$$

$$R = 1 \tag{69}$$

To concurrently accomplish pendulum balancing as well as control of the cart position, we applied a larger penalty on state x_4 (which represents the integral of positional error), whereas state x_3 (which represents cart position) received a penalty term of zero.

7.4. Designing the Luenberger observer

To determine the Luenberger gain L, we again employed the MATLAB lqr command. We refrained from using the observer to predict the cart's velocity or position since estimating these is straightforward, given that velocity is directly used as the control signal.

As with the determination of state feedback controller gains, the leading diagonal entries of the 2x2 Q matrix and the single value in the R matrix were selected to penalize the system states and the control action, respectively. Suitable parameter values for these matrices were ascertained through trial and error.

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{70}$$

$$R = 1 \tag{71}$$

7.5. Gain scheduling

Transitioning between velocity control of the pendulum and position control of the pendulum was realized by selecting their respective system gain K and pre-processing term $\bar{N}$ (as discussed later and presented in Table 2). We note that resetting the integral error state to zero was necessary each time the controller was switched from velocity to position mode, to ensure processing started with a zero positional error.

In position control mode, we make use of integral action. In this case, the reference position input can have a zero value to preserve the cart's current position on the track.

During the velocity control of the cart, the velocity of the cart is required to track the reference input. To ensure this takes place

the controller requires an appropriate feedforward pre-emphasis term, denoted as $\bar{N}$:

$$\bar{N} = -[\mathrm{C}(\mathrm{A} - \mathrm{BK})^{-1}\,\mathrm{B}]^{-1} \tag{72}$$

For velocity control, we note that $\bar{N}$ has a value of 1 (that is, a gain of unity). Thus, the reference input directly sets the cart velocity. This is achieved by a slight update to the calculation of the control signal, which now incorporates a non-zero reference input, referred to as 'ref':

$$\mathrm{U} = -\mathrm{KX} + \bar{N}\mathrm{ref} \tag{73}$$

8. Laplace Analysis of the Inverted Pendulum Dynamics

8.1. System transfer function

Laplace analysis can provide useful insights into system behaviour. Neglecting the non-linear effect of friction, the linearized differential equation that describes the pendulum can be expressed as:

$$(I + ml^2)\frac{d^2\theta}{dt^2} + \mu\frac{d\theta}{dt} = mgl\theta + ml\frac{d^2x_P}{dt^2} \tag{74}$$

Applying the Laplace transform, and assuming initial conditions of zero we have:

$$(I + ml^2)s^2\Phi(s) + \mu s\Phi(s) = mgl\Phi(s) + mls^2X_P(s) \tag{75}$$

$$\Longrightarrow \big((I + ml^2)s^2 + \mu s - mgl\big)\Phi(s) = mls^2X_P(s) \tag{76}$$

This leads to the s-domain transfer function relating output pole angle $\Phi(s)$ to cart position $X_P(s)$:

$$\Longrightarrow \frac{\Phi(s)}{X_P(s)} = \frac{s^2ml}{\big((I + ml^2)s^2 + \mu s - mgl\big)} \tag{77}$$

We now rearrange terms in the denominator and use the relationship $\mathrm{V}(s) = sX_P(s)$. This leads to the expression relating output pole angle $\Phi(s)$ to cart velocity $\mathrm{V}(s)$:

$$\Longrightarrow \frac{\Phi(s)}{\mathrm{V}(s)} = \frac{\dfrac{sml}{(I + ml^2)}}{\left(s^2 + s\dfrac{\mu}{(I + ml^2)} - \dfrac{mgl}{(I + ml^2)}\right)} \tag{78}$$

8.2. 2nd order canonical form

Comparing the expression with the second-order canonical form, we can identify the coefficients and characteristics of the system. This comparison allows us to further analyze the dynamics of the inverted pendulum and gain deeper insights into its behavior and control requirements.

$$\frac{sk}{(s^2 + 2\xi\omega_n s - \omega_n^2)} \Leftrightarrow \frac{\dfrac{sml}{(I + ml^2)}}{\left(s^2 + s\dfrac{\mu}{(I + ml^2)} - \dfrac{mgl}{(I + ml^2)}\right)} \tag{79}$$

Here, k represents a simple gain factor. Upon examining the given expression, we find that it allows us to determine the natural frequency of the system as follows:

$$\omega_{n=}\sqrt{\frac{mgl}{(I + ml^2)}} \tag{80}$$

Given that angular frequency (ω_n) is related to frequency (f_n) in cycles per second through the relationship $\omega_n = 2\pi f_n$ we can write the expression:

$$f_{n=}\frac{1}{2\pi}\sqrt{\frac{mgl}{(I + ml^2)}} \tag{81}$$

Similarly, by inspection, we can write down an expression for the damping ratio of the system:

$$\xi = \frac{\dfrac{\mu}{(I + ml^2)}}{2\omega_n} \tag{82}$$

9. Numeric integration to implement real-time control

9.1. Euler integration

To implement real-time state feedback control, some form of numerical integration is needed. Such integration can often be carried out satisfactorily on a digital computer using Euler's methods. Forward Euler integration works by incrementally calculating contributions to the integral that arise from the differential term.

The basic idea is as follows. Consider a function y=f(x) such that when $x=x_0$ then $y=y_0$. This is illustrated in Figure 15. As we increase the value of x by Δx we reach a point where $x_1=x_0+\Delta x$ and similarly this increases y by Δy reaching the value $y_1=y_0+\Delta y$. Therefore:

$$(x_1, y_1) = (x_0 + \Delta x,\ y_0 + \Delta y) \tag{83}$$

The gradient of the curve at (x_0, y_0) is the tangent at this point. From Figure 15, it is seen that the gradient at this point can be approximated by the ratio of a small change in y divided by a small change in x:

$$\frac{dy}{dx}|(x_0, y_0) \approx \frac{\Delta y}{\Delta x} \tag{84}$$

This is only strictly true in the limit where Δx tends to zero. In practical numerical methods, this limit is approximated by choosing a sufficiently small Δx. We also see that we can use this relationship to iteratively estimate y_1 from y_0 by replacing the Δy term by two very close and successive y values:

$$\frac{dy}{dx}|(x_0, y_0) \approx \frac{(y_1 - y_0)}{\Delta x} \tag{85}$$

Re-arranging this equation and writing Δx as step size *h* gives:

$$y_1 = y_0 + h\frac{dy}{dx}|(x_0, y_0) \quad (86)$$

Writing the x-axis in terms of a time variable t, the gradient is given by:

$$\frac{dy}{dt} = f(t, y) \quad (87)$$

And initial conditions are given by:

$$f(t_0, y) = y_0 \quad (88)$$

We then obtain the recurrence relation for step n:

$$y_{n+1} = y_n + hf(t_n, y_n) \quad (89)$$

Where t is time, at initial time to the output is y_0, at future time $t_{(n+1)}$ the output is $y_{(n+1)}$, and h is the step size. This expression can be used to iteratively find the next estimate of y_1 if we know x_o and the gradient at (x_0, y_0). This approach provides a method for integrating the differential term, which is generally satisfactory if a sufficiently small temporal step size, h, is used. This method is easily extended to vector form to perform the integration stage needed in state space models.

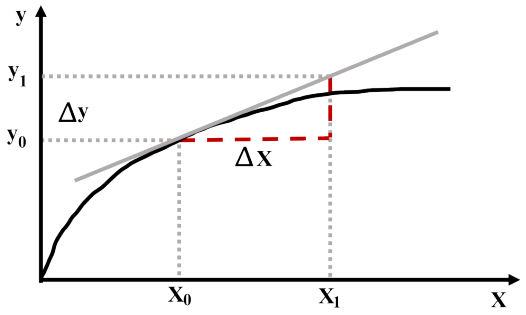


Figure 15: The gradient dy/dx of a curve y=f(x) can be locally approximated at the point (x_0,y_0) as the ratio of a small change in the value of y to a corresponding small change in the value of x.

9.2. Higher order numerical integration

Euler integration is the simplest fixed-step numerical method that can be adopted. However, other more complex integration rules can also be used. These include the midpoint, trapezoidal, and Runge-Kutta methods, which, though requiring more computational steps in the estimation of the integral, offer higher accuracy. Additional methods utilize dynamic selection of step size, such as the ode45 function in MATLAB. See [67] for a discussion of these methods.

Here, we use MATLAB's ode45 for simulations of the uncontrolled stable pendulum configuration. We use Euler Integration to model the controlled pendulum because of the method's simplicity and its ease of implementation on a microcontroller, especially considering its low computational requirements.

10. System identification

10.1. Large angle pendulum oscillatory behavior

Approximately estimating observable parameters of a pendulum, such as length and weight, can be done with relative ease. However, assessing other parameters is considerably more challenging, and in some cases, impossible, solely based on observations of the static mechanical system. To accurately determine values for viscous and sliding friction, it is necessary to conduct measurements during pendulum movement.

To examine the large-angle oscillatory behavior of the pendulum, we raised it from its resting, vertically hanging (non-inverted) position to a horizontal alignment, corresponding to an angle of approximately 90°, before releasing it. The pendulum then oscillated until viscous damping and friction gradually brought it to a standstill in its vertical, non-inverted position.

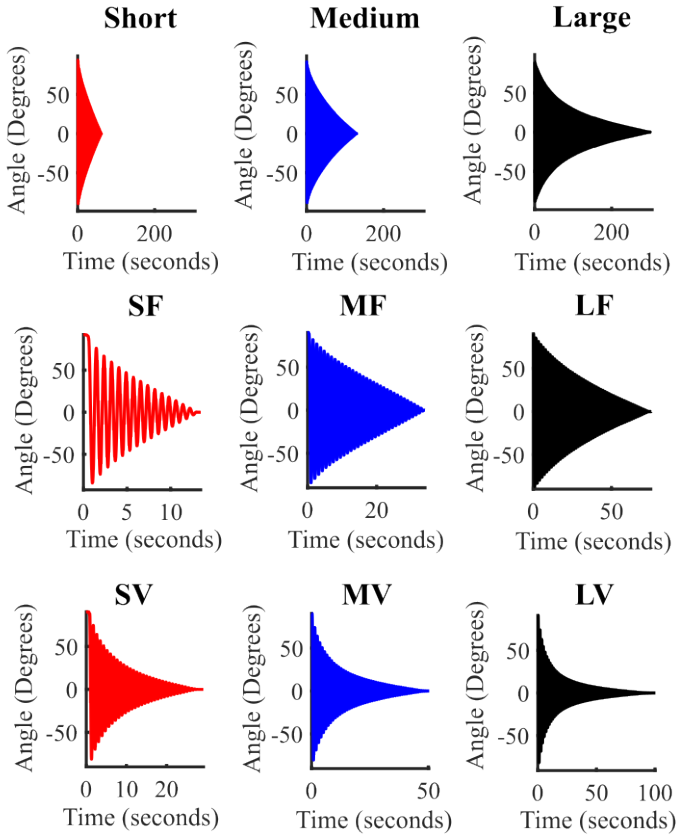


Figure 16: Pendulum large angle oscillation decay over time for three pendulum lengths. Data captured using an incremental encoder. Top row: No added viscosity or friction for short, medium, and long pole lengths. Middle row: Effect of added friction for short, medium, and long pole lengths (SF, MF, and LF, respectively). Lower row: Effect of added viscosity for short, medium, and long pole lengths (SV, MV, and LV, respectively).

10.2. Data logging

To examine the pendulum's oscillation decay over time, we collected time-stamped pole angle data from its shaft encoder as the pendulum swung. In addition, the pendulum cart position was recorded using readings from the encoder mounted on the rear of the cart drive stepper motor. The data were gathered using a program running on an Arduino Mega, which transmitted the time and angular measurements to a host PC equipped with Microsoft Excel. The Excel program was used to record the data at a 25Hz rate and save it to the hard disk in Excel file format. Subsequently, the data were imported into MATLAB for analysis. This allowed

for the generation of a plot depicting the pendulum's decaying oscillations under various conditions, as well as further analyses.

The mechanical pendulum system features three interchangeable poles, and for each pole, the level of friction could be adjusted from low to high. Similarly, the viscous damping could be altered from low to high by manipulating the wind resistance experienced by the paddle mechanism. This configuration led to a total of 9 different experimental conditions.

Figure 16 top row shows the temporal response waveforms for the undamped cases with no added friction for all three pendulum lengths, plotted on the same scale. It is observed that a longer pendulum length significantly increases the time required for the pendulum angle to decay to zero. Figure 16 middle and lower rows illustrate the responses of pendulums of three different lengths with additional friction and additional viscosity introduced, respectively. It is apparent that incorporating viscosity into the system accentuates the exponential decay. However, it is noteworthy that when friction is the dominant factor, the decay is linear rather than exponential.

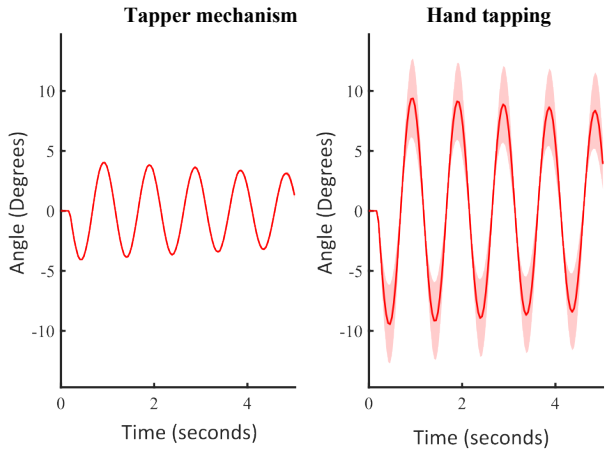


Figure 17: Testing the consistency of the tapping mechanism. Mean and standard deviation of the pendulum angular response averaged over 8 trials are shown using the tapper mechanism and hand excitations of the pendulum pole.

10.3. Small angle pendulum oscillatory behavior

The primary interest of this study is the examination of the balancing behavior of the pendulum system in its inverted configuration; therefore, large angle behavior is not of particular relevance. In a balancing configuration, the pendulum pole is maintained close to its unstable equilibrium position by the controller. In this case, the angular deviation from the 0° position is small, which is also essential for the validity of the linear approximation made in the observer model. Therefore, to estimate the parameters of the linear model accurately, it is necessary to examine the pendulum operation at small angles of deflection and to perform system identification for all pendulum parameters under these conditions. To generate consistent excitation to the pendulum, we utilized a mechanical RC servo tapping mechanism.

11. Using the tapping mechanism

11.1. Evaluating tapper consistency

To evaluate the consistency of the tapping mechanism's operation, we conducted tests on the medium pendulum pole without added friction or viscosity. Figure 17 illustrates that the tapper provides very consistent excitation of the pendulum, particularly when compared to the variability typically observed with manual tapping by hand.

11.2. Excitation of non-inverted pendulums

We examined the pendulums in their normal, stable, hanging-down mode to characterize the effects of the tapping. Figure 18 illustrate the responses of pendulums of different lengths driven by the tapper mechanism, both with and without added friction and viscosity. In comparison to the large angle oscillation tests, it is noteworthy that at small angles, the paddle has only a minor effect, and additional friction more rapidly damps out pendulum oscillation.

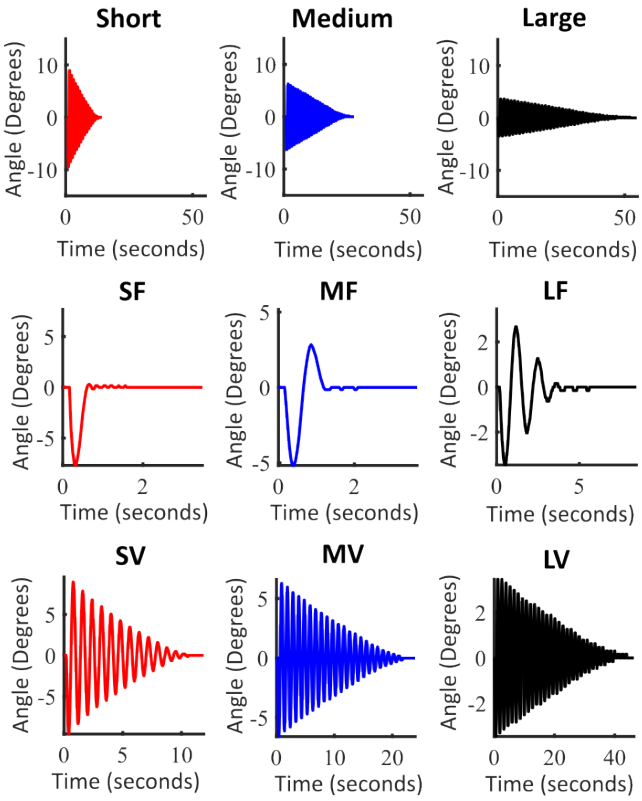


Figure 18: Pendulum small angle oscillation decay over time. Top row: No added viscosity or friction. Middle row: Effect of added friction for short, medium, and long pole lengths (SF, MF, and LF, respectively). Lower row: Effect of added viscosity for short, medium, and long pole lengths (SV, MV, and LV, respectively).

11.3. Estimating pendulum parameters

We performed grey-box system identification of the physical pendulum mechanism to identify parameters of viscous and static friction, and to fine-tune others, including pendulum length, weight, and moment of inertia.

To fit the small angle pendulum response data, we focused on six of the nine configurations: the three pole lengths both with and without added viscosity. This fitting was accomplished with a simulation of its nonlinear dynamics, employing an optimization procedure using the MATLAB fmincon function. We discarded the configurations with extra friction due to the dramatic impact it had on the system's behavior, which resulted in a limited amount of useful temporal data. This procedure optimized the mass,

effective pole length, pole moment of inertia, sliding friction, and viscous friction parameters of the pendulum system.

We designed an objective function that minimized the sum of squared distances between the predicted oscillations and the measured data. To align the simulation with the measured data for comparison, we first trimmed the measured data to begin at its first positive peak in the pendulum's oscillation. The pendulum's angular velocity at this point is zero, and its corresponding angle was used to set the initial angular state in the non-linear simulation of the pendulum, based on Eqn. (2), that includes both viscosity and friction. This was carried out using the MATLAB ode45 solve.

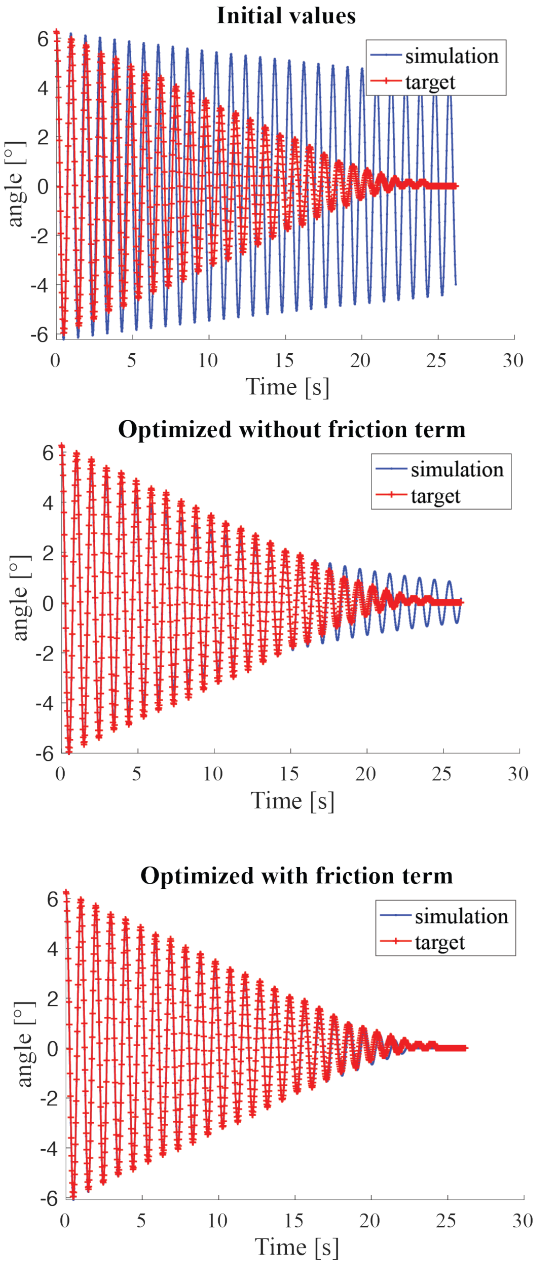


Figure 19: Running optimization to fit the measured response of the medium-length non-inverted pendulum without additional viscosity or friction. Top plot shows the predicted response based on an initial rough guess. Middle plot shows the estimated response after running the optimization algorithm without including the friction term. Lower plot shows the estimated response when the friction term is present. It is evident that accounting for friction leads to a significantly better fit.

We initialized the simulation parameters based on direct measurements of pendulum pole length and mass. Initially, we estimated the values of the friction and viscosity parameters through a process of trial and error, which was aided by careful observation of the simulated responses. During the fitting procedure, we allowed the optimization algorithm to refine all parameter values. However, it was necessary to constrain the parameter solutions to prevent fits that deviated substantially from the known ground truth values for mass and pendulum length. To this end, the mass and length were constrained to values between 0.9 and 1.1 times their measured values. The other parameters, for which we had less grounded certainty, were allowed to vary from 0.1 to 10 times their initial estimated values.

Table 1: Physical Measurements of the three different pendulums and estimated values found by system identification. Values marked with 'F' represent estimates with friction included in the second-order nonlinear differential equation model of the pendulum. The bold values represent estimates made when only a viscous damping term is present.

	Short Pendulum		Medium Pendulum		Long Pendulum	
	Normal	Viscous	Normal	Viscous	Normal	Viscous
Measured Length to CoG [m]	0.233/2 = **0.117**		0.335/2 = **0.168**		0.635/2 = **0.318**	
Measured Weight [Kg]	**0.174**		**0.226**		**0.336**	
Estimated Half-length to CoG [m]	0.116(F) **0.106**	0.117(F) **0.105**	0.167(F) **0.149**	0.164(F) **0.149**	0.322(F) **0.317**	0.308(F) **0.316**
Estimated Weight [Kg]	0.178(F) **0.158**	0.174(F) **0.157**	0.207(F) **0.207**	0.202(F) **0.207**	0.341(F) **0.338**	0.327(F)x **0.338**
Estimated MoI [Kg-m2]	8.97e-04(F) **9.02e-04**	8.47e-04(F) **8.83e-04**	0.0024v **0.0027**	0.0024(F) **0.0027**	0.0142(F) **0.0144**	0.0147(F) **0.0146**
Estimated Viscosity [N-m-s/rad]	2.21e-04(F) **9.90e-04**	2.22e-04(F) **1.00e-03**	3.36e-04(F) **1.00e-04**	2.35e-04(F) **1.00e-04**	2.22e-04(F) **0.0038**	5.20e-04(F) **0.0048**
Estimated Friction [N/Rads-1]	5.05e-04(F) **n/a**	6.16e-04(F) **n/a**	3.07e-04(F) **n/a**	3.87e-04(F) **n/a**	3.87e-04(F) **n/a**	5.46e-04(F) **n/a**

Figure 19 shows the results of using system identification to fit model parameters to the measured data for the medium-length, low-friction, low-viscosity pendulum condition. The initial guess, which incorporated insufficient damping, is shown in Figure 19 top plot. A reasonably good fit is achieved by fitting the model with only viscous damping, as demonstrated in Figure 19 middle

plot. An even better fit is obtained when the model also includes sliding friction, depicted in Figure 19 1ower plot.

For each pendulum length, measured and estimated values were obtained, both with and without friction, leading to a total of six conditions. These conditions are presented in Table 1. The corresponding (no-friction) state space matrices and Luenberger gain L are shown in Table 2. The computed state feedback controller system gains K and reference pre-scaling factor $\bar{N}$, are shown in Table 3.

11.4. Reality check using canonical form

To provide a ballpark estimate of the main parameters of the three pendulums in their low-damping modes, each pendulum was first nudged using the tapper mechanism and then allowed to sway freely, eventually settling into its stable position, as depicted in Figure 20. Measured values are shown in Table 4.

Table 2: Values of the state-space matrix and Luenberger observer gain, presented in MATLAB syntax.

Parameter	Value
A	[0 1 0 0; 41.4440 −0.1692 0 0; 0 0 0 0; 0 0 1 0;]
B	[44.2247; −0.7147; 1; 0;]
C	[1 0 0 0]
L	[12.7461 80.7319]

Table 3: Computed state feedback control K gains and $\bar{N}$ values used for the two control mode (in MATLAB syntax).

Control 1zModes	K Gain	$\bar{N}$
Position Control	[5.0514 0.7587; −4.4412; −4.4721;]	*0*
Velocity Control	[3.2478; 0.4734; 0; 0;]	*1*

Table 4: Measured decay and nearest integer number of cycles as a function of time for all three pendulum lengths in undamped conditions. The values in brackets were found by the system identification procedure.

	Short Pendulum	Medium Pendulum	Long Pendulum
Measured Maximum Time [s]	13.50	27.0	54.8
Measured 50% Decay Time [s]	5.99	12.15	26.2
Measured 50% Decay Cycles [integer count]	7	12	19
Observable frequency of oscillation f_d [Hz]	1.17 Hz **(1.25 Hz)**	0.99 Hz **(1.03)**	0.73 Hz **(0.74 Hz)**
Estimated damping ratio ζ	0.016 **(0.024)**	0.0091 **(0.011)**	0.0057 **(0.0084)**

We used Equations (80 and 81) to calculate the corresponding natural frequencies, which were essentially the same as the damped frequencies due to the very small damping ratio. These values aligned well (within 10%) with those obtained through system identification, confirming the appropriateness of the values found by the fitting procedure. We calculated the damping ratio from the decay to 50% in a time t_{50} using the equation:

$$\zeta = \frac{-\ln(0.5)}{\omega_n t_{50}} \tag{90}$$

From Table 4 it can be seen that the system identification yielded similar values.

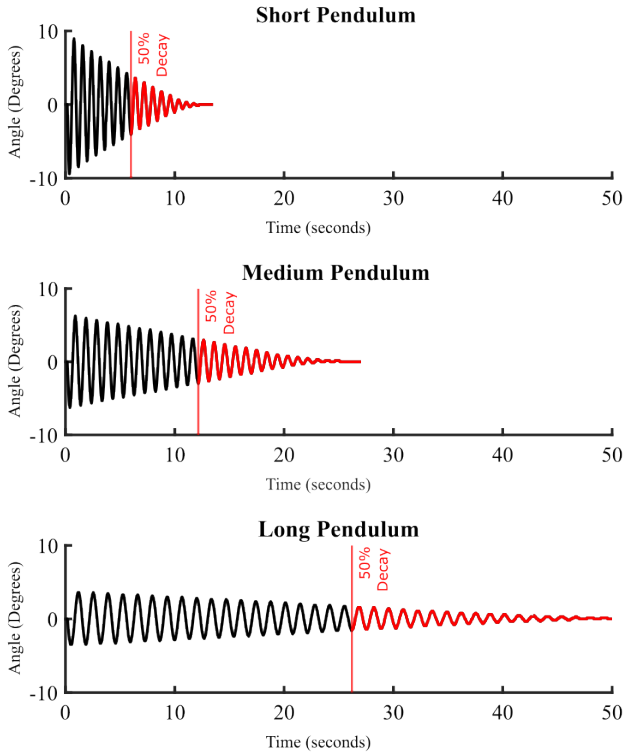


Figure 20: Approximate estimation of the canonical parameters for the pendulum system by observing the frequency of oscillation and the amplitude decay to 50% of the initial value. Plots are shown here for the short, medium, and long-length pendulums (with no extra damping), respectively.

12. MATLAB SFC controller simulations

12.1. Simulating the pendulum

In MATLAB, we simulated the pendulum balancing using both velocity and position controllers. This involved the deployment of a non-linear state space model to simulate the pendulum plant. Additionally, a linear model was implemented for Luenberger observer full state feedback control.

Following initial design of the two inverted pendulum controllers, they underwent iterative tuning and testing within MATLAB simulations. This methodology included assessing the behavior and stability of the inverted pendulum in response to cart movement and simulated impacts to the pole while it was balanced in its inverted position. Simulations carried out in both velocity control and position control modes, also involve modifying the reference input to maneuver the cart velocity or position.

12.2. Reference tracking

Simulated results of velocity tracking are shown in Figure 21 left column. The cart nearly achieved the reference velocity using

a feedforward velocity command applied directly to it. In this scenario, feedback control of the cart's velocity was not utilized. As a result, the tracking was not entirely accurate; this is partly because the feedback control required for stabilizing the balance tends to counteract the cart's movement.

Simulated results of position tracking are presented in Figure 21 right column. Utilizing integral action to correct the cart's position error proved effective for tracking the desired cart position. However, the response to the target location was not immediate, with a noticeable delay before the cart aligned with the desired position.

12.3. Recovery from impulsive disturbance

To simulate kick disturbances, we manipulated the state of the pendulum to mimic the effect of an elastic collision with another hard object. This involved setting the pendulum's angular state (x_1) to zero and its second state (x_2), which comprises an angular velocity component and a control input term, to a small positive value. The results of this simulation are shown in Figure 22.

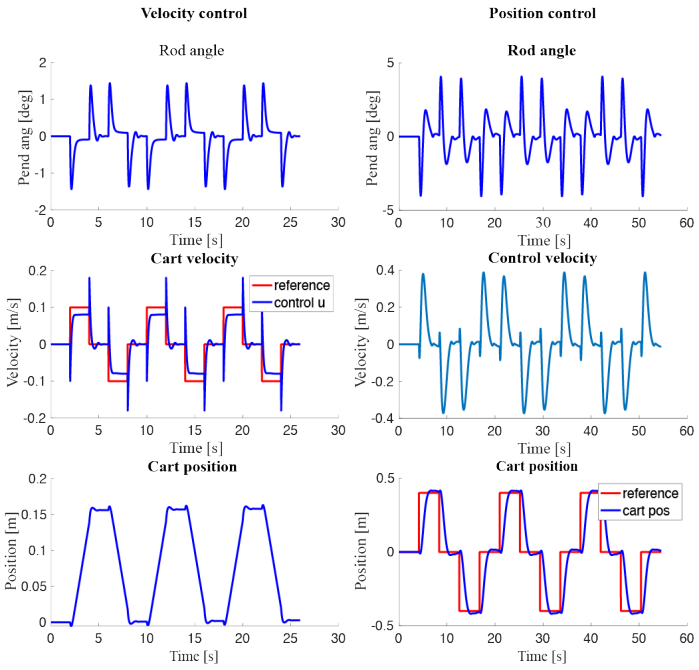


Figure 21: The pendulum angle and cart position are depicted in response to the cart movement, driven by changes in the cart position reference setpoint. Left column: MATLAB simulation of the pendulum with velocity control. The pendulum angle and cart position are shown in response to moving the cart, driven by changes in the velocity reference input. Right column: MATLAB simulation of the pendulum under position control of the cart with integral action on cart positional error.

The left column of Figure 22 shows the response of the inverted pendulum to an impulsive disturbance while operating in velocity control mode. It demonstrates a successful recovery to the initial pole deviation from the vertical position. However, it is notable that the cart position shifts and does not return to its initial starting position.

The right column of Figure 22 shows the behavior of the inverted pendulum to the same disturbance while operating in position control mode. Again it can be seen that a good recovery to the initial pole deviation from the vertical orientation is achieved. However, although the cart initially drives away from its initial location, is slowly moves back after a few seconds.

13. DIN Rail Panel Construction

13.1. Connections to controller

A DIN rail panel was tailor-made to control the inverted pendulum system. Figures 23 and 24 show the connectors and their corresponding wiring diagrams. The controller interface, located at the top of the panel enclosure, features a female USB-B port. This port connects the Arduino Mega 2560, located inside the controller assembly, to an external computer, which is used for software development and relaying control commands.

The panel features seven female D-connectors. The interfaces are configured to connect with four stepper motors, an I²C device, an SPI device, and two encoders. The use of these D-connectors allows for easy and quick connection of the panel to the inverted pendulum apparatus or other equipment.

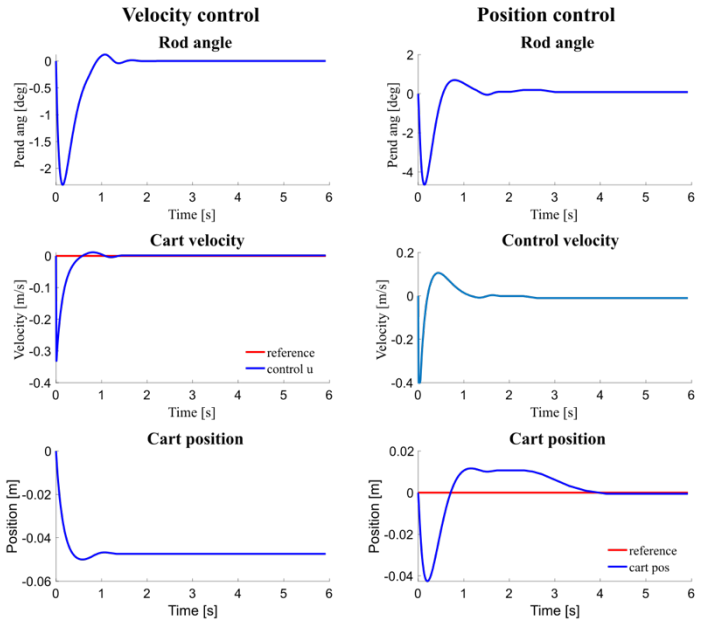


Figure 22: The pendulum angle and cart position are depicted in response to a simulated tap. Left column: MATLAB simulation of the pendulum with velocity control. The pendulum angle and cart position are shown in response to a 40 deg/s initial velocity disturbance. Right column: MATLAB simulation of the pendulum under position control of the cart, with integral action on cart positional error.

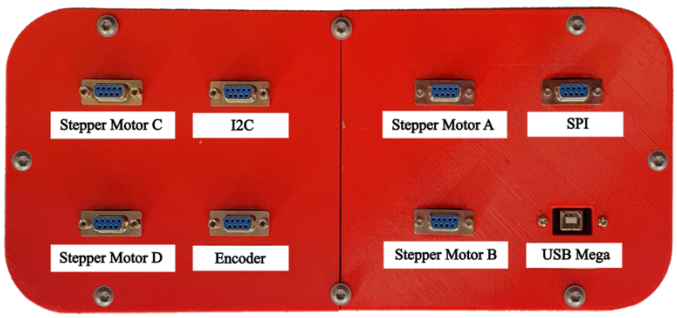


Figure 23: View of the actual controller panel cover, which was designed in Autodesk Fusion 360 and then 3D printed.

Special care was taken in the wiring of the plugs for each attached component, and they were configured with a unique connection pattern. This design feature serves to prevent accidental inappropriate connections that could result in electrical damage. The primary focus was on the power pin assignments, ensuring their isolation from other signal inputs and outputs. This safeguard ensures that potential connection errors will not result in component damage.

13.2. Panel internal layout

The controller panel's internal layout comprises three specific DIN rails, as depicted in Figure 25, which shows a photograph of the assembled unit. It consists of a microcontroller rail, a rail for motor drivers and power converters, and a power supply rail. Additionally, wiring conduit was strategically placed to manage the cabling in a tidy manner. This arrangement not only enhanced the orderly appearance but also reduced the risk of unintended cable contact with other components or causing interference.

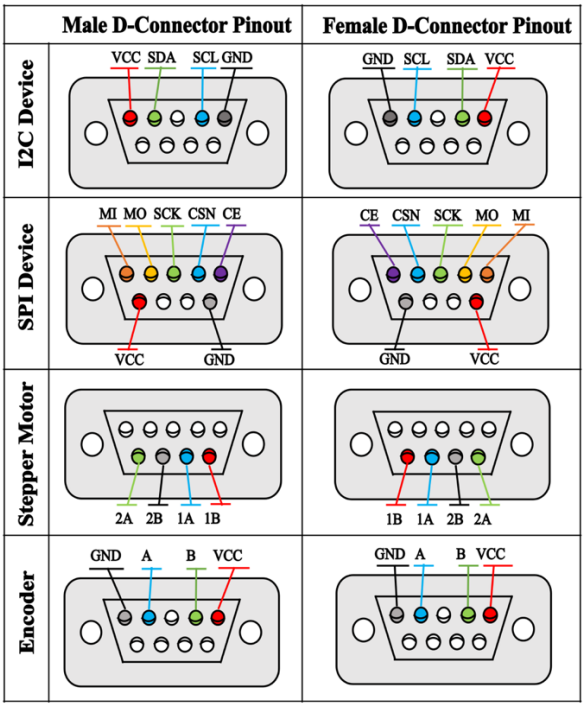


Figure 24: Rear view of soldered D-Sub connections on the male plugs (left-hand column) and the female sockets (right-hand column) mounted on the panel. The top row shows the Inter-Integrated Circuit (I²C) pinout, the second row is for Serial Peripheral Interface (SPI), the third row is dedicated to stepper motors, and the bottom row is configured for encoders

13.3. Panel DIN rail components

In the selection of panel components, preference was given to components equipped with rear mounts for DIN rail attachment. For other components, custom 3D printed support structures were made and fitted with clips to attach them to the DIN rails. The power rail within the controller panel was compartmentalized into three principal sections:

1. An AC circuit breaker section that is responsible for regulating overcurrent conditions and interrupting the main supply.
2. A region featuring a PULS Dimension DIN rail power supply, which delivers 24 volts at 5A, and meets the voltage and current prerequisites of the system.
3. A region containing three discrete DC circuit breakers, to permit individual disconnection of each control component. This feature provides additional protection against potential damage due to short circuits.

13.4. Stepper motor drivers

The power converter and motor driver rails were engineered to facilitate the operation of up to four stepper motors, each capable of being directly interfaced with the control panel's motor driver rail. The motor driver rail featured two 3D-printed carriers, each housing a pair of A4988 stepper motor drivers. These motor drivers were powered by a 24-volt supply operating on mains power.

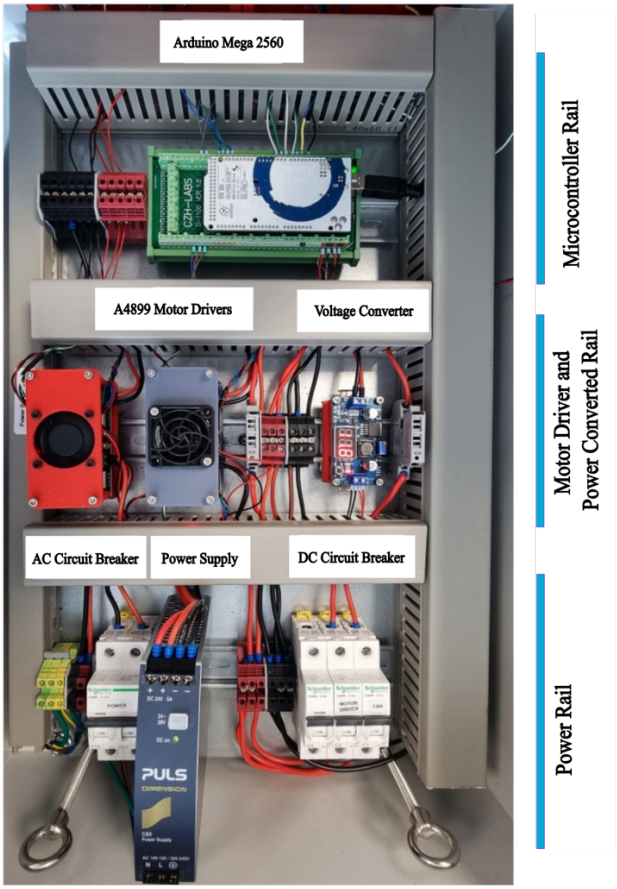


Figure 25: View displaying the internal layout of the controller panel, which is partitioned into distinct sections for power management, motor control, and microcontroller functions.

The A4988 driver offers eight distinct micro-stepping resolutions, providing flexibility in stepper motor control. For this project, we employed a one-fourth step resolution for all motors. To realize the one-fourth step resolution setting on the expansion board, the DIL switches were configured appropriately. The A4988 driver can operate within a voltage range of 8V to 35V and can deliver a current ranging from 1.5A to 2.2A. A potentiometer is incorporated into the driver breakout board, enabling hands-on fine-tuning of the current supplied to the stepper motor. To

determine the current limit, I_{lim}, it's essential to gauge the reference voltage (V_{ref}) on the potentiometer and apply the given formula to compute its desired value:

$$I_{lim} = 2V_{ref} \tag{90}$$

This facilitates precise control over the motor current, optimizing performance and efficiency. Our adjustment of the current limit was a critical step to ensure the proper functioning of our NEMA23 stepper motors. A current limit of 1.5A was determined to be ideal for this specific application. Exceeding this current threshold may damage both the motors and the drivers. Conversely, setting the current below 1.5A could compromise the motors' operation, raising the likelihood of stalling. Thus, appropriate calibration is paramount to ensure good performance while safeguarding the components from potential harm.

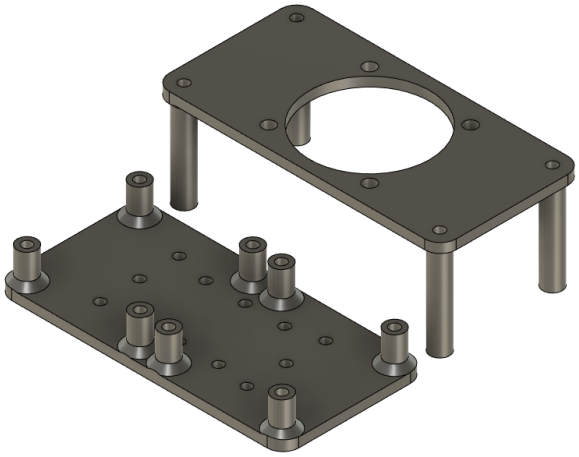


Figure 26: DIN Rail holder for the A4988 Driver Module. This setup mounts two stepper motor controller breakout boards onto a DIN rail. Notably, it includes a cooling fan to ensure the controllers remain at optimal temperatures and prevent overheating.

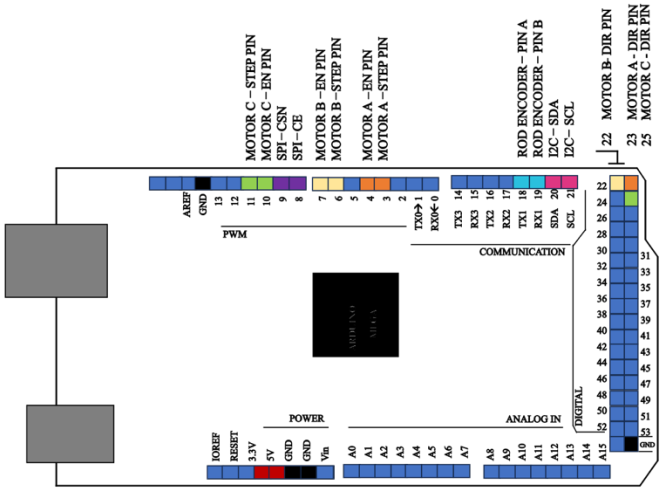


Figure 27: Pinouts for the Arduino Mega 2560. The connections are configured to control up to three stepper motors and communicate with peripherals via the I²C and SPI data buses.

To prevent overheating of the motor drivers, the incorporation of a cooling solution was essential. A pair of compact DC axial fans were positioned above them to prevent thermal damage to the chips. These fans required a supply voltage of 12V, which was furnished through a Buck converter operating from the 24V power line (refer to Figure 26). This setup ensured that the motor drivers stayed at a safe temperature during operation, enhancing both their performance and longevity.

At the heart of the controller panel, on the microcontroller rail, the Arduino Mega 2560 is prominently featured. Selected for its robust I/O capabilities, it offers 54 digital input/output pins, including 15 suitable for PWM, 256KB of flash memory, and a clock speed of 16MHz (refer to Figures 27 and 28). The Arduino Mega facilitates seamless communication with the host PC, aligning perfectly with the project's requirements.

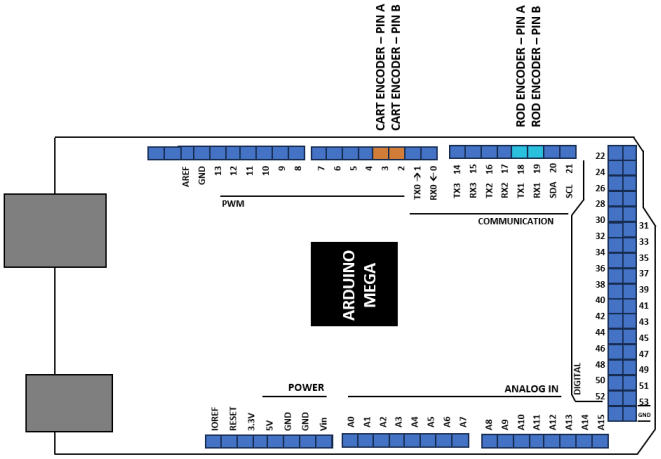


Figure 28: Arduino Mega 2560 pinout connections used for reading the system's two encoders.

14. Software Implementation in Arduino

14.1. Arduino software overview

Control over the pendulum was overseen by the Arduino Mega, which was employed to implement a state feedback controller function within its primary polling loop. During the polling operation, the control process begins by reading the encoder to ascertain the pendulum pole's angle relative to the vertical. For the state feedback controller to operate, it required the pendulum system's full state. Since only the angle was directly available, the second state was estimated using a Luenberger observer. Using the full state estimate, the control command was generated by multiplying it with the specified feedback gain. Based on the measured angle, the feedback controller then calculated the stepper motor velocity control signal. This signal produced an output pulse train, driving the stepper motor and allowing the cart to move at the ideal speed to maintain balance. State updates were computed utilizing Euler integration.

14.2. Arduino command menu

Within the polling cycle, the system processes and interprets incoming commands received through the serial interface. The suite of accessible commands encompasses the following:

- Activate control: Starts balancing when the pendulum is brought up to an inverted configuration.
- Deactivate control.
- Display help menu: View controller parameters.

14.3. Main Arduino loop pseudocode

The pseudocode for the main Arduino poll loop is shown below. This makes use of function calls the can reset the encoder, calculate the motor command, and drive the stepper motor. The loop is continuously run to control the system, ensuring that the pendulum maintains its desired position or follows a certain trajectory. In addition, there are command responses from calls to the menu system so that the program can be operated from a serial monitor over a USB connection.

Main Arduino Poll loop

Result: Balances pendulum on track

Initialization of SFC parameters and flags

while *program running* **do**

Call the menu object for input commands, act accordingly

Read time

Read pendulum pole angle

Read reference value

Call SFC function to compute control u

Generate stepper control pulses to

Drive stepper motor

end

14.4. Pseudocode to implement SFC

The state feedback controller is implemented as a C++ class. Its constructor sets up the parameters for the state space model, as well as the SFC gain K and the Luenberger gain L. A SFC function is called with the current time and the measured pendulum pole position. It returns the control value U, which is subsequently used to set the rotational velocity of the stepper motor and drive the pendulum cart.

State Feedback Controller

Result: state feedback control variable u

Initialization state space matrices A, B, C
Initialization SFC gain K
Initialization Luenberger gain L
Initialization state estimate; $\hat{X}$ = [0; 0; 0; 0];

while *balancing pendulum* **do**

Read reference value ref

Read pendulum output angle y

Calculate time step: h. = time - lastTime

Update last time: lastTime = time

Calculate control: u = - K$\hat{X}$ + ref * $\bar{N}$

Calculate Output prediction error: yErr = y – C$\hat{X}$

Calculate 1st pendulum state derivative:
$\dot{\hat{X}}(0)$ = A[0][0] * $\hat{X}$[0] + A[0][1] * $\hat{X}$[1] + B[0] * u + L[0] * yErr

Calculate 2nd pendulum state derivative:
$\dot{\hat{X}}(1)$ = A[1][0] * $\hat{X}$[0] + A[1][1] * $\hat{X}$[1] + B[1] * u + L[1] * yErr

Calculate 3rd cart position state derivative:
$\dot{\hat{X}}(2)$ = B[2] * u

Calculate 4th integral position error state derivative:
$\dot{\hat{X}}(3)$ = $\hat{X}(2)$ = B[2] – ref

Perform Euler integration:
$\hat{X}$ = $\hat{X}$ + h$\dot{\hat{X}}$

Return u

end

15. Experimental results

15.1. Online demonstration videos

Viewers can watch the inverted pendulum operations, including various tests, on the YouTube channel 'Robotics, Control and Machine Learning'. All related videos are grouped under the 'ASTESJ Inverted Pendulum' playlist. Please follow the provided link for direct access to these videos:

https://youtu.be/pvF0Zhs501U?si=P7vknbwL11tvBksE

15.2. Perturbation tests

We use a straightforward approach to assess control law performance by simultaneously monitoring the cart's position and the pendulum's angle during minor system disturbances. We obtain these measurements using encoders on the cart's stepper motor and at the pendulum's axis for angle measurement. During the balance test, disturbances are introduced by systematically applying an impulse to the pendulum using a tapping mechanism.

Experiments were conducted in both position and velocity control modes, varying the pole length from shorter to longer dimensions. These tests demonstrated the system's ability to maintain stability, even when subjected to additional loads altering its dynamic characteristics. Video demonstrations of these experiments can be found on YouTube, as per the link provided above.

15.3. Position mode results

Figure 29 shows the results of the test conducted in position control mode. The pendulum was initially stabilized and held stationary in a balanced position. Following the disturbance, immediate and noticeable changes occurred in both the pendulum angle and the cart's position. The pendulum angle plot showed minor fluctuations around zero. The motor encoder data revealed that the cart moved quickly to restore balance in the system and then gradually returned to its original position. This behavior, referring to the cart's movement and pendulum stabilization, is also evident in the YouTube videos.

During tests with the long-length pole, we observed a small oscillation of the cart, even in the absence of disturbance. This oscillation likely stems from a significant mismatch between the long-length pole's parameters and those of the medium-length pole, for which the controller was originally designed. Based on these observations, modifying the control law appears necessary to effectively manage the pendulum's behavior with its increased length.

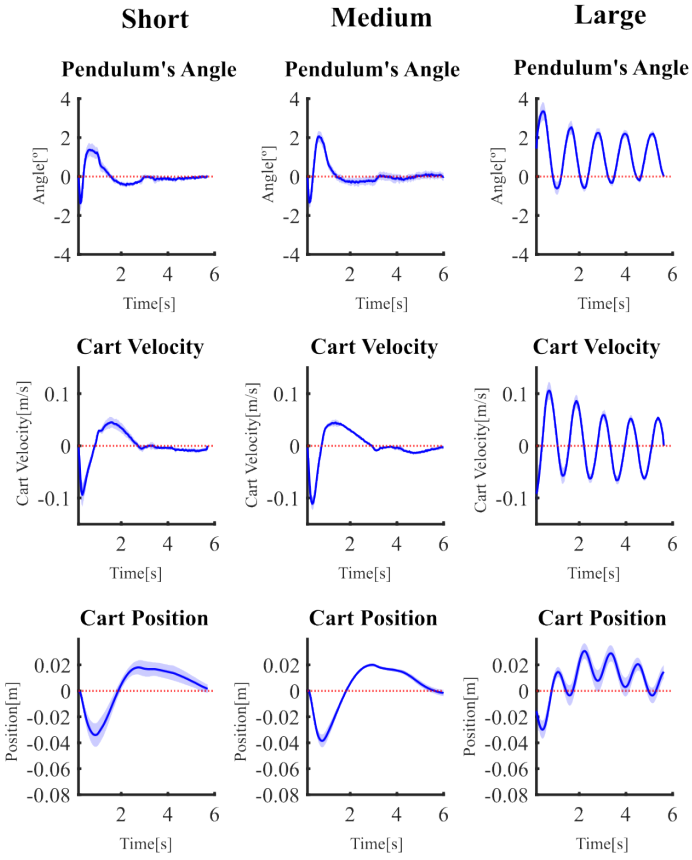


Figure 29: Responses of real, physically position-controlled inverted pendulums (using integral action on cart position) to impulsive disturbances delivered by the tapping mechanism. Results are for all three pendulum lengths in their low damping configurations. The solid blue line represents the mean response averaged across eight aligned recordings, and the light blue shading indicates the corresponding standard deviation.

15.4. Velocity mode results

The pendulum's performance was closely examined while operating under velocity control, as illustrated in Figure 30. The velocity control operation was initiated via the keyboard on the main PC. Users could select a cart velocity, resulting in the cart moving as specified in both left and right directions.

To further evaluate the system's resilience against knocks, representing impulsive disturbances, we conducted an additional test. In this test, the pole was gently tapped while the velocity was set to zero. In all instances, the pendulum system demonstrated notable stability and effectiveness in mitigating the disturbance. Compared to the position control mode, disturbances in the velocity control mode led to uncompensated cart movements, particularly for the short pendulum, as shown in the lowest row of Figure 30. Again, this behavior is also apparent in the YouTube videos. This is also worth noting that using a longer pole again resulted in a modest reduction in performance in velocity control mode, with some cart oscillations observed after the impulsive disturbance. However, this effect was less pronounced than in position control mode.

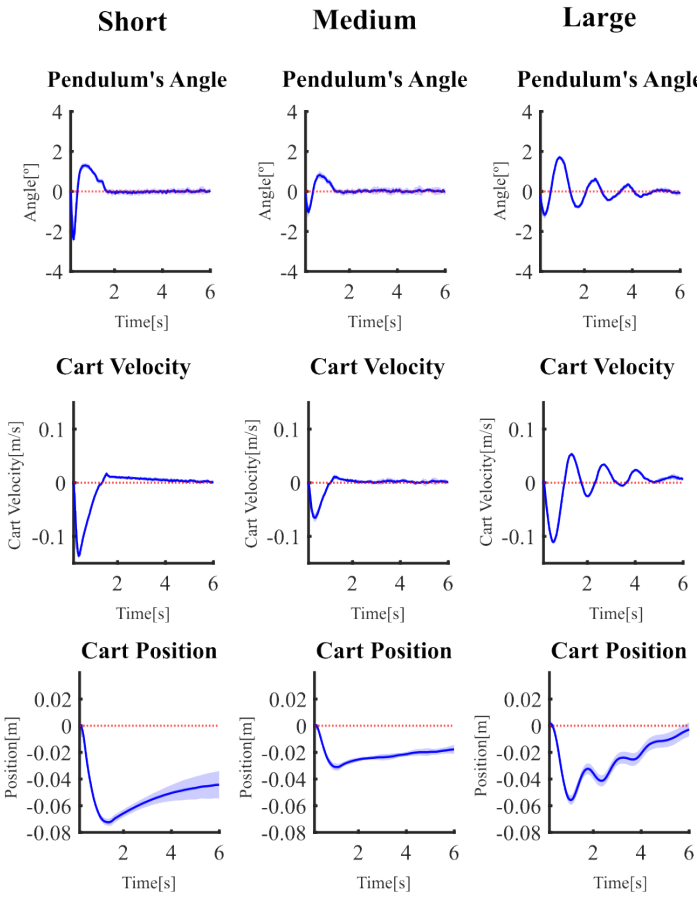


Figure 30: Responses of real, physically velocity-controlled inverted pendulums to impulsive disturbances delivered by the tapping mechanism. Results are for all three pendulum lengths in their low damping configurations. The solid blue line represents the mean response averaged across eight aligned recordings, and the light blue shading indicates the corresponding standard deviation.

16. Discussion

16.1. Summary

In this study, we demonstrated the design, implementation, analysis, simulation, and testing of an inverted pendulum. The mechanical system was built utilizing V-slot rail components and custom 3D printed parts.

Novel features of this work included:

1. The construction of a physical pendulum in which the pole length, viscosity, and resistive friction could be easily modified.
2. The construction of a testing rig, designed to apply consistent taps to the pendulum pole, to evaluate its reaction to impulsive disturbances.
3. Use of system identification to investigate and quantify the pendulum's uncontrolled dynamics as the physical characteristics of the pendulum system were changed.
4. The design of two linear controllers, based on observer-based full state feedback control modes, supporting either velocity or position control of the pendulum cart while balancing the pendulum pole.
5. Performance testing to investigate the pendulum's controlled dynamics as its physical characteristics were changed.

16.2. Uncontrolled pendulum oscillations

We first investigated the behavior of the pendulum in its stable downward inverted configurations. This was done for all three test pendulum pole lengths, as well as under conditions of both high and low static friction and high and low viscous damping. We examined the pendulum's free oscillations following an initial large angle displacement of approximately 90°, and also after initial small angle displacements of about 5°, caused by taps from the tapping mechanism. As expected, increasing the pole length reduced the oscillatory frequency of the pendulum.

Increasing viscosity using the paddle mechanism significantly reduced the decay time of oscillations in the large angle case. In this scenario, the added viscous resistance dominated the pendulum's oscillatory decay, resulting in an exponential reduction in amplitude over time. However, in the small angle condition, where static sliding friction was a significant source of decay, the paddle had minimal effect. This was evidenced by a more linear decay of oscillatory amplitude.

When we increased static friction equally in both the large and small angle cases, the effects differed. In the large angle case, this increase led to a noticeable, albeit modestly more linear, decay of oscillations. Conversely, in the small angle condition, the same level of friction had a dramatic effect, rapidly decelerating oscillations and bringing them to a standstill within just a few cycles.

These observations indicate that the viscous resistance from the paddle is better represented by the square of the movement velocity than by the linear model commonly used in mathematical modeling. This is particularly true since the paddle's effect is significant only during faster movements. Furthermore, at small amplitudes, where the effect of gravity acting on the pole produces minimal torque, any substantial static frictional resistance can dominate the damping behavior.

16.3. System identification

To estimate the pendulums' parameters, we applied a system identification procedure to the small angle dataset. This data best represents the small angle condition occurring during the balancing of the inverted pendulum. We used optimization to fit the predicted pendulum damped oscillatory decay waveform with data recorded from the pendulum in different configurations. We noted that including a friction term in the mathematical model of the non-inverted pendulum was necessary in this process. This addition accounts for the linear aspect of decay and achieves a good fit. The fitting procedure enabled the estimation of the viscous and static friction terms, as well as an updated estimate of the effective pendulum pole length, mass, and moment of inertia.

16.4. Controlled pendulum results

We undertook a sequence of system tests, starting with the standard pendulum pole length for which the controllers were developed. We then investigated how the stabilized pendulum pole responded to light taps from the tapper mechanism, creating impulses. This testing was carried out in both velocity and position control modes. In velocity mode, the cart could be driven with a feedforward velocity command to move left or right while maintaining balance. In position control mode, which operates with integral action on the cart's position, the control system maintains the pendulum's cart at a specified location. We observed distinct behavioral differences between the two control modes.

Unsurprisingly, velocity control, unconcerned with cart position, responded to disturbances to the pendulum pole with balancing movements of the cart, typically causing a positional shift. In position control mode, a disturbance similarly resulted in balancing movement, but the cart gradually returned to its initial position.

Since the cart's movement velocity is limited, any disturbance requiring faster movement would naturally result in a loss of balance. This limitation also affects the balancing robustness in velocity control mode. If the cart is already moving in one direction, its capacity for additional corrective velocity in that direction is restricted.

16.5. Transfer of controller operation to other pole lengths

To assess controller performance, we conducted balancing system tests using both shorter and longer pendulum pole lengths, comparing them to the standard length for which the controllers were developed. We found that the controllers operated well with the shorter length pendulum pole. However, with the longer pendulum pole, we observed some oscillatory behavior of the cart, particularly in position control mode. Despite this, the inverted pendulum balance was still maintained.

We note that the gains for both controller modes, derived from the Linear Quadratic Regulator (LQR) design process, were based solely on experimentation with the standard pendulum. Naturally, there remains a strong likelihood that more optimal controllers for both modes could exist, particularly if they were specifically designed for these varying pendulum lengths.

16.6. Future work

In the current study, we examined the effects of a relatively low-intensity, fixed impulsive disturbance on pendulums using a tapper mechanism. We utilized this setup to assess how the pendulums responded as their characteristics were altered and to observe the behaviors of velocity and position control. These tests did not result in a loss of balance; rather, they only evaluated the reactions necessary to maintain equilibrium. It would be enlightening to apply a wider range of impulse intensities and compare the behaviors related to loss of balance across the various conditions and controllers examined in this study. The existing

tapper mechanism could not be easily adjusted to provide a range of impulse intensities. For future research, constructing a tapper device that employs a motor to drive the tapper rod would be a valuable exercise, allowing for precise control over the impulse intensity.

Currently, we have examined the behavior of two different state feedback controller architectures for balancing a pendulum. Many other control approaches exist, and a comparison with methods such as PID (Proportional-Integral-Derivative) [16,17], and reinforcement learning [22–24], would be informative.

The inverted pendulum has been valuable in understanding human balance while standing [44–52]. Future studies within the framework of our pendulum system could further explore and model human behavior in such tasks. These studies could investigate factors that constrain human performance, including sensory feedback latency, noise in the control and sensory systems, as well as the force, stiffness, and speed of movement characteristics of muscles [68]. Such issues could be readily incorporated into the MATLAB simulations as well as real-time control of the physical pendulum system.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

Support for LAH was provided by the Engineering and Physical Sciences Research Council and The University of Plymouth. ISH was supported by The University of Plymouth. We also thank Jonathan Marsden and Gunnar Schmidtmann in the School of Health Professions at the University of Plymouth, for insightful discussions on this work.

References

[1] L. Alvarez-Hidalgo, I.S. Howard, “Gain scheduling for state space control of a dual-mode inverted pendulum,” in 2022 International Conference on System Science and Engineering (ICSSE), IEEE: 39–46, 2022, doi:10.1109/ICSSE55923.2022.9947361.

[2] P. Horáček, “Laboratory experiments for control theory courses: A survey,” Annual Reviews in Control, **24**, 151–162, 2000, doi:https://doi.org/10.1016/S1367-5788(00)90029-4.

[3] K.H. Lundberg, T.W. Barton, “History of inverted-pendulum systems,” IFAC Proceedings Volumes, **42**(24), 131–135, 2010, doi:https://doi.org/10.3182/20091021-3-JP-2009.00025.

[4] H. Wang, H. Dong, L. He, Y. Shi, Y. Zhang, “Design and simulation of LQR controller with the linear inverted pendulum,” in 2010 international conference on electrical and control engineering, IEEE: 699–702, 2010, doi: 10.1109/iCECE.2010.178.

[5] I. Kafetzis, L. Moysis, “Inverted Pendulum: A system with innumerable applications,” School of Mathematical Sciences, 2017.

[6] O. Boubaker, “The inverted pendulum benchmark in nonlinear control theory: a survey,” International Journal of Advanced Robotic Systems, **10**(5), 233, 2013, doi:https://doi.org/10.5772/55058.

[7] N. Muskinja, B. Tovornik, “Swinging up and stabilization of a real inverted pendulum,” IEEE Transactions on Industrial Electronics, **53**(2), 631–639, 2006, doi:10.1109/TIE.2006.870667.

[8] M. Turner, T.R. Cooley, “A Low-cost and Flexible Open-source Inverted Pendulum for Feedback Control Laboratory Courses,” in 2015 ASEE Annual Conference & Exposition, 26.63. 1-26.63. 13, 2015, doi:DOI:10.18260/p.23404.

[9] K. Kaheman, U. Fasel, J.J. Bramburger, B. Strom, J.N. Kutz, S.L. Brunton, “The experimental multi-arm pendulum on a cart: A benchmark system for chaos, learning, and control,” ArXiv Preprint ArXiv:2205.06231, 2022, doi:https://doi.org/10.48550/arXiv.2205.06231.

[10] F. Grasser, A. D’arrigo, S. Colombi, A.C. Rufer, “JOE: a mobile, inverted pendulum,” IEEE Transactions on Industrial Electronics, **49**(1), 107–114, 2002, doi:10.1109/41.982254.

[11] H. Vasudevan, A.M. Dollar, J.B. Morrell, “Design for control of wheeled inverted pendulum platforms,” Journal of Mechanisms and Robotics, **7**(4), 41005, 2015, doi:https://doi.org/10.1115/1.4029401.

[12] M. Hehn, R. D’Andrea, “A flying inverted pendulum,” in 2011 IEEE International Conference on Robotics and Automation, IEEE: 763–770, 2011, doi:10.1109/ICRA.2011.5980244.

[13] S. Awtar, N. King, T. Allen, I. Bang, M. Hagan, D. Skidmore, K. Craig, “Inverted pendulum systems: rotary and arm-driven-a mechatronic system design case study,” Mechatronics, **12**(2), 357–370, 2002, doi:https://doi.org/10.1016/S0957-4158(01)00075-7.

[14] B. Tomar, N. Kumar, M. Sreejeth, “Optimal Control of Rotary Inverted Pendulum Using Continuous Linear Quadratic Gaussian (LQG) Controller,” in 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE: 1–6, 2023, doi:10.1109/ICCCNT56998.2023.10306449.

[15] P. Strakoš, J. Tůma, “Mathematical modelling and controller design of inverted pendulum,” in 2017 18th International Carpathian Control Conference (ICCC), IEEE: 388–393, 2017, doi:10.1109/CarpathianCC.2017.7970431.

[16] K. Sultan, A. Mirza, “Inverted Pendulum, Analysis, Design and Implementation,” Visionaries Document, 2003.

[17] K. Razzaghi, A.A. Jalali, “A new approach on stabilization control of an inverted pendulum, using PID controller,” Advanced Materials Research, **403**, 4674–4680, 2012, doi:https://doi.org/10.4028/www.scientific.net/AMR.403-408.4674.

[18] I.S. Howard, “A modular 3D-printed inverted pendulum,” in Towards Autonomous Robotic Systems: 20th Annual Conference, TAROS 2019, London, UK, July 3–5, 2019, Proceedings, Part I 20, Springer: 413–424, 2019, doi:https://doi.org/10.1007/978-3-030-23807-0_34.

[19] C.A. Ibanez, O.G. Frias, M.S. Castanon, “Lyapunov-based controller for the inverted pendulum cart system,” Nonlinear Dynamics, **40**, 367–374, 2005, doi:https://doi.org/10.1007/s11071-005-7290-y.

[20] Y. Yang, P. Wang, T. Zhang, “Modelling and controller design of planar inverted pendulum system,” in 2014 International Conference on Mechatronics, Electronic, Industrial and Control Engineering (MEIC-14), Atlantis Press: 57–61, 2014, doi:10.2991/meic-14.2014.14.

[21] H. Lipson, Robots on the run, 2019, doi:doi: https://doi.org/10.1038/d41586-019-00999-w.

[22] M. Safeea, P. Neto, “A Q-learning approach to the continuous control problem of robot inverted pendulum balancing,” Intelligent Systems with Applications, 200313, 2023, doi:https://doi.org/10.1016/j.iswa.2023.200313.

[23] Y. Ma, D. Xu, J. Huang, Y. Li, “Robust Control of An Inverted Pendulum System Based on Policy Iteration in Reinforcement Learning,” Applied Sciences, **13**(24), 13181, 2023, doi:https://doi.org/10.3390/app132413181.

[24] S. Israilov, L. Fu, J. Sánchez-Rodríguez, F. Fusco, G. Allibert, C. Raufaste, M. Argentina, “Reinforcement learning approach to control an inverted pendulum: A general framework for educational purposes,” PLoS One, **18**(2), e0280071, 2023, doi:https://doi.org/10.1371/journal.pone.0280071.

[25] M. Deisenroth, C.E. Rasmussen, “PILCO: A model-based and data-efficient approach to policy search,” in Proceedings of the 28th International Conference on machine learning (ICML-11), 465–472, 2011.

[26] C.W. Anderson, “Learning to control an inverted pendulum using neural networks,” IEEE Control Systems Magazine, **9**(3), 31–37, 1989, doi:10.1109/37.24809.

[27] V. Mladenov, G. Tsenov, L. Ekonomou, N. Harkiolakis, P. Karampelas, “Neural network control of an inverted pendulum on a cart,” in WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering, World Scientific and Engineering Academy and Society, 2009.

[28] A. Suresh, M.P.F. Queen, V.A.T.P. Symon, A. Linsely, “Control and stabilization of inverted pendulum using GA based controller,” International Journal of Mechanical Engineering and Technology, **8**(8), 748–756, 2017.

[29] S. Omatu, S. Deris, “Stabilization of inverted pendulum by the genetic algorithm,” in Proceedings of IEEE international conference on evolutionary computation, IEEE: 700–705, 1996, doi:10.1109/ICEC.1996.542687.

[30] E.S. Sazonov, P. Klinkhachorn, R.L. Klein, "Hybrid LQG-neural controller for inverted pendulum system," in Proceedings of the 35th Southeastern Symposium on System Theory, 2003., IEEE: 206–210, 2003, doi:10.1109/SSST.2003.1194559.

[31] P. Kumar, K. Chakraborty, R.R. Mukherjee, S. Mukherjee, "Modelling and controller design of inverted pendulum," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), **2**(1), 2013.

[32] W. Zhong, H. Rock, "Energy and passivity based control of the double inverted pendulum on a cart," in Proceedings of the 2001 IEEE International Conference on Control Applications (CCA'01)(Cat. No. 01CH37204), IEEE: 896–901, 2001, doi:10.1109/CCA.2001.973983.

[33] J. Baek, C. Lee, Y.S. Lee, S. Jeon, S. Han, "Reinforcement learning to achieve real-time control of triple inverted pendulum," Engineering Applications of Artificial Intelligence, **128**, 107518, 2024, doi:https://doi.org/10.1016/j.engappai.2023.107518.

[34] W. An, Y. Li, "Simulation and control of a two-wheeled self-balancing robot," in 2013 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE: 456–461, 2013, doi:DOI: 10.1109/ROBIO.2013.6739501.

[35] M. Hasan, C. Saha, M.M. Rahman, M.R.I. Sarker, S.K. Aditya, "Balancing of an inverted pendulum using PD controller," Dhaka University Journal of Science, **60**(1), 115–120, 2012, doi:DOI: 10.3329/dujs.v60i1.10348.

[36] H.-S. Juang, K.-Y. Lum, "Design and control of a two-wheel self-balancing robot using the arduino microcontroller board," in 2013 10th IEEE International Conference on Control and Automation (ICCA), IEEE: 634–639, 2013, doi:DOI: 10.1109/ICCA.2013.6565146.

[37] S. Wang, L. Cui, J. Zhang, J. Lai, D. Zhang, K. Chen, Y. Zheng, Z. Zhang, Z.-P. Jiang, "Balance control of a novel wheel-legged robot: Design and experiments," in 2021 IEEE International Conference on Robotics and Automation (ICRA), IEEE: 6782–6788, 2021, doi:10.1109/ICRA48506.2021.9561579.

[38] K. Darvish, L. Penco, J. Ramos, R. Cisneros, J. Pratt, E. Yoshida, S. Ivaldi, D. Pucci, "Teleoperation of humanoid robots: A survey," IEEE Transactions on Robotics, 2023, doi:10.1109/TRO.2023.3236952.

[39] A. Bratta, M. Focchi, N. Rathod, C. Semini, "Optimization-Based Reference Generator for Nonlinear Model Predictive Control of Legged Robots," Robotics, **12**(1), 6, 2023, doi:https://doi.org/10.3390/robotics12010006.

[40] J.A. Castano, J. Humphreys, E. Mingo Hoffman, N. Fernández Talavera, M.C. Rodriguez Sanchez, C. Zhou, "Benchmarking Dynamic Balancing Controllers for Humanoid Robots," Robotics, **11**(5), 114, 2022, doi:https://doi.org/10.3390/robotics11050114.

[41] S. Kajita, M. Morisawa, K. Miura, S. Nakaoka, K. Harada, K. Kaneko, F. Kanehiro, K. Yokoi, "Biped walking stabilization based on linear inverted pendulum tracking," in 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE: 4489–4496, 2010, doi:10.1109/IROS.2010.5651082.

[42] D.A. Winter, "Human balance and posture control during standing and walking," Gait & Posture, **3**(4), 193–214, 1995, doi:https://doi.org/10.1016/0966-6362(96)82849-9.

[43] F. Quijoux, A. Nicolaï, I. Chairi, I. Bargiotas, D. Ricard, A. Yelnik, L. Oudre, F. Bertin-Hugault, P. Vidal, N. Vayatis, "A review of center of pressure (COP) variables to quantify standing balance in elderly people: Algorithms and open-access code," Physiological Reports, **9**(22), e15067, 2021, doi:https://doi.org/10.14814/phy2.15067.

[44] P. Gawthrop, I. Loram, M. Lakie, H. Gollee, "Intermittent control: a computational theory of human control," Biological Cybernetics, **104**, 31–51, 2011, doi:https://doi.org/10.1007/s00422-010-0416-4.

[45] I.D. Loram, H. Gollee, C. van de Kamp, P.J. Gawthrop, "Is Intermittent Control the Source of the Non-Linear Oscillatory Component (0.2–2Hz) in Human Balance Control?," IEEE Transactions on Biomedical Engineering, **69**(12), 3623–3634, 2022, doi:10.1109/TBME.2022.3174927.

[46] J.G. Milton, "Time delays and the control of biological systems: An overview," IFAC-PapersOnLine, **48**(12), 87–92, 2015.

[47] J. Milton, J.L. Cabrera, T. Ohira, S. Tajima, Y. Tonosaki, C.W. Eurich, S.A. Campbell, "The time-delayed inverted pendulum: implications for human balance control," Chaos: An Interdisciplinary Journal of Nonlinear Science, **19**(2), 2009, doi:https://doi.org/10.1063/1.3141429.

[48] A. Kot, A. Nawrocka, "Modeling of human balance as an inverted pendulum," in Proceedings of the 2014 15th International Carpathian Control Conference (ICCC), IEEE: 254–257, 2014, doi:10.1109/CarpathianCC.2014.6843607.

[49] I.D. Loram, M. Lakie, "Human balancing of an inverted pendulum: position control by small, ballistic-like, throw and catch movements," The Journal of Physiology, **540**(3), 1111–1124, 2002, doi:https://doi.org/10.1113/jphysiol.2001.013077.

[50] J.-L. Sung, C.-Y. Hong, C.-H. Liu, P. Lee, L.-Y. Guo, N.-H. Lin, C.-W. Yen, L.-J. Liaw, "Characterizing the validity of the inverted pendulum model for quiet standing," Journal of Healthcare Engineering, **2021**, 1–6, 2021, doi:https://doi.org/10.1155/2021/8884614.

[51] I.D. Loram, H. Gollee, M. Lakie, P.J. Gawthrop, "Human control of an inverted pendulum: is continuous control necessary? Is intermittent control effective? Is intermittent control physiological?," The Journal of Physiology, **589**(2), 307–324, 2011, doi:https://doi.org/10.1113/jphysiol.2010.194712.

[52] P. Morasso, A. Cherif, J. Zenzeri, "Quiet standing: The single inverted pendulum model is not so bad after all," PLoS One, **14**(3), e0213870, 2019, doi:https://doi.org/10.1371/journal.pone.0213870.

[53] S. Franklin, J. Česonis, D.W. Franklin, "Influence of visual feedback on the sensorimotor control of an inverted pendulum," in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE: 5170–5173, 2018, doi:10.1109/EMBC.2018.8513461.

[54] J. Česonis, S. Franklin, D.W. Franklin, "A simulated inverted pendulum to investigate human sensorimotor control," in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE: 5166–5169, 2018, doi:10.1109/EMBC.2018.8513434.

[55] R. Leib, J. Česonis, S. Franklin, D.W. Franklin, "LQG framework explains performance of balancing inverted pendulum with incongruent visual feedback," in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE: 1940–1943, 2019.

[56] S. Franklin, J. Česonis, R. Leib, D.W. Franklin, "Feedback delay changes the control of an inverted pendulum," in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE: 1517–1520, 2019, doi:10.1109/EMBC.2019.8856897.

[57] F. Ghorbani, M.A. Shooredeli, M. Teshnehlab, "Fault tolerant improvement with chaos synchronization using Fuzzy-PID control," in 2013 13th Iranian Conference on Fuzzy Systems (IFSC), IEEE: 1–5, 2013, doi:10.1109/IFSC.2013.6675645.

[58] M.F. Hamza, H.J. Yap, I.A. Choudhury, A.I. Isa, A.Y. Zimit, T. Kumbasar, "Current development on using Rotary Inverted Pendulum as a benchmark for testing linear and nonlinear control algorithms," Mechanical Systems and Signal Processing, **116**, 347–369, 2019, doi:https://doi.org/10.1016/j.ymssp.2018.06.054.

[59] M.H. Raibert, H.B. Brown Jr, "Experiments in balance with a 2D one-legged hopping machine," 1984, doi:https://doi.org/10.1115/1.3149668.

[60] H.-S. Juang, K.-Y. Lum, "Design and control of a two-wheel self-balancing robot using the arduino microcontroller board," in 2013 10th IEEE International Conference on Control and Automation (ICCA), IEEE: 634–639, 2013.

[61] S. Monteleone, F. Negrello, G. Grioli, M.G. Catalano, A. Bicchi, M. Garabini, "A method to benchmark the balance resilience of robots," Frontiers in Robotics and AI, **9**, 2022, doi:https://doi.org/10.3389/frobt.2022.817870.

[62] J.C. Simbach, J. Priest, "Another look at a damped physical pendulum," American Journal of Physics, **73**(11), 1079–1080, 2005, doi:http://dx.doi.org/10.1119/1.1858488.

[63] L.F. da C. Zonetti, A.S.S. Camargo, J. Sartori, D.F. De Sousa, L.A. de O. Nunes, "A demonstration of dry and viscous damping of an oscillating pendulum," European Journal of Physics, **20**(2), 85, 1999, doi:DOI: 10.1088/0143-0807/20/2/004.

[64] I.R. Lapidus, "Motion of a harmonic oscillator with sliding friction," Am. J. Phys, **38**(11), 1360–1361, 1970, doi: http://dx.doi.org/10.1119/1.1976111.

[65] M.I. Molina, "Exponential versus linear amplitude decay in damped oscillators," The Physics Teacher, **42**(8), 485–487, 2004, doi:http://dx.doi.org/10.1119/1.1814324.

[66] K.J. Åström, R.M. Murray, Feedback systems: an introduction for scientists and engineers, Princeton university press, 2021.

[67] C.B. Moler, Numerical computing with MATLAB, SIAM, 2004.

[68] D.W. Franklin, D.M. Wolpert, "Computational mechanisms of sensorimotor control," Neuron, **72**(3), 425–442, 2011, doi:10.1016/j.neuron.2011.10.006.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 9, No. 1, 133-142 (2024)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

A Novel Metric for Evaluating the Stability of XAI Explanations

Falko Gawantka[*,1], Franz Just[1], Marina Savelyeva[1], Markus Wappler[2], Jörg Lässig[1,2]

[1] *University of Applied Sciences Zittau/Görlitz, Faculty of Electrical Engineering and Computer Science, Görlitz, 02826, Germany*

[2] *Fraunhofer IOSB, Advanced System Technology (AST), Görlitz, 02826, Germany*

ARTICLE INFO

Article history:
Received: 15 November, 2023
Revised: 16 January, 2024
Accepted: 17 January, 2024
Online: 21 February, 2024

Keywords:
eXplainable AI
Evaluation
Stability of explanations

ABSTRACT

Automated systems are increasingly exerting influence on our lives, evident in scenarios like AI-driven candidate screening for jobs or loan applications. These scenarios often rely on eXplainable Artificial Intelligence (XAI) algorithms to meet legal requirements and provide understandable insights into critical processes. However, a significant challenge arises when some XAI methods lack determinism, resulting in the generation of different explanations for identical inputs *(i.e., the same data instances and prediction model). The question of explanation stability becomes paramount in such cases. In this study, we introduce two intuitive methods for assessing the stability of XAI algorithms. A taxonomy was developed to categorize the evaluation criteria and the ideas were expanded to create an objective metric to classify the XAI algorithms based on their explanation stability.*

1 Introduction

As artificial intelligence (AI) becomes increasingly integrated into people's lives, a notable trend can be observed: automation of processes and domains that significantly impact the well-being and future prospects of individuals. For example, the initial screening of job applicants can be considered. Given the substantial human resource expenditure involved in this process, a compelling case can be made for its automation. Candidate scores can be calculated using artificial intelligence, as demonstrated in previous research [1]. Refer to Figure 1 for more details. Nevertheless, the decision-making process utilized by the underlying machine learning model presents a challenge. The accumulation of applicant information often results in a lack of transparency, a concern shared by HR managers, companies, applicants, and regulators alike.

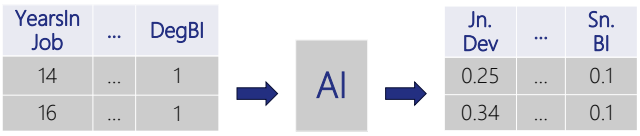


Figure 1: The job application pipeline is depicted here. On the left, the input values predicted by the ML model are displayed, while on the right, the corresponding results are presented. Transparency issues arise when the information density is reduced to a single numeric value. This lack of transparency makes it neither intuitive nor comprehensible for the stakeholders involved. The question of: „What impact does feature x had on the prediction?" arise.

One viable approach to address this concern is the integration of explanation algorithms, often referred to as XAI algorithms. This strategy enables visualization of model decisions in an understandable way. Additionally, both the transparency and the validity of the predictive model can be improved, ensuring that the needs and expectations of the relevant stakeholders are met.

However, it is essential to recognize that not all explanatory algorithms can be used without restrictions in high-stake areas [1]. This journal paper extends the work originally presented in IEEE ICCI*CC'22 [1]. The stability of the explanations was investigated through different XAI approaches and ranked according to their results. This was achieved by quantifying the feature importance values (FIVs). In this work, the state of the art for the evaluation of XAI explanations is presented, and the taxonomies found are expanded or modified in terms of the aspect: *stability*. The corresponding research questions are as follows.

- RQ1: What does the *stability* of an explanation mean?
- RQ2: How can the explanation of an XAI algorithm be measured in terms of *stability*?

By that, the following hypothesis can be stated: „High *stability* of an explanation is observed when the explanation undergoes minimal changes in response to minor variations in unimportant features of the data instance."

[*]Corresponding Author:University of Applied Sciences Zittau/Görlitz, Faculty of Electrical Engineering and Computer Science, Görlitz, 02826, Germany, Email: falko.gawantka@hszg.de

[1]This domains could harm people or the future of a person, for instance here the HR use case represents an automated decision that has an impact on a person.

For example, in the case of an HR applicant selection process, a change in the input features (i.e. 14 vs 16 *YearsInJob*) should not produce a substantially different explanation.

The paper is organized as follows. In section 2, metrics and evaluation techniques for XAI algorithms are presented, and the foundations of XAI and representative methods are discussed in detail. In section 3, an approach to quantify explanation stability is introduced. The details and limitations of previous approaches serve as a basis for a new scoring metric. The results are shown in section 4. Potential future research directions are outlined in the discussion in section 5.

2 Related Work and Background

For a deeper understanding and examination of XAI algorithms, an overview of evaluation taxonomies, methodologies, and the underlying calculation models is provided, forming the theoretical foundations. In addition to that, some background information are provided to enhance the understanding of XAI. This work aims to be considered as best practices in the selection process of XAI algorithms for real-world use cases, with the objective of uncovering a comprehensive understanding of the model with high *stability* to the respective explanation of XAI.

2.1 Related Work on XAI evaluation

To explore the stability of an explanation result, this subsection first defines the relevant properties/criteria. It then provides an overview of the measuring techniques and presents the mathematical foundations.

2.1.1 Notions by different Taxonomies

In the work of [2], a comprehensive taxonomy is introduced, a distinction is made between *the user aspect*, *the explanation aspect* and *the model aspect*. The focus of this work lies on *the explanation aspect*. In this aspect there are two further sub-criteria defined by [2], which have to do with stability: *Identity* and *Separability*. *Identity* is defined by: " identical instances should have identical explanations" and *Separability* is defined by: "non-identical instances should not have identical explanations" [2]. To describe these two aspects, one can think of the ends of a scale.

A paper cited by [2] was that of [3] that defines *Stability* as: It represents how similar are the explanations for similar instances. While consistency compares explanations between different models, stability compares explanations between similar instances for a fixed model. High stability means that slight variations in the feature values of an instance do not substantially change the explanation, unless these slight variations also strongly change the prediction.

Nonetheless, a lack of stability can also be created by non-deterministic components of the explanation method, such as a data sampling step. Regardless of that, high stability is always desirable. This definition brings another dimension into play. The author in [2] speaks only of "identical instances", and [3] also mentioned the predictor by constructing a context around the instance and saying that there is also a predictor and an instance.

Following the definitions in [2], the author in [4] describes in the work the so-called *Co-12 properties*. The *Consistency* property coincides with the *Identity* criterion from [2]. *Consistency* is defined as: "Identical inputs should have identical explanations" [4]. The term inputs can be defined by the data instance as well as the ML model. In this context, the model's perspective is integrated, with the rest remaining consistent with the definition in [2]. And there is also a second aspect in [4], which can be equated with *Stability* in [3]. The term *Continuity* introduced by [4] is defined as follows: "Similar inputs should have similar explanations" [4]. The use of similar inputs could mean both the model view and the data view.

Besides the *Stability* notion by [3] and the term *Continuity* by [4] there is a another work [5], where the author measured the *Stability* of XAI algorithms. The underlying assumption was that: If the inputs are almost identical, then it is expected that only minor changes appear in the explanation. This can be used as a scoring metric for the stability of an XAI algorithm. The term *almost identical* refers hereby to *similar* and therefore matches the understanding of *Stability* by [3] and the notion *Continuity* from the work of [4]. The author in [5] gives no explicit definition, but: "similar instances and an identical model should have similar explanations", could be suggested as a common understanding for the property of *Stability*. Another property or criterion investigated in [1] was the *Reproducibility/Stability of the explanation*. This refers to an aspect in which an identical model predicts an identical data instance several times (i.e., 10 times, 100 times, and 1000 times). In the context of the reproducibility of explanation stability, *Reliability* would be a more appropriate term than the one chosen by the author.

The entire notions of the taxonomies are grouped and summarized on the *Stability Scale* and could be seen in Figure 2.

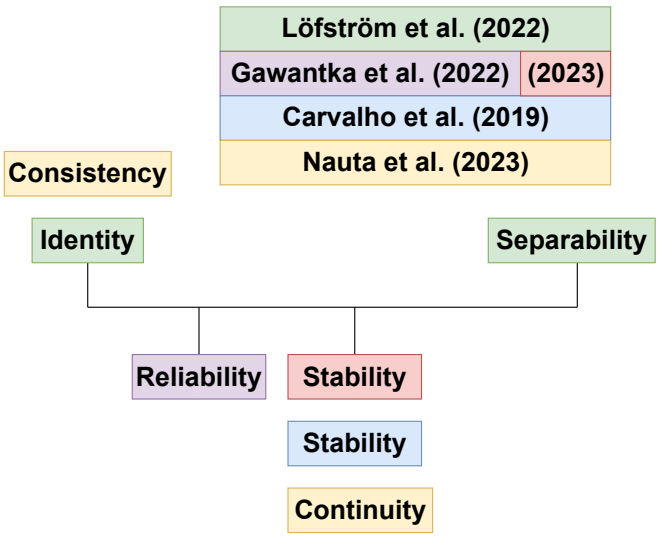


Figure 2: Stability Scale: The scale visualize the different terms and group them on the scale to show what terms could be used interchangeable.

The focus of this work lies in unifying different schools and terms, proposing concrete measuring techniques, and establishing a strong foundation from the mathematical perspective of XAI evaluation metrics.

2.1.2 Quantifying XAI through Feature Importance Values

A formal introduction into machine learning is initiated, drawing inspiration from a more detailed section in [6]. A comprehensive introduction to artificial intelligence is available in the textbook [7]. The notion of explanation method is taken and adapted from [8].

Feature importance values are used to quantify the explanations. These are numerical scores that help to evaluate the relative significance of individual features of the model's predictions, thus offering transparency and interpretability. It is important to note that the choice of the feature importance method matters and that different methods may lead to varying results. Therefore, careful consideration is needed to select the most appropriate method for a particular application. Adopting the notions presented in subsubsection 2.1.1 to the feature importance values will provide evaluation metrics to support the selection of data-driven methods.

In the context of supervised learning, XAI and the feature importance values will be considered. This is based on two fundamental components: the model and the data. Given an input domain X and an output domain Y, supervised learning aims to predict an output $y \in Y$ for given $x \in X$. In the example of classifying job applicants, the input domain X reflects the data structure of the applicant profile, while Y encodes the job profiles. The prediction mechanism is described through a mathematical model that can be deterministic or probabilistic.

A *deterministic model* is a function $f : X \to Y$ that predicts exactly for a given input $x \in X$ an output $f(x) \in Y$. This can be a classification prediction, in case Y is a discrete set, or this can be a regression prediction, if Y is a continuous set. For example f assigns a job profile $y \in Y$ to each applicant for a job $x \in X$.

A *probabilistic model* is a probability distribution $p : X \times Y \to [0, 1]$. It can be recognized as a full distribution, giving the probability of assignment of $x \in X$ and $y \in Y$ by $p(x, y)$, or it can be recognized as a conditional distribution, providing the probability of output $y \in Y$ for chosen $x \in X$ by $p(y|x)$. In the latter case for example, p would encode the probability that an applicant's profile $x \in X$ will be classified as job profile $y \in Y$.

In either deterministic model or probabilistic model, the underlying function is usually parameterized. Training the model means optimizing the parameters to achieve good predictions, based on a data set $\{(x^1, y^1), \ldots, (x^m, y^n)\} \subseteq X \times Y$.

The input data are commonly structured data $X = X_1 \times \cdots \times X_n$ with the notion of a *feature* being applied context-dependent to any primitive domain X_k or to any data entry $x_l \in X_k$ or instance $x \in X$. The output $y \in Y$ is called the label.

In the remainder of the paper, all models are considered to be deterministic. The set of all (deterministic) models is denoted by $\mathcal{F} = X \to Y$.

Having a (trained) model f, an explanation method $\Phi : \mathcal{F} \times (X = (X_1, \ldots, X_n)) \to \mathbb{R}^n$ attributes an importance score $\Phi(f, x)_i$ to each feature x_i describing its impact on prediction $f(x)$. The range of feature importance values is method-dependent. Common choices are $(-\infty, \infty)$ for SHAP, $(0, 1)$ for Gini importance, $(0, \infty)$ for Gain, $(-1, 1)$ for Correlation Coefficients.

Using this notion, *Consistency* means that $\Phi(f, x)$ evaluates to the same value, regardless of how often it is calculated. Obviously, this is always fulfilled, making the term irrelevant in the context of feature importance values. But still, a similar metric might be useful, as certain XAI methods, e.g. surrogate explainer methods, train a model and try to approximate the original ML model, e.g. LIME is such a method. In fact, each LIME instance $\Phi(f, x)$ might behave differently depending on the training data points used. Demanding that explanations at least do not deviate much would be a natural extension of *Consistency*.

Thus, in [1], another evaluation criterion known as *Reliability* is mentioned here. For every call of Φ an element is added to the result set $\{\Phi^{(r)}(f, \cdot)\}_r$. One can expect that a statistical quantity that measures the deviation, such as the length of confidence intervals, will decrease with increasing r. However, its rate of decrease provides a measure for *Reliability*.

The definition of *Reliability* is shown in the following equation:

$$\{\Phi^{(r)}(f, \cdot)\}_r = \left\{ \begin{pmatrix} fiv_{i=1} \\ \vdots \\ fiv_{i=s} \end{pmatrix}_{r=1}, \cdots, \begin{pmatrix} fiv_{i=1} \\ \vdots \\ fiv_{i=s} \end{pmatrix}_{r=c} \right\} \tag{1}$$

In Equation 1 the result of a single call from Φ is shown as vector. For every call r an explanation vector ($\vec{F}$) is calculated, so $fiv_i \in \vec{F}$. The result is a set of vectors that represents a matrix.

Based on this aspect, the mean over the 1st feature importance value from r-repetitions could be defined by:

$$\mu_{F_{(i=1)}} = \frac{1}{c} \sum_{r=1}^{c} fiv_{(i=1)r} \tag{2}$$

Therefore the variance could be calculated by the mean:

$$\sigma^2_{F_{(i=1)}} = \frac{1}{c} \sum_{r=1}^{c} (fiv_{(i=1)r} - \mu_{F_{(i=1)}})^2 \tag{3}$$

The standard deviation for a single feature importance value is then calculated from:

$$\sigma_{F_{(i=1)}} = \sqrt{\sigma^2_{F_{(i=1)}}} \tag{4}$$

Finally, the deviation can establish the foundation for the confidence interval of a single feature importance value.

$$CI_{F_{(i=1)}} = \mu_{F_{(i=1)}} \pm (z = 1.96) \cdot \frac{\sigma_{F_{(i=1)}}}{\sqrt{c}} \tag{5}$$

In the previous equations, an approach was presented to calculate the confidence interval of the single feature importance value of the feature importance vector $\vec{F}$. When this is done over the entire feature importance vector and then the mean is built on top of it, this matches the metric from [1]. This calculation process is denoted by $R(x)$.

The intuitive definition of *Stability* involves small perturbations of an instance that result in only minor changes to the explanations under the same model. To be numerically tangible this definition has to be measurable, hence the notion of correlated small changes needs to be quantified. In [9] the explanation methods SHAP and LIME are evaluated in terms of Lipschitz continuity, using the best Lipschitz constant as a stability measure. This approach is generalized while keeping it local (depending on the chosen data point). Furthermore, it will be demonstrated that the different stability measures previously used fall under this umbrella.

Lipschitz continuity can be defined in a general sense for metric spaces as follows: A function $f : X \to Y$ between metric spaces (X, d_X) and (Y, d_Y) is called *Lipschitz continuous* if there is a real constant $L \geq 0$ (referred to as *Lipschitz constant*) such that for all $x, \tilde{x} \in X$,

$$d_Y(f(x), f(\tilde{x})) \leq L d_X(x, \tilde{x}) . \tag{6}$$

This is, the distance of the images of f is bounded by the distance of the respective arguments up to a constant factor. In case of local explanation methods, e.g. as SHAP and LIME are, an instance-independent constant is not attainable. Therefore, we will base the definition on a data point-dependent Lipschitz factor within a neighborhood of the data point.

Let X be an input domain, equipped with the metric d_X. Let Y be an output domain equipped with the metric d_Y. Let $f \in \mathcal{F} = X \to Y$ be a model, and $\Phi : \mathcal{F} \times X \to X$ be an explanation method. An explanation $\Phi(f, x)$ can be termed *ε-stable* if there is a local real constant $L(x) \geq 0$ such that

$$d_Y(\Phi(f, x), \Phi(f, \tilde{x})) \leq L(x) d_X(x, \tilde{x}) \tag{7}$$

for all $\tilde{x} \in N_\varepsilon(x) = \{\tilde{x} \in X : d_X(x, \tilde{x}) \leq \varepsilon\}$. The Lipschitz factor $L(x)$ can be seen as a stability measure. The lower $L(x)$, the smaller the deviation, the higher the stability. Finding small Lipschitz factors can be done using the Monte Carlo method. Examples from the literature will be provided that specify the notion of stability.

Example 2.1. In [9] the metrics are induced by norms, resulting in $N_\varepsilon(x) = \{\tilde{x} \in X : \|x - \tilde{x}\| \leq \varepsilon\}$. The best Lipschitz constant is given by

$$L(x) = \max_{\tilde{x} \in N_\varepsilon(x) \setminus \{x\}} \frac{\|\Phi(f, x) - \Phi(f, \tilde{x})\|}{\|x - \tilde{x}\|} . \tag{8}$$

An approximation can be done by sampling with $\{x_1, \ldots, x_n\} \subset N_\varepsilon(x)$.

Example 2.2. In [10] an evaluation metric for XAI methods called Max-Sensitivity was introduced. The concept is taken up by [8]. To derive the term, the inequality is revisited (7) and relaxed to

$$d_Y(\Phi(f, x), \Phi(f, \tilde{x})) \leq \varepsilon L(x) = L'(x) . \tag{9}$$

Now, again using norms as distance metrics, $L'(x)$ can be obtained by

$$L'(x) = \max_{\tilde{x} : \|x - \tilde{x}\|_\infty \leq \varepsilon} \|\Phi(f, x) - \Phi(f, \tilde{x})\|_2 . \tag{10}$$

This is in fact the Max-Sensitivity metric. Again, it can be approximated by sampling.

Example 2.3. The work [5] considers a similarity measure by first generating data points by perturbing a feature around 1% and then evaluating the ratio of original feature importance value and perturbed feature importance value for that feature. Furthermore, the feature importance values are normalized so that they form a probability distribution.

Modifying this approach by allowing perturbations of at least ε, which bounds the values around one percent, allows to extend this measure to

$$S_i(x) = \max_{\tilde{x} \in N_\varepsilon(x)} \left| 1 - \frac{\Phi(f, x)_i}{\Phi(f, \tilde{x})_i} \right| . \tag{11}$$

Therefore, $S_i(x)$ is the best similarity to the feature i that all $\tilde{x} \in N_\varepsilon(x)$ fulfill. As a combined score for all features could be

$$S(x) = \max_{\tilde{x} \in N_\varepsilon(x)} \frac{\|\Phi(f, x) - \Phi(f, \tilde{x})\|_\infty}{\|\Phi(f, \tilde{x})\|_\infty} . \tag{12}$$

On one hand, this is similar to Lipschitz-based stability criteria; on the other hand, it effectively demonstrates that the criteria proposed in [5] normalize the feature importance values.

2.2 Background of XAI Taxonomies

At this point, eXplainable Artificial Intelligence is introduced as a field of study within the domain of artificial intelligence, along with an explanation of its associated focuses and terminology.

In [11, p. 3511] the author defines XAI as „the study of explainability and transparency for socio-technical systems, including AI.“ The 2019 XAI Taxonomy by [12] introduces important terms in explainable artificial intelligence. Later, in [13] used this taxonomy as a decision-making tool to choose the appropriate explanatory algorithms for the IBM AIX 360 tool. The tree structure of the taxonomy of [13] was transformed into a tabular format and can be seen in Table 1. The focus of this work is to provide an overview of explanatory models, which is highlighted in the table. The *data* leaf is not shown in the table and the interactive path is actually empty. The explainability of the *model* has a *local* and a *global path*; that is, there is a distinction between explaining the data sample and explaining the predictive model. Local interpretations of a single data instance can be done *ante-hoc*, i.e., the predictor is so comprehensible that a closer look serves as an explanation, for example, when the ML model is a simple decision tree. The other path, called XAI *post-hoc* methods, needs at least one execution of the predictor to produce information about the decision process of the ML model. The *global* model explanation is divided into two parts. The *direct* methods have the notion of direct interpretable and could be seen as *ante-hoc*. *Post-hoc* refers, in contrast to that, to algorithms that can create an explanation.

Due to rapid development in the field of XAI, the results of Table 1 are supplemented by other essential terms. The simple XAI taxonomy of [14], as well as the more sophisticated categories of [15], served as the basis for the following taxonomy, which has three sub-categories. The first sub-category is shown in Table 2.

Table 1: This IBM XAI Taxonomy is derived from the decision tree presented in [13], and serves as a guide to find a suitable XAI method. This overview focuses specifically on the model explanation aspect and does not present specific algorithms.

model					
local			***global***		
ante-hoc	***post-hoc***		direct	***post-hoc***	
	samples	features		surrogate	visualize

In Table 2, the distinction between XAI algorithms is made based on the creation of the explanation. For example, *Local Perturbation* methods try to modify the input and find important features as well as their corresponding feature importance values [14, 15]. The methods in *Leveraging Structure* use internal information from the ML model, such as gradients, like *Backprop* in the paper by

[14]. *Meta Explanations* abbreviated with *Meta-Expl.* exploit explanations from different explanatory methods, and *Architecture Modification* abbreviated with *Arch. Mod.* algorithms alter the predictor with the goal of finding simpler representations [15]. *Examples* Methods derive explanations from input data and use the output as an interpretation for similar inputs.

Table 2: This table presents the functional part by the XAI Taxonomy according to [14] and [15]. The methods are classified based on their functional interaction with the predictor model.

funtional approach				
Local Perturbation (Perturbation)	Leveraging Structure	Meta-Expl.	Arch. Mod. (Backprob)	Examples

Another classification uses the output of the explanation approaches, which is shown in Table 3. The first sub-category is the weighting of feature importance values that provides information on which features have an influence and how great this influence is. *Surrogate Models* try to approximate more complex models by focusing only locally or by generally using *ante hoc* explanatory models. In addition, example-based methods are also used here. [15]

Table 3: The scope of this part from the introduced XAI Taxonomy according to [14] and [15] is focused on the results that different explanation models provide. As it can be seen there are *Feature Importance* as well as *Example* explanations and *Surrogate Models* that can explain black box models.

result approach		
Feature Importance	Surrogate Models	Examples

The conceptual approach, shown in Table 4, is the last category introduced by [15] and covers all aspects that were introduced by [14] in 2020. Further breakdowns are made into the *usage*, the *scope* of the explanation, and the *other dimensions*. The *usage* includes procedures that can be explained from the outside or XAI algorithms that provide a *model agnostic* (*model-agno.*) or *model specific* (*model-spec.*) explanation after the predictor model. The *scope* describes the range of what is explained, e.g., the complete/*global* ML model or individual data instances. *Other dimensions* include outputs, among others [15]. IBM researchers suggest the following approaches for a feature-based explanation: CEM [16], LIME [17], and SHAP [18]. To understand the decision underlying the model, it is not enough to explain the sample (data instance), but it is fundamental to explain the impact of each feature on the decision. Based on the explanations for each feature, it may also be possible to identify bias. Considering the subject of *evaluation*, it is necessary to generate nearly equal outputs, falling into the sub-category of *Feature Importances*. In the proposed use case shown in section 1, and given the axiom that explanations are an advantage of decision support systems (DSS), the easiest way to implement this is to use *local post-hoc* Methods. A summary of suitable methods and their limitations is presented.

Table 4: This view of the XAI taxonomy from [14] and [15] highlight the explanatory capability of XAI methods. The main questions addressed here are: what is being explained and whether the explanation can be obtained from the XAI method before the machine learning model makes predictions. Additional information related to the XAI approach and the problem domain is also provided.

conceptual approach						
stage (usage)			***scope***		***other dim.***	
ante-hoc	***post-hoc***		glo.	loc.	res.	issues
	model-agno.	model-spec./ (intrinsic)				

2.3 Background of XAI Algorithms

In this sub-section are reviewed three XAI methods such as LIME (Local Interpretable Model-Agnostic Explanations by [17]), SHAP (SHapley Additive exPlanations by [18]) and CIU (Contextual Importance and Utility by [19]). To describe these algorithms, two definitions are needed, namely deterministic and nondeterministic algorithm.

An algorithm whose behavior depends entirely on the input data is called *deterministic*. Thus, processing the same input data always leads to the same result. A *nondeterministic* algorithm is an algorithm that specifies several ways to process the same input data - without any specification of which option will be chosen, which can lead to either the same or different output [20].

2.3.1 LIME

The idea behind LIME is to consider the local model as a black box model. The mode of operation of LIME is based on perturbing an original data point as input into the black box model and using the resulting predictions to train an interpretable surrogate model, which locally approximates the predictions of the black box model. The explanation provided by LIME is defined by:

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \, \mathcal{L}(f, g, \pi_x) + \Omega(g) \tag{13}$$

In Equation 13, ξ is the explanation of instance x, which is obtained through an optimization task. The function g is an interpretable local model, and G is a class of potentially interpretable models. The function f is the original predictor and π_x defines the radius of the neighborhood around instance x. $\mathcal{L}$ is the loss function that measures the accuracy of the prediction of the instance x with respect to the interpretable model g and the original prediction of f in the area of π_x around the original prediction. Ω is a complexity measure of g and serves as a penalty function.

LIME calculates feature importance values that show the contribution of each feature for and against a prediction in a certain class. LIME values are numerical, where a negative numerical value indicates that this feature is not in favor of the prediction. On the other hand, a positive numerical value shows that this feature has a positive influence on the prediction.

The advantages of LIME are the simple explanations diagrams and the ability to process various types of input [21]. However, LIME has some disadvantages: it provides only local explanations and does not show this to the end user. The interpretation of the

importance plots is often hard to understand by nonexperts. The algorithm has consistency issues and can be attacked by adversarial examples [21].

The consistency issues with LIME are not only related to minimal changes in features but also to the fact that such changes can lead to completely different explanations. In previous experiments with LIME, it has been shown that by iterating with LIME over the same data instance 10, 100 and 1000 times, the inconsistency is also present [1]. These results and [22] show that the LIME algorithm is nondeterministic.

2.3.2 SHAP

The idea of the SHAP algorithm has its origins in the game theory. Calculate the extent to which a coalition (set of features) contributes or does not contribute to a particular classification based on the so-called Shapley values. The implementation used was the implementation by [18], known as *SHap Additive exPlanations*. The following definition describes the generation of explanations by the algorithm [18, 23]:

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i \tag{14}$$

The function g describes the explanatory model and z' describes the data instance to be interpreted. The variable, z' may have only a subset of all features. The explanation is generated by a linear model, where $\phi_i \in \mathbb{R}$ and z'_i are either zero or one, to represent the presence or absence of a value from the feature set z' [18].

The author in [18] presents various explainer models in the framework, and the model used in [1] was the kernel explainer. The computational model approximates Shapley values with perturbations of z'. Thus, the complexity problem of computing Shapley values could be solved. SHAP also generates feature importance values such as LIME. A higher FIV of SHAP means that it contributes more to a prediction than a smaller one. A negative feature value argues against a prediction in a class and a positive value argues for a particular prediction.

SHAP also has some advantages: first, there is the sophisticated mathematical model for greater consistency and accuracy [17, 21, 23]. Besides that, the better intuition about feature weights, which is more related to humans, is another reason mentioned in [21]. Even though SHAP comes in hand with several advantages, the algorithm is non-deterministic [24], the approximation time of the Kernel SHAP is an issue [21, 23] and SHAP is also vulnerable to adversarial attacks [21].

2.3.3 CIU

The third model-agnostic algorithm is based on Decision Theory, more specifically, based on the subdomain of Multiple Criteria Decision Making (MCDM). In contrast to LIME and SHAP, this approach distinguishes between the measured importance and the utility of an attribute. Based on the relevance of the features, the focus lies on contextual importance (CI). This is described in [19, 25] as follows:

$$CI_j(\vec{C}, \{i\}) = \frac{C_{max_j}(\vec{C}, \{i\}) - C_{min_j}(\vec{C}, \{i\})}{absmax_j - absmin_j} \tag{15}$$

The explanation model CIU calculates with the function CI_j the importance of feature i in the feature vector $\vec{C}$ for an output label (value) j. The function C_{max_j} determines the maximum output of the prediction j for a certain feature i. The C_{min_j} calculation follows a similar approach. The functions $absmax_j$ and $absmin_j$ determine the highest and lowest prediction from the given data set. More details are provided in [19, 25, 26].

The CIU values are in the interval between 0 and 1, where a higher numerical value represents a greater influence on the prediction and vice versa. A value of 0 means that the attribute does not influence the prediction. In contrast to the other methods, in which negative values argue against a classification, this is not the case here (i.e., values are between 0 and 1).

The advantages of the CIU algorithm are the different working calculation model and the absence in the questions of consistency (such as in LIME), as well as the calculation time as in SHAP [1]. Thus, this algorithm is suitable as a control method in the group of model-agnostic methods like in [9]. In [1] it is shown that CIU has the best stability during different runs with the same instances to explain, i.e. in terms of repeatability it is a good benchmark for SHAP and LIME. The problem with the value interval of 0 to 1 was solved in [9] by introducing a threshold value as a decision boundary. If feature values exceed this limit, the value has a positive influence and vica versa. Despite the advantages of the CIU algorithm, it is also non-deterministic, due to its randomized data sampling for numeric values, which the HR use case currently is [25].

3 The Idea

According to [2] there are two key criteria when you want to quantify the stability of explanations. The first aspect is the *Identity* criteria, i.e., that you should have identical explanations for identical instances. The second criterion is the *Separability*, which means that when the data instances are not identical, the explanation should not be identical.

The initial proposal of this work is to understand that these two criteria are at different ends of the same scale. The next step is to expand this scale by the criteria that are also presented in subsubsection 2.1.1. A representation of the so-called *Stability Scale* is shown in Figure 3. The following sub-sections explain the development of the scale in greater detail.

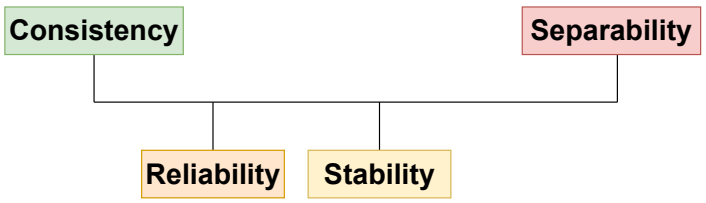


Figure 3: Stability Scale and their aspects (sub-criteria, properties)

3.1 Unification of Notions

For the standardization of notions (terms) the left end of the scale in Figure 3, serves as a fix point, from that other properties (terms), or criteria are explained. In Figure 2 there are two notions shown at

this fix point: *Consistency* and *Identity*. According to the definition of *Identity* in [2] it is mainly focused on the instances and this could lead to the idea that only the data instances are mentioned. In contrast, the definition of *Consistency* by [4] includes the data instances and the predictor is considered. That means, the scope is broader and not only data-centered. Due to this, the notion *Consistency* is preferred, because of its wider scope.

Since the term *Reliability*, which comes next to *Consistency*, is not considered in the other works, it is not necessary to unify this notion with another.

Stability comes next to *Reliability* and in Figure 2 it is shown that there are three notions. The definitions of the notions are very similar, according to the majority usage of the term *Stability*, it could be used as a common notion.

The right end of the scale, which is denoted with *Separability*, is only mentioned by [2] and there is also no need for a standardization.

It could be summarized that the initial idea, denoted in this work as *Explanation-Stability*, could be seen as an aspect of an explanation produced by an XAI algorithm. When it comes to details, it could be stated that this aspect of stability has sub-criteria and the initial notion is not sufficient, that means the *Stability Scale* represents all kinds of aspects related to the proposed notion *Explanation-Stability*.

3.2 The Notions of this work in Detail

In the following, it is explained why the notions (sub-criteria) are ordered as seen in Figure 3. As a foundation, the arrangement of the *sub-criteria of Explanation Stability* in Table 5 was employed.

Table 5: Comparison of sub-criteria from the Explanation Stability according to [1, 2, 3, 4, 5]. The abbreviations used in the table are: id. as *identical*, sim as *similar*. The aspects of the comparison are the input, the output (that is, the explanation), and the constraints. The yellow-highlighted terms indicate the differences from the previous row, specifically in comparison to the stricter criterion.

Notion	Inputs		Explanation	Constraints
	model	data		
Consistency	id.	id.	id.	determinism
Reliability	id.	id.	highly sim.	–
Stability	id.	sim.	sim.	–
Separability	id.	non-id.	non-id.	different instances

The sub-criteria *Consistency* is the most strict criterion because of the common definitions of [2] and [4]: "identical inputs should result in identical outputs". Despite this, [3] mentioned in their definition of *Stability* that: "Nonetheless, a lack of stability can also be created by non-deterministic components of the explanation method, such as a data sampling step. Regardless of that, high stability is always desirable". In this work, where *Stability* is examined in detail, the constraint of determinism is assigned to the *Consistency* criteria as the strongest requirement for stability.

A weaker requirement in terms of stability is *Reliability*. In Table 5, the second row summarizes this criterion. The inputs are the same as in the *Consistency* row, with the explanation being only *highly similar* compared to the first criterion. Therefore, this stability aspect could be fulfilled by nondeterministic XAI algorithms. To ensure uniform inputs and solely vary the output – specifically, the explanation property – the stability aspect here is weaker compared to *Consistency*.

Another stability property denoted by *Stability* is not as strong as the *Reliability* criterion. The only identical input is the model (i.e., the predictor), while the data instances are only similar, representing slight variations of an original data point. Consequently, the explanations are only *similar*, not *identical* or *highly similar*. This implies that the stability aspect, denoted as *Stability*, has less stringent requirements than both *Consistency* and *Reliability*. Consequently, it can be concluded that *Stability* is the least demanding stability criterion represented in Figure 3.

The *Separability* property, situated at the opposite end of the scale, asserts that two fundamentally different instances should not produce the same explanation on the same ML model. This high instability contrasts with or complements *Consistency*, measuring an opposite aspect on the *Stability Scale*. The summarized information is visualized in Figure 4.

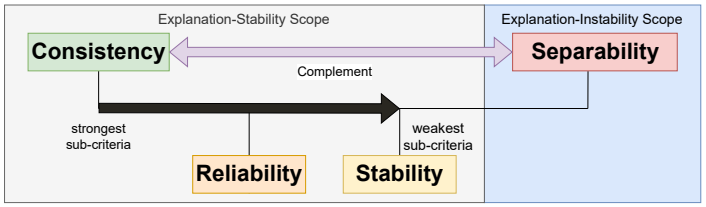


Figure 4: The proposed *Stability Scale* is shown here with additional context, starting from the left with the strongest criterion, up to the right with the complement of stability. The relations between the stability aspects: *Consistency, Reliability and Stability* are presented in this overview.

3.3 Mapping of Stability Criteria and Measures

At this point, it can be stated that there is an overarching concept for the stability of an explanation produced by an XAI approach. This term was introduced as *Explanation-Stability* and has three sub-criteria of stability. In order to quantify the *Explanation-Stability* objectively and to take all aspects into account, one should measure all sub-criteria of stability.

However, the properties of the XAI algorithms should be taken into account (subsection 2.3). In the proposed use case of [1], all XAI procedures are post-hoc approaches that generate local explanations. The explanation is of the type where each predicted feature $f(x_i)$ is associated with a feature importance value $\Phi(f,x)_i$. Therefore, for every explanation method Φ, the stability has been measured.

In case that all XAI approaches lack in terms of determinism, this stability criterion is very hard to fulfill by the used approaches. Therefore, more realistic stability measures such as *Reliability* and *Stability* are focused when measuring nondeterministic XAI algorithms. An overview of the measurement approaches assigned to the XAI methods used is shown in Table 6.

After assigning concrete stability measures, it is now denoted as *Explanation-Stability-Metric* due to the use of single stability-metric approaches.

Table 6: Assigning the proposed stability measures to non-deterministic, post-hoc explanation methods is discussed in detail in section 2.

	Explanation-Stability		
XAI Algorithm	**Consistency**	**Reliability**	**Stability**
LIME (2.3.1)	violate determinism	R(x)	S(x)
SHAP (2.3.2)			
CIU (2.3.3)			

3.4 Problems of Reliability and Stability

Even though the proposed approaches of the reliability measurement, presented in [1], as well as the stability analysis in [5] deliver highly valuable information about the quality of a certain XAI algorithm, the information are limited to one sub-criteria of the explainable models. The confidence interval method in [1] focuses on approaching the problem of *Reliability*, by conducting a test with the same data samples/models and with the expectation of highly similar data explanations. However, this approach lacks information regarding the stability of an XAI model.

To solve that, the paper [5] implements a method to evaluate the stability of an XAI model, by conducting changes on the input data. In contrast to [1], however, information regarding the reliability are missing. In addition, outliers are treated as regular explanations, which negatively effects the overall stability performance of the algorithm. An example for that can be seen in Figure 5, where several XAI algorithms are compared based on the similarity of the explanations in the feature "Income". By that, it can be seen that for instance LIME suffer from a high data dispersion because of several outliers. Due to a 1% change in one of the input features, the explanation similarity is expected to be in a range of 99% to 101%, where 100% refers to an identical explanation. However, algorithms like LIME seem to be prone to outliers. To avoid an overall biased score, those outliers needs to be handled accordingly.

Besides those problems, both papers do not provide a reference point, specifically an optimality criterion that serves as a baseline for gauging the algorithm's performance at its optimum. Instead, the only available approach is to compare algorithm A with another XAI method to determine whether algorithm A outperforms or underperforms in relation to algorithm B.

It can be concluded that stability measures are highly valuable when, combined because they address both stability aspects. Additionally, the use of an optimality criterion for the reliability metric is particularly interesting when evaluating the stability of an individual XAI algorithm. When considering the stability metric, it is important to take outliers into account.

3.5 Improvements of Reliability and Stability

To solve the problems mentioned in subsection 3.4, the current article presents a novel metric to evaluate current XAI solutions by combining the strengths of several XAI criteria. In particular, based on the approaches presented in equations $R(x)$ and Equation 12, the metric returns a vector of the corresponding scores to evaluate the XAI models. Since multiple aspects are considered, the metric provides an objective overall perspective on the performance of an XAI algorithm. In addition to that, the result vector enables comparability to other XAI models to find the best solution for the corresponding use-case.

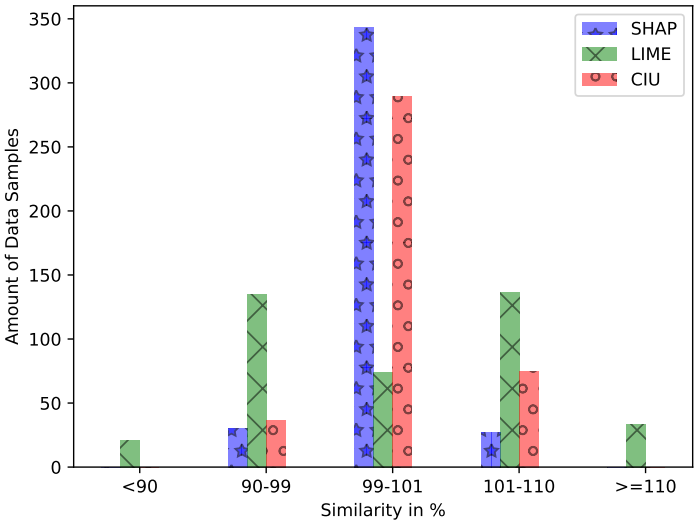


Figure 5: Algorithm Comparison about the similarity of the explanations for the feature „Income" - representation of the feature importance similarities for the data pairs. The similarity measures the percentage match between the explanation of the original data sample as well as the one with a feature change of around 1%. The results of *SHAP*, *LIME* and *CIU* are compared.

4 Results

The motivation for the novel approach to assess the stability aspects of XAI explanations was derived from the need to unify various notions. A comprehensive foundation was established through an extensive review of survey papers and articles on stability, including references such as [1]–[5]. This research aided in the identification and consolidation of the most relevant concepts and terms, forming the basis for the proposed *Stability Scale*, as depicted in Figure 2.

After gathering all key ideas and notions, a suitable process for unifying terms is detailed in subsection 3.1. The proposed and consistent scale is presented in Figure 3. The suggested notion of the XAI explanation stability is *Explanation-Stability*, it is important to emphasize that this is an overarching concept. For a more detailed view, the *Explanation Stability* is defined by three sub-criteria. These are visualized on a scale and ordered by their strictness according to Table 5. In fact, by defining the table and identifying these sub-criteria, research question RQ1 is fully answered.

Considering the real-world scenario presented in [1], it can be concluded that the *Explanation-Stability* construct of nondeterministic XAI approaches deteriorates in only two of the three criteria, namely *Reliability* and *Stability*. Furthermore, the outstanding idea of this work is to propose a concrete measurement approach for every stability aspect (sub-criterion). The sub-criterion of *Reliability* is measured by $R(x)$ and *Stability* is measured by $S(x)$, this concept is denoted as *Explanation-Stability-Metric*. Measures are described in depth in subsubsection 2.1.2 and improvements are identified in subsection 3.4.

This level of explicitness is unprecedented; unlike [2, 3], which provided only evaluation criteria and definitions. These works and

considerations were pivotal when addressing the overarching question: "How can we evaluate XAI in general?" When the focus narrows to evaluating specific aspects, such as the actual explanation, and the scope is tightly defined, sources are rarely limited. Developing a new standard for evaluating XAI methods entails constructing a robust foundation from an initial and intuitive idea. This process involves being highly specific and thereby bridging existing gaps.

In addressing research question RQ2, the newly proposed approach, known as *Explanation-Stability-Metric*, has the potential to provide at least a partial answer to nondeterministic XAI approaches. With additional effort, expanding the metric to incorporate the stability aspect of *Consistency*, it could also be applied to deterministic XAI methods.

The key idea is to provide a holistic measure of *Explanation-Stability*; therefore, it is necessary to furnish information about stability. Upon the development of an initial outcome from the execution of this metric, a stability vector would provide all stakeholders with sufficient information to better trust, build, and comprehend AI models. Furthermore, this work should contribute to making autonomous systems applicable in high-risk domains.

5 Discussion

Future work and assumptions are considered to enhance the new stability assessment approach outlined in this study. Firstly, the question arises: "How useful is the addition of the stability metric for *Consistency* concerning nondeterministic XAI methods?". This can be achieved by incorporating a distance measure that supports multidimensional data, such as the Euclidean distance. This enhancement provides a more objective overview of stability, which is valuable even when these methods produce less accurate scores than deterministic methods.

In the proposed stability metric, a complementary aspect to *Consistency*, referred to as *Separability*, was identified. Its addition could prove beneficial not only for measuring stability aspects but also for assessing *instability*, as mentioned in subsection 3.2. This broadens the scope of stability and introduces a related aspect into the metric. However, the inclusion of these two additional criteria requires further exploration in future work.

Considering the scope of stability, measuring all four sub-criteria could prove valuable, potentially making this approach applicable to deterministic XAI algorithms.

When considering the representation of the results, a second significant question arises: how to combine the measures of all sub-criteria into a single score. Several reasons justify this additional processing step. The fundamental concept of a metric is to distill relevant information into a real-valued single score. This approach offers a significant advantage, preventing information overload for those seeking insights.

As AI systems become more prevalent in our daily lives, ensuring information about stability is accessible to non-experts is essential. A single-score approach facilitates this by offering a concise and easily understandable representation.

However, aggregating the information also introduces a few disadvantages as it may lead to the loss of detailed information when reducing the sub-criteria scores of *Explanation-Stability* into a single score. When the resulting scores from each sub-criterion are condensed into a single value, questions such as "Could *Stability sub-criterion A* potentially compensate for the lower performance of *Stability sub-criterion B*, and to what extent does this influence exist?" arise.

Other important considerations include "How strong is the correlation between these stability aspects?", "Is there a dominant stability criterion?" and "Are there domain-dependent sub-criteria that should be prioritized, such as in the medical field or other sensitive use cases?" These are the research questions that will need to be explored as this metric approach is implemented in real-world use cases.

6 Conclusion

In conclusion, the development of autonomous systems must adhere to specific quality criteria. Given that these systems frequently employ artificial intelligence for critical tasks, assessing the stability of complex machine learning models is imperative for comprehension, creation, and improvement. This work aims to lay the groundwork for future endeavors in verifying the stability of such models and their applicability in real-world scenarios.

Conflict of Interest The authors declare no conflict of interest.

References

[1] F. Gawantka, A. Schulz, J. Lässig, F. Just, "SkillDB - An Evaluation on the stability of XAI algorithms for a HR decision support system and the legal context," in 2022 IEEE 21st International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), 183–190, 2022, doi: 10.1109/ICCICC57084.2022.10101657.

[2] H. Löfström, K. Hammar, U. Johansson, "A Meta Survey of Quality Evaluation Criteria in Explanation Methods," in J. De Weerdt, A. Polyvyanyy, editors, Intelligent Information Systems, 55–63, Springer International Publishing, Cham, 2022.

[3] D. V. Carvalho, E. M. Pereira, J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," Electronics, **8**(8), 832, 2019.

[4] M. Nauta, J. Trienes, S. Pathak, E. Nguyen, M. Peters, Y. Schmitt, J. Schlötterer, M. van Keulen, C. Seifert, "From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating explainable ai," ACM Computing Surveys, **55**(13s), 1–42, 2023.

[5] F. Gawantka, F. Just, M. Ullrich, M. Savelyeva, J. Lässig, "Evaluation of XAI Methods in a FinTech Context," in International Workshop on Artificial Intelligence and Pattern Recognition, 143–154, Springer, 2023.

[6] M. Schuld, F. Petruccione, Machine Learning with Quantum Computers, Quantum Science and Technology, Springer International Publishing, 2021.

[7] S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach (4th Edition), Pearson, 2020.

[8] I. Kakogeorgiou, K. Karantzalos, "Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing," International Journal of Applied Earth Observation and Geoinformation, **103**, 102520, 2021, doi:https://doi.org/10.1016/j.jag.2021.102520.

[9] S. Hariharan, R. Rejimol Robinson, R. R. Prasad, C. Thomas, N. Balakrishnan, "XAI for intrusion detection system: comparing explanations based on global and local scope," Journal of Computer Virology and Hacking Techniques, 1–23, 2022.

[10] C.-K. Yeh, C.-Y. Hsieh, A. Suggala, D. I. Inouye, P. K. Ravikumar, "On the (In)fidelity and Sensitivity of Explanations," in H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett, editors, Advances in Neural Information Processing Systems, volume 32, Curran Associates, Inc., 2019.

[11] D. Minh, H. X. Wang, Y. F. Li, T. N. Nguyen, "Explainable artificial intelligence: a comprehensive review," Artificial Intelligence Review, 1–66, 2021.

[12] V. Arya, R. K. E. Bellamy, P. Chen, A. Dhurandhar, M. Hind, S. C. Hoffman, S. Houde, Q. V. Liao, R. Luss, A. Mojsilovic, S. Mourad, P. Pedemonte, R. Raghavendra, J. T. Richards, P. Sattigeri, K. Shanmugam, M. Singh, K. R. Varshney, D. Wei, Y. Zhang, "One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques," CoRR, **abs/1909.03012**, 2019.

[13] Q. V. Liao, M. Singh, Y. Zhang, R. Bellamy, "Introduction to explainable AI," in Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, 1–3, 2021.

[14] A. Das, P. Rad, "Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey," 2020, doi:10.48550/ARXIV.2006.11371.

[15] T. Speith, "A review of taxonomies of explainable artificial intelligence (XAI) methods," in 2022 ACM Conference on Fairness, Accountability, and Transparency, 2239–2250, 2022.

[16] A. Dhurandhar, P.-Y. Chen, R. Luss, C.-C. Tu, P. Ting, K. Shanmugam, P. Das, "Explanations based on the missing: Towards contrastive explanations with pertinent negatives," Advances in neural information processing systems, **31**, 2018.

[17] M. T. Ribeiro, S. Singh, C. Guestrin, ""Why Should I Trust You?": Explaining the Predictions of Any Classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, 1135–1144, Association for Computing Machinery, New York, NY, USA, 2016, doi:10.1145/2939672.2939778.

[18] S. M. Lundberg, S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, 4768–4777, Curran Associates Inc., Red Hook, NY, USA, 2017.

[19] K. Främling, "Decision theory meets explainable AI," in Explainable, Transparent Autonomous Agents and Multi-Agent Systems: Second International Workshop, EXTRAAMAS 2020, Auckland, New Zealand, May 9–13, 2020, Revised Selected Papers 2, 57–74, Springer, 2020.

[20] R. W. Floyd, "Nondeterministic algorithms," Journal of the ACM (JACM), **14**(4), 636–644, 1967.

[21] M. Knap, "Model-Agnostic XAI Models: Benefits, Limitations and Research Directions," 2022.

[22] A. Holzinger, A. Saranti, C. Molnar, P. Biecek, W. Samek, "Explainable AI methods-a brief overview," in International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers, 13–38, Springer, 2020.

[23] C. Molnar, "Interpretable machine learning," https://christophm.github.io/interpretable-ml-book/, 2022, accessed on 04 16, 2023.

[24] R. Heese, S. Mücke, M. Jakobs, T. Gerlach, N. Piatkowski, "Shapley Values with Uncertain Value Functions," in International Symposium on Intelligent Data Analysis, 156–168, Springer, 2023.

[25] K. Främling, "Contextual Importance and Utility: A Theoretical Foundation," in G. Long, X. Yu, S. Wang, editors, AI 2021: Advances in Artificial Intelligence, 117–128, Springer International Publishing, Cham, 2022.

[26] K. Främling, "Explainable AI without Interpretable Model," CoRR, **abs/2009.13996**, 2020.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 9, No. 1, 143-148 (2024)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Investigating Heart Rate Variability Index Classification in Macaca fascicularis and Humans: Exploring Applications for Personal Identification and Anonymization Studies

Daisuke Hirahara[1], Itaru Kaneko[*,2], Junji Nishino[3], Junichiro Hayano[4], Oscar Martinez Mozos[5], Emi Yuda[2]

[1]*Department of AI Research Lab, Harada Academy, 2-54-4 Higashitaniyama, Kagoshima, Kagoshima 891-0113, Japan*

[2]*Tohoku University, GSIS, Sendai, 980-8579, Japan*

[3]*The University of Electro-Communication, Tokyo, 182-8585, Japan*

[4] *Nagoya City University, Nagoya, 467-8601, Japan*

[5]*Örebro University, Institutionen för Naturvetenskap och Teknik, Örebro, 701 82, Sweden*

ARTICLE INFO

Article history:
Received: 09 November, 2023
Revised: 25 January, 2024
Accepted: 26 January, 2024
Online: 22 February, 2024

Keywords:
Macaca fascicularis
Privacy risks
Heart rate patters

ABSTRACT

In this paper, we determine the feasibility of differentiating between the heart rate patterns of Macaca fascicularis and human infants by comparing pertinent hyperparameters. This verification process was undertaken to ascertain the suitability of Macaca fascicularis heart rate data as a testbed for evaluating heart rate parameter privacy safeguarding methodologies. The biological characteristics of Macaca fascicularis bear significant resemblance to those of humans, which consequently renders them useful subjects in medical experiments alongside other laboratory animals. The process of capturing heartbeat data from Macaca fascicularis is notably akin to the methodologies used to record human cardiac activity. In other hand, the recent years have witnessed the construction of extensive heart rate databases, thus raising important considerations surrounding privacy in their usage. Heartbeat recordings, indeed, can provide a wealth of diverse information, necessitating careful handling to maintain data privacy. Specifically, a Holter monitor, a type of electrocardiogram device, can record cardiac electrical activity for over 24 hours. The statistical indices derived from these recordings prove useful for various types of analysis, and simultaneously hold information relating to individual behaviors and health conditions. The extent to which individuals can be identified within such expansive databases is a topic warranting exploration; however, few individuals have granted consent for their data to be used for such research purposes. Given this scenario, since the protection of personal data is not a requisite for Macaca fascicularis, the proposition of employing Macaca fascicularis data to investigate the potential for individual identification appears to be a plausible approach. The experiment verified the similarity of cynomolgus monkey heart rate data to human heart rate data. The results are similar, suggesting that it is appropriate to use cynomolgus monkey heart rate data for personality identification experiments.

1. Introduction

Macaca fascicularis exhibit a significant similarity to humans in terms of biological characteristics, making them suitable subjects in medical experiments alongside other laboratory animals [1] and [2]. Heartbeat data from Macaca fascicularis can also be recorded using methodologies similar to those employed for humans [3]. Concurrently, a large heart rate variability database has been developed in recent years [4], [5], [6] and [7] . Personal identity can be an issue when using large heart rate databases. Therefore, research is needed to verify individual identifiability, which is our overall goal.

[*]Corresponding Author: Itaru Kaneko, Tohoku University, GSIS, Sendai, 980-8579, Japan, kyhsubmit@it-aru.com

Heart rate recordings are rich sources of data. Notably, wearable Holter electrocardiogram devices have the capacity to record cardiac electrical activity over a span exceeding 24 hours. The statistical indices derived from these recordings serve multiple analytical purposes and can also aid in predicting individual behaviors and health conditions.

Analysis of these databases revealed that various personal attributes can be extracted from heart rate [8], [9] and [10]. It is therefore essential to consider the protection of personal data contained in heart rate variability. Therefore, it is important to ascertain the identifiability of heartbeats and explore methods to enhance their anonymization.

Information pertaining to such health conditions, particularly cardiac diseases, is of a private nature and ought to be safeguarded from public disclosure.

Hence, when disseminating such expansive databases for research purposes, it is imperative to apply processing techniques ensuring the heartbeat data contained therein cannot be attributed to a specific individual. Various methodologies have been proposed to hinder individual identification, including anonymization strategies.

There is no absolute method to confirm that personal identification has been entirely thwarted. The effectiveness of identification prevention measures is typically evaluated empirically. Public tests, wherein attempts are made to identify specific individuals using test data, will be carried out, with the resultant outcomes undergoing evaluation. This necessitates the availability of appropriate test data.

Nevertheless, employing human data for such public testing initiatives is untenable due to inherent privacy implications.

The crux of this matter lies in the fact that the data that needs to remain unidentified pertains to personal health characteristics—specifically diseases—that the individual would prefer to keep private. Should the test results lead to the identification of both the disease and the individual, this would pose serious ethical concerns. Contributing data to such tests would bring about substantial risks and minimal benefits to the subjects. Consequently, procuring consent for such use is virtually unfeasible.

Consequently, we propose the use of Macaca fascicularis data for identification trials. Given that there is no requirement for the protection of personal data pertaining to Macaca fascicularis, there exist no ethical issues associated with identifying individual entities or symptoms within this data set.

Nonetheless, if the heart rate data of Macaca fascicularis can be statistically distinguished from that of humans, it would be improper to incorporate the data of Macaca fascicularis into human data and investigate the feasibility of individual identification. This is due to the fact that if we can discern the data of Macaca fascicularis, it would inevitably lead to the recognition of data specific to an individual entity.

Hence, it is crucial to establish that distinguishing between Macaca fascicularis and humans is sufficiently challenging. The primary aim of this study is to undertake this verification.

In essence, within our research purpose, the data of Macaca fascicularis and humans do not have to be entirely indistinguishable. If we select human data that bears similarity to the data of Macaca fascicularis, our objective is to ensure that distinguishing between Macaca fascicularis and humans proves challenging. Should this validation be successful, we can draw the conclusion that the Macaca fascicularis data, which represents the ultimate objective of this study, can be effectively utilized.

2. Proposed method

2.1. Data preparation

Thirty seven neonatal human samples and Macaca fascicularis ECGs, a single case of ECG data obtained from a Macaca fascicularis were utilized for this study.

Thirty seven neonatal Human heartbeat samples were obtaine from ALLSTAR (Allostatic State Mapping by Ambulatory ECG Repository) database, those are all available data in the database. ALLSTAR is database based in Japan (https:/allstar.jpn.org/). The data contained within the ALLSTAR database was gathered utilizing a Holter electrocardiograph, a product of Suzuken Co., Ltd., Japan.

The Macaca fascicularis used in this study was male and it is the only data we could use. The heart rate data for these specimens were collected using a two-lead electrocardiograph provided by Tokyo University of Agriculture and Technology, Japan.

2.2. Analysis Method

A prevalent technique for the detection of R-waves in both the Macaca fascicularis and human neonatal data involves the extraction of baseline fluctuation components, followed by the isolation of feature points through waveform detection processing. For this study, we opted for a device that is capable of automatically calculating RRI.

Utilizing the heart rate variability index as a high-dimensional feature, we sought to determine whether it could be visualized using three representative methods of classification visualization.

Visualization using PCA, t-SNE, UMAP were executed.

PCA (Principal Component Analysis):

PCA stands for Principal Component Analysis and is a classical technique used for dimensionality reduction and information extraction of data. It transforms the data into a low-dimensional space, mainly by finding new axes (principal components) that maximize the variance of the data.

Since PCA extracts features in the range of linear transformations, it is specificly useful in this study to see the relationship between features and original variables in linear transformations.

t-SNE (t-Distributed Stochastic Neighbor Embedding) is a nonlinear dimensionality reduction technique. High-dimensional data can be mapped into a low-dimensional space while maintaining data similarity. This is particularly suitable for visualizing cluster structures. Neighborhoods in high-dimensional space are also mapped to neighborhoods in low-dimensional space.

This transformation not only provides visualization, but also gives insight into how each piece of data is categorized. In this study, it is useful to visualize nonlinear structures.

UMAP (Uniform Manifold Approximation and Projection) is a newer technique than t-SNE. It performs non-linear dimensionality reduction like t-SNE, but makes it possible to capture the overall structure while preserving the local structure of high-dimensional data. It is used for visualization and clustering of high-dimensional data.

It reflects the local and global structures of high-dimensional data in a well-balaned manner. In addition, it is resistant to noise and provides stable and good results.

In this experiment, we expected that PCA could confirm the relationship between the feature value and the original space in the linear transformation relationship, that visualization of the cluster structure by t-SNE and that UMAP would provide a well-balanced view of local and global structures.

Commonly utilized features of heart rate variability were computed and employed for multi-dimensional analysis. These features encompass HR, MNN, SDNN, SDNN, rMSSD, TPW, ULF, VLF, LF, HF, VHF, , and LFHF. These features are widely employed in various heart rate variability analyses[7][8].

3. 3D visualization

3.1. 3D visualization using PCA

Figure 1 presents a three-dimensional visualization created through the principal component analysis (PCA) method. In this visualization, the principal components are derived through PCA, with the three most significant components being transformed into a three-dimensional format. The data points corresponding to Macaca fascicularis are circled in red.

Since PCA is a technique used for converting high-dimensional data into lower dimensions, it amalgamates dimensions in a high-dimensional linear space and extracts the dimension contributing most significantly to the data's variability. Consequently, dimensions that do not illustrate data differences are minimized, allowing for a reduction in dimensions while preserving information vital for data identification to the maximum possible extent.

In the figure 1, Macaca fascicularis is situated peripherally within the overall distribution. Nonetheless, numerous human data points exist in proximity to those of Macaca fascicularis, including closely neighboring human samples, thereby rendering the differentiation between the two challenging.

3.2. 3D visualization using t-SNE

Figure 2 presents the data visualization in three dimensions via the t-SNE method.

Since t-SNE (t-Distributed Stochastic Neighbor Embedding) is a dimensionality reduction technique that is designed to maintain data differences while reducing the dimensionality of high-dimensional data to low dimensionality, a key feature of t-SNE is its inclusion of non-linear transformations, which enhances its performance in data clustering and the preservation of similarity.

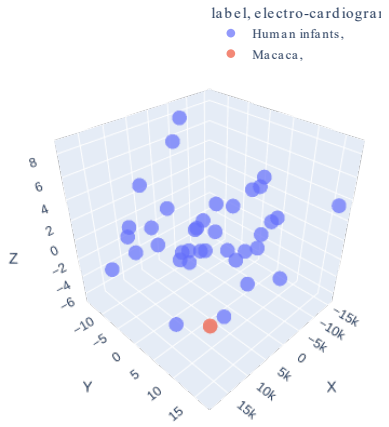


Figure 1: 3D visualization using PCA. 3-dimensional representation is obtained by selecting the top three eigenvectors associated with the largest eigenvalues of the covariance matrix. These eigenvectors capture the most significant variance in the original data, ensuring that the essential information is retained while reducing computational complexity and facilitating data interpretation. The selected 3-dimensional subspace preserves the main patterns and relationships present in the original data, enabling efficient data analysis and visualization. Macaca fascicularis is circled in red. It can be seen that there are human samples in the vicinity and it is difficult to distinguish them. Local structure in the cluster is not clear in case of PCA analysis.

In this analysis, t-SNE compresses from a high dimension to a 3-dimensional space.

This technique performs non-linear transformations to preserve the similarity of data. In other words, data points from the same class are located in close proximity to each other even in a three-dimensional space, while data points from different classes are more distantly spaced. Hence, some degree of clustering can be discerned in the figure, making it easier to understand the relationships between different groups of data in the original high-dimensional space.

Given that t-SNE reduces dimensions while preserving similarities between data points in the high-dimensional space, the figure tends to reflect some relative distances and similarities between the data in the original high-dimensional space.

Moreover, patterns and similarities within the data can be discerned more clearly with t-SNE than with PCA. This is attributed to t-SNE's ability to emphasize local data structures. Data points that were closely situated in high-dimensional space should also tend to be close together in 3D space, allowing local features to be more clearly discerned in the figure.

However, it should be noted that dimensionality reduction by t-SNE does not necessarily preserve the global structure. Therefore, it should be considered that the global arrangement in the 3D space may not necessarily reflect the global arrangement in the higher dimensions.

3.3. 3D visualization using UMAP

Figure 3 shows a three-dimensional visualization by UMAP.

UMAP embeds data in a three-dimensional space while preserving data similarity, resulting in a distribution that reflects the features and clusters of the original high-dimensional data.

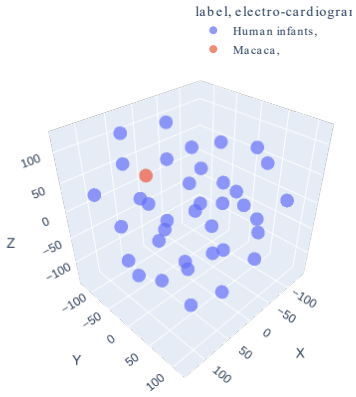


Figure 2: Visualization using t-SNE. Macaca fascicularis is circled in red. It can be seen that there are three or four clusters but local structure of the clusters are punctuated. It can be seen that there are human samples in all around it and it is difficult to distinguish from them.

Similar to t-SNA, data belonging to the same class are plotted close together in 3D space and data of different classes are plotted far apart. As a result, clusters of data are clearly visible in the figure, making it easier to understand the relationships between groups of data.

Also, like t-SNA, UMAP emphasizes local structures, so data that were close together are arranged close together in 3D space. Data that are plotted close together on the figure are likely to be nearby even in a high-dimensional space, making it easier to read the cluster structure.

Unlike t-SNA, UMAP also reflects global data placement to some extent. The structure between clusters can also be read in the figure, but in the case of UMAP, it is thought that this reflects the structure in a high-dimensional space.

This figure seems to be useful for understanding the characteristics of high-dimensional data by performing dimensionality reduction and visualization of data, but it is not true that there is no information loss due to dimensionality reduction.

4. 2D visualizations

Figures 4 to 6 provide two-dimensional (2D) visualizations, achieved through further dimensionality reduction from the previous three-dimensional (3D) representations. While 2D inherently contains fewer dimensions than 3D, this can actually be advantageous when visualizing data. When 3D visualizations are projected onto a 2D plane (such as a computer screen or printed page), some information can be lost due to the lack of depth perception. This could potentially obscure important features or relationships in the data. Therefore, if the 3D visualization isn't explored with a 3D viewer or software that allows for rotation and inspection from different angles, it can sometimes be more insightful to create a 2D visualization from the outset. This can ensure that the most significant data relationships are immediately visible, and are not hidden by the 3D to 2D transformation at the.

4.1. 2D visualization using PCA

Figure 4 presents a two-dimensional (2D) visualization using Principal Component Analysis (PCA). Here, the dimension that contributes the least to the variability of the data has been removed, effectively reducing the data from three dimensions down to two.

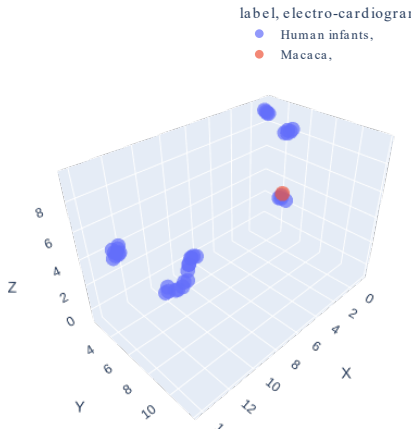


Figure 3: 3D visualization using UMAP. There are three or four clusters and the local structure of those are also visible. Macaca fascicularis is circled in red. It can be seen that it is in one of the clusters and there are human samples in the all-around it and it is difficult to distinguish from them.

The reduction process in PCA is designed to retain the features that account for the most variability in the data. Consequently, this 2D visualization is better suited to display the overall, or global, structure of the data. While PCA is very effective at illustrating these global trends and differences, it's worth noting that it may not accurately portray the more nuanced, local structures within the data set. These may become obscured or lost during the dimensionality reduction process. Thus, while PCA is an invaluable tool for examining broad trends in a data set, it may not fully capture the complexity of the data if there are intricate local patterns or clusters.

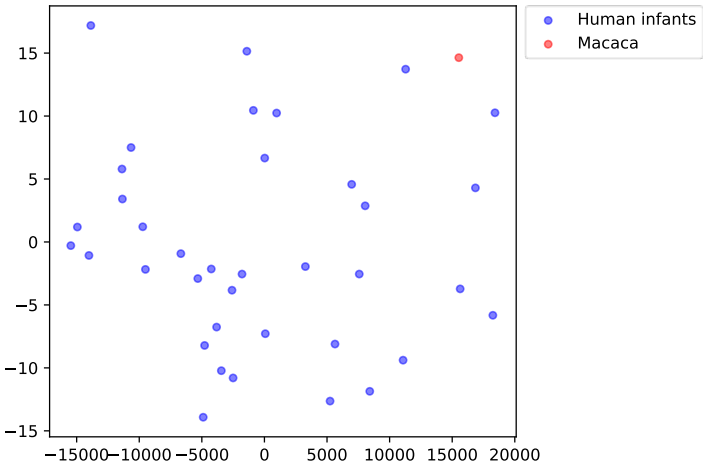


Figure 4: 2D visualization using PCA. Acquire the maximum 2D after PCA and convert it to 2D. Macaca fascicularis is circled in red. It can be seen that there are human samples in the vicinity and it is difficult to distinguish them. Local structure in the cluster is not clear.

4.2. 2D visualization using t-SNE

Figure 5 represents the data's compression into two dimensions using t-SNE (t-Distributed Stochastic Neighbor Embedding). Like in the case of Figure 4, this 2D visualization method emphasizes the preservation of local structures and similarities between data points. Therefore, it provides a more detailed view of the internal structure within each cluster, often producing a more intuitive representation than PCA.

This 2D visualization helps reveal four discernible clusters, with data from Macaca fascicularis appearing within one of these clusters. However, as previously mentioned, t-SNE may not necessarily preserve the global structure of the data. When comparing this t-SNE visualization with the PCA representation, this limitation becomes evident.

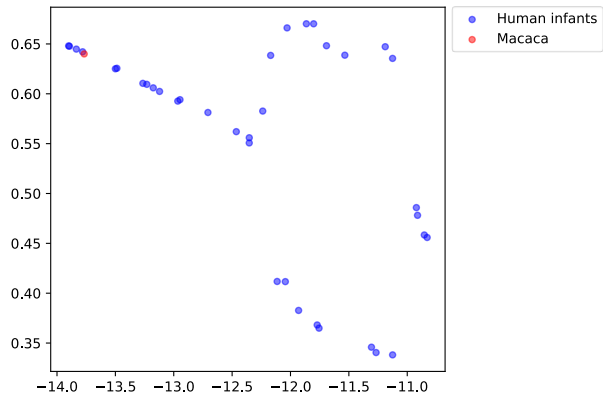


Figure 5: 2D visualization using t-SNE. It can be seen that there are three or four clusters but local structure of the clusters are punctuated. Macaca fascicularis is circled in red. It is difficult to distinguish Macaca fascicularis.

It's important to note that while t-SNE generally requires higher computational resources compared to PCA, in this case, due to the relatively small size of the dataset, this computational cost is not a significant concern.

4.3. 2D visualization using UMAP

Figure 6 illustrates a two-dimensional visualization using the UMAP (Uniform Manifold Approximation and Projection) method. UMAP is particularly adept at preserving both local and global structures in data, leading to robust and informative visualizations.

In this figure, UMAP outperforms t-SNE by capturing intricate local cluster structures in a more comprehensible manner. While t-SNE may cause some clusters to collapse, UMAP is able to differentiate within-cluster data points and maintain clearer demarcations.

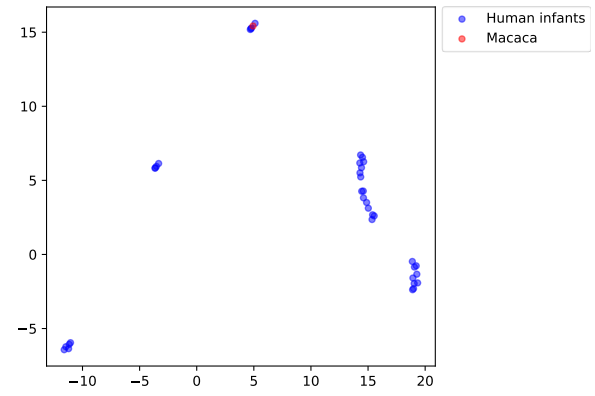


Figure 6: 2D visualization using UMAP. It can be seen that three are three or four clusters and the local structure of those are also visible. Macaca fascicularis is circled in red. It can be seen that it is in one of the clusters and there are human samples in the all-around it and it is difficult to distinguish from them.

Moreover, data points that were dispersed across separate clusters in the t-SNE visualization appear more integrated in this UMAP visualization. Consequently, it's easier to discern three or four distinct clusters. The data points for Macaca fascicularis are situated within one of these clusters, showcasing a fairly even distribution.

Since, UMAP is less sensitive to variations in parameters, allowing the global structure of the visualization to better represent the actual distribution of data in the high-dimensional space, global structure is best visualized in this figure.

5. PCA after UMAP

Figure 7 shows a 2D visualization of UMAP (Uniform Manifold Approximation and Projection) followed by PCA. UMAP is particularly good at preserving both local and global structure in your data, resulting in robust and informative visualizations. PCA can derive the principal factors by linear transformation of the obtained dimensions. By combining these, you can see the structure visualized by UMAP on the space where the factors of the overall structure are orthogonalized in this figure.

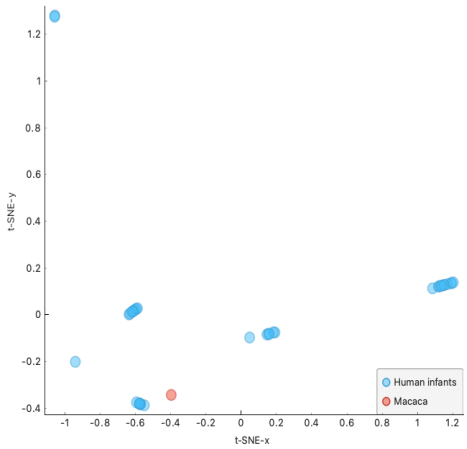


Figure 7: 2D visualization using UMAP (Uniform Manifold Approximation and Projection) followed by PCA. UMAP is preserving local and global structure in data, resulting in robust and informative visualizations. PCA can derive the principal factors by linear transformation of the obtained dimensions.

6. Discussion

Throughout the conducted analyses, our findings consistently validate that the Macaca fascicularis dataset exhibits a distribution pattern that closely aligns with that of the human data, posing difficulties in effectively distinguishing the Macaca fascicularis data. Furthermore, the cynomolgus dataset demonstrates integration within this structure, upon the cluster structure observed in the human data.

Therefore, the findings of this study indicate that straightforward visual features and dimensionality reduction techniques may not be sufficient to differentiate between the heart rate variability indices of Macaca fascicularis and humans when their heart rates are similar. This implies that when the mechanical pulsation mechanism (the mechanical model) of the heart is consistent, the heart rate variability index exhibits analogous characteristics, making it challenging to distinguish between Macaca fascicularis and humans using conventional methods.

These results provide an intriguing insight into the similar physiological patterns shared between certain primates and humans when observed through the specific lens of heart rate

variability indices. Furthermore, they also underscore the need for more nuanced or advanced methods for distinguishing between such closely aligned data sets.

This study might have significant implications for a range of applications where such differentiation could be crucial. For instance, in medical research involving primate models, or in the development of biometric identification systems that rely on heart rate variability as a distinguishing factor. Further research is needed to identify and develop more sophisticated techniques that can make this crucial differentiation more accurately and efficiently.

A clear limitation of this study is the scant amount of data for both Macaca fascicularis and humans. However, the ultimate aim of this research is to utilize Macaca fascicularis data to verify the individuality of human heartbeats. For this purpose, it is not strictly necessary to establish whether there is discrimination between humans and Macaca fascicularis; it is sufficient to demonstrate that the level of discrimination between humans and Macaca fascicularis is considerably low.

If Macaca fascicularis possess traits that humans do not, it's assumed that it would be particularly easy to identify individuals in human heartbeat data. Therefore, it could actually be beneficial for the experiments on individual identification and anonymization. This opens up opportunities for future research and experimentation with larger datasets, more sophisticated analytical techniques, and alternative data sources.

To bolster the reliability of these results, future research will need to gather a larger sample size and consider an approach that incorporates a variety of feature values. Techniques such as data co-mingling and masking are commonly employed to safeguard personally identifiable information, and it's imperative that datasets remain devoid of any personally identifiable information or personal identifiers.

Additionally, integrating different types of machine learning and statistical techniques may also improve the ability to differentiate between the two species. Involving more complex techniques, such as deep learning or support vector machines, could also be considered to help refine the model. Overall, this research provides an interesting starting point, but more extensive work is required to validate these results and to explore the potential implications further.

To bolster the reliability of these findings, our future work will encompass the collection of additional data samples and exploration of a methodology that integrates a variety of feature values. It is crucial to highlight that in the handling of such data, methods such as data commingling and masking are often employed to safeguard personally identifiable information. This ensures that the datasets remain void of personal identifiers and any information that can be traced back to individuals.

In summary, the objective of this experiment was that cynomolgus monkeys are highly similar to humans, and a highly reliable answer was obtained through a multifaceted investigation. In the future, as we acquire more data, we will be able to use these results to advance research toward the broader goal of identifying individual heartbeats.

7. Conclusion

This research substantiated the claim that the cardiac data of Macaca fascicularis are substantially analogous to human heart rhythms and thus can be aptly employed to validate various method intended to prevent personal identification of cardiac data. This research provided highly reliable answer, derived through a multifaceted investigation. The deployment of Macaca fascicularis cardiac rhythm data as a test set for such validation is expected to enable research on more secure processing of cardiac rhythm data. And once the privacy preservation technique for heart rhythm data is possible to be provided, large and efficient use of large-scale shared heart rhythm data repositories will be possible. Clear limitation of this study is the small number of Macaca fascicularis and we need to address this limitation in future research. Even current result is possible to support such result, still it is possible to improve precision and reliability of result by extending data size and analysis method.

References

[1] B. Koo, D. Lee1, P. Kang, K. Jeong, S. Lee, K. Kim, Y. Lee, J. Huh, Y. Kim, S. Park, Y.B. Jin, S. Kim, J. Kim, Y. Son and S. Lee, "Reference values of hematological and biochemical parameters in young-adult cynomolgus monkey (Macaca fascicularis) and rhesus monkey (Macaca mulatta) anesthetized with ketamine hydrochloride," PMID: 32257895 PMCID: PMC7081622, 2019.

[2] Y. Uno, S. Uehara, H. Yamazaki, "Genetic polymorphisms of drug-metabolizing cytochrome P450 enzymes in cynomolgus and rhesus monkeys and common marmosets in preclinical studies for humans," PMID: 29277691, DOI: 10.1186/s42826-019-0006-0.

[3] Y. Noguchi, I. Tawara, K. Kondo, H. Nigi, T. Tanaka, "Electrocardiographic studies in the Japanese monkey (Macaca fuscata) with special reference to the effect of anesthesia with barbiturates," Primates 10, 273–283, 1969, DOI: 10.1007/BF01730348.

[4] S. Aziz, S. Ahmed, M. Alouini, "ECG-based machine-learning algorithms for heartbeat classification," Sci Rep 11, 18738 (2021), DOI: 10.1038/s41598-021-97118-5.

[5] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. Peng, H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," Circulation 101(23):e215-e220, 2000 (June 13), doi: 10.1161/01.CIR.101.23.e215.

[6] E. Yuda, Y. Furukawa, Y.a Yoshida, J. Hayano & ALLSTAR Research Group, "Association Between Regional Difference in Heart Rate Variability and Inter-prefecture Ranking of Healthy Life Expectancy: ALLSTAR Big Data Project in Japan," Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering book series (LNICST,volume 194), 2017.

[7] E. Yuda, M. Kisohara, Y. Yoshida & J. Hayano, "Constituent factors of heart rate variability ALLSTAR big data analysis," Wireless Networks, **28**(3):1287–1292, 2022, DOI: 10.1007/s11276-018-01898-0.

[8] T. Yukishita, K. Lee, S. Kim,Y. Yumoto, A. Kobayashi, T. Shirasawa, "Age and Sex-Dependent Alterations in Heart Rate Variability Profiling the Characteristics of Men and Women in Their 30s," ANTI-AGING MEDICINE. 2010; 7: 94-99, 2010.

[9] R. K. Tripathy, A. Acharya, S. K. Choudhary, "Gender Classification from ECG Signal Analysis using Least Square Support Vector Machine," American Journal of Signal Processing, **2**(5): 145-149, 2012 doi: 10.5923/j.ajsp.20120205.08.

[10] I. Kaneko, J. Hayano, E. Yuda, "How can gender be identified from heart rate data? Evaluation using ALLSTAR heart rate variability big data analysis," BMC Research Notes **16**, Article number: 5, 2023, doi:10.1186/s13104-022-06270-2.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 6, 144-154 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Optimizing the Performance of Network Anomaly Detection Using Bidirectional Long Short-Term Memory (Bi-LSTM) and Over-sampling for Imbalance Network Traffic Data

Toya Acharya*, Annamalai Annamalai, Mohamed F Chouikha

Electrical and Computer Engineering, Prairie View A & M University, Prairie View, Texas,77446, USA

ARTICLE INFO

Article history:
Received: 07 November, 2023
Accepted: 22 December, 2023
Online: 30 December, 2023

Keywords:
Network Anomaly Detection
Sampling
Machine Learning
Deep learning
Bidirectional-LSTM
NSL-KDD
Random Under Sampling (RUS)
Random Over Sampling (ROS)
SMOTE
Data Imbalance

ABSTRACT

Cybercriminal exploits integrity, confidentiality, and availability of information resources. Cyberattacks are typically invisible to the naked eye, even though they target a wide range of our digital assets, such as internet-connected smart devices, computers, and networking devices. Implementing network anomaly detection proves to be an effective method for identifying these malicious activities. The traditional anomaly detection model cannot detect zero-day attacks. Hence, the implementation of the artificial intelligence method overcomes those problems. A specialized model, known as a recurrent neural network (RNN), is specifically crafted to identify and utilize sequential data patterns to forecast upcoming scenarios. The random selection of hyperparameters does not provide an efficient result for the selected dataset. We examined seven distinct optimizers: Nadam, Adam, RMSprop, Adamax, SGD, Adagrad, and Ftrl, with variations in values of batch size, epochs, and the data split ratio. Our goal is to optimize the performance of the bidirectional long short-term memory (Bi-LSTM) anomaly detection model. This optimization resulted in an exceptional network anomaly detection accuracy of 98.52% on the binary NSL-KDD dataset. Sampling techniques deal with the data imbalance problem. Random under-sampling, which involved removing data from the majority classes to create a smaller dataset, was less efficient for deep learning models. In contrast, the Synthetic Minority Oversampling Technique (SMOTE) successfully generated random data related to the minority class, resulting in a balanced NSL-KDD multiclass dataset with 99.83% Bi-LSTM model detection accuracy. Our analysis discovered that our Bidirectional LSTM anomaly detection model outperformed existing anomaly detection models compared to the performance metrics, including precision, f1-score, and accuracy.

1. Introduction

This paper is an extension of "Efficacy of Bidirectional LSTM Model for Network-Based Anomaly Detection" [1] presented at the conference IEEE 13th Symposium on Computer Applications & Industrial Electronics (ISCAIE) in 2023.

Information technology has revolutionized how essential data is conveyed, utilizing bits to transfer a wide range of information from one point to another. This transmitted data can encompass diverse forms, such as voice, images, or data, including sensitive details like banking information, personal records, and network traffic. Numerous tools and techniques are available to identify and thwart unauthorized access.

Anomaly is an unusual pattern present in the dataset. Some techniques or methods are required to detect those anomalies from the dataset. The anomaly is also called the outliers during the study of anomaly detection. Anomaly detection is used in large fields to sense abnormal patterns, such as in business, network attack detection, monitoring health conditions, detecting fraud credit card transactions, and detecting malicious activities in mission-critical systems. Detecting anomalies is critical in cyber security for achieving solid safeguards against cyber criminals. Figure 1. provides brief information about the taxonomy of anomaly detection methods [2].

The security of information resources is ensured when the three fundamental principles of computer security—confidentiality, integrity, and availability (CIA)—are appropriately obeyed [3]. An intrusion detection system is a mechanism used to monitor and scrutinize computer or network- related activities to identify potential threats by assessing the frequency at which computer security guidelines are violated based on confidentiality, integrity, and availability. Intrusion is any unwelcome and illegal activity within an organization's internet-connected end terminal or

*Corresponding Author: Toya Acharya, tacharya@pvamu.edu

https://dx.doi.org/10.25046/aj080614

network-connected devices. These illegal activities aim to gain entry to a business computer or network device. An alternative term for intrusion is a malicious activity that disrupts the fundamental principles of information resource protection known as the CIA triad. An intrusion detection system examines computer network and host activities, pinpointing dubious traffic and abnormalities. Intrusion detection and prevention systems scrutinize traffic from both internal and external sources to identify potentially malicious actions.

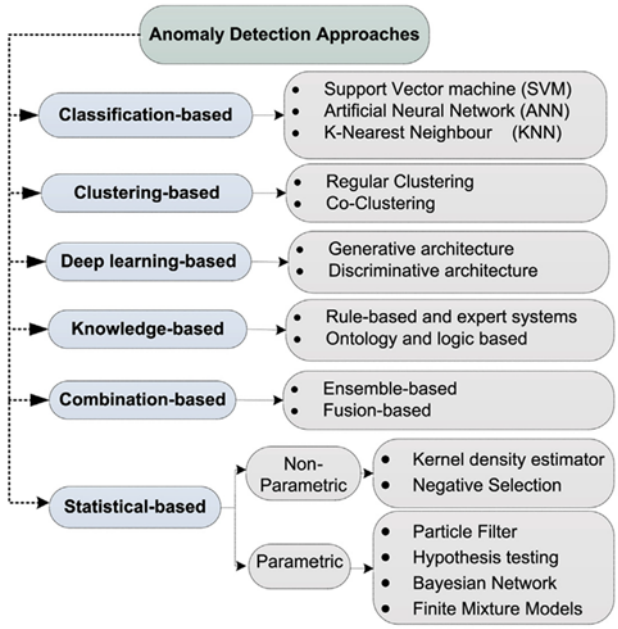


Figure 1: Taxonomy of Anomaly Detection [2]

Detection of misuse and/or intrusion involves identifying potentially suspicious activities within a network or on hosts. Misuse detection focuses on recognizing deviations from established rules by individuals with valid system access rather than actual intrusions. For instance, when an employee uses the Internet for personal purposes in violation of company policy, it constitutes a misuse intrusion. In contrast, intrusion detection is designed to identify unauthorized individuals, such as external hackers or government spies, who lack authorized system access. The intrusion detection system primarily focuses on spotting ongoing intrusions within the system or network but does not proactively prevent malicious activities.

There are two main types of methods for detecting intrusions: signature-based systems, known as SIDS, and anomaly-based systems, referred to as AIDS. Anomaly detection systems are further categorized into network-related and host-related intrusion detection systems. The identification of normal or anomalous data in anomaly detection techniques is achieved through the utilization of labels.

A SIDS identifies suspicious activities through pattern matching with known external attack patterns, which fall into two categories: misuse detection and knowledge-based detection. This anomaly detection model compares the recent signature with a previously stored signature in its database. When a match occurs between these signatures, the IDS signals the presence of malicious activities within the network. Regular updates to the signature database are crucial to effectively detect malicious activities in a network. However, it's important to note that this type of detection system cannot identify zero-day attacks, as these novel attack types may not yet be contained in the signature archive. This anomaly detection model delivers optimal detection outcomes for recognized signatures associated with malicious activities. This type of anomaly detection model is known for its straightforward configuration and comprehensibility. Widely adopted intrusion detection systems include Snort and NetSTAT. In the traditional setup, the SIDS examines network packets and matches them against stored signatures. However, newly introduced attacks not yet included in the signature database can reduce intrusion detection accuracy. To address these limitations, the anomaly-based anomaly detection model offers enhancements and boosts the overall anomaly detection rate. The anomaly-based intrusion detection approach is designed to identify malicious and unreliable network exploitation activities within the corporation.

AIDS effectively addresses the limitations included in SIDS approaches. Likewise, the anomaly-based intrusion model leverages statistical-based, machine learning, and knowledge-based techniques to model the typical behaviors of network traffic. An "anomaly" refers to any behavior that deviates from these established norms, and such traffic anomalies can harm computers and network devices. Anomaly-based detection can occasionally yield false results due to shifts in user behavior. AIDS can generate errors even when legitimate users alter their usual habits. This approach comprises two key stages: the testing and the training stages. In the training stage, the model is trained using normal traffic data to establish a baseline or "normal profile." In the testing stage, previously unseen data is employed to evaluate the model's performance. The primary benefit of this methodology is its ability to detect zero-day attacks.

Three distinct anomaly detection methods include unsupervised, semi-supervised, and supervised anomaly detection methods based on the target class. AIDS addresses the limitations of SIDS by employing knowledge-based methods, machine learning, and statistical-based to model normal behaviors. Figure 1 outlines the various approaches to anomaly detection [2].

Deep learning has the capability to generate improved representations, enhancing the development of effective anomaly detection models. In contrast, conventional machine learning algorithms for network-related abnormality detection are more appropriate for smaller datasets and often rely on performance influenced by the implementation of feature engineering. The model benchmark indicators of conventional anomaly detection models are significantly influenced by the split ratio. While these conventional ML methods are uncomplicated and require minimal resources, they face limitations when dealing with extensive datasets and large feature sets, making them unsuitable for tasks such as machine vision, image translation, natural language processing, and similar applications.

The convolutional neural network is primarily employed for computer vision using image datasets, with the lower layers' neurons responsible for feature reduction. These lower layers typically recognize image corners, boundaries, and intensity called small-scale features. As the information progresses to higher

layers, the network integrates these lower-level features to create forms, basic shapes, and partial objects. The final layer of the network amalgamates these lower-level features to generate the output. LSTM operates distinctively from a CNN as it is commonly applied for processing and predicting outcomes based on sequential data. Unlike CNNs, Recurrent Neural Networks (RNNs), including LSTMs, were specifically designed to preserve long-range information within a sequence, preventing the loss of important details in lengthy sequences. The Bidirectional LSTM (BiLSTM) enhances this by introducing an additional LSTM layer that reverses the flow of information, effectively addressing issues such as vanishing gradients.

The deep learning methodology tackles challenges inherent in conventional machine learning, specifically its ability to handle extensive datasets and numerous features. The efficacy of anomaly detection algorithms based on deep learning depends on various factors, including the choice of hidden layers, determination of activation function, neurons, batch size, and epochs during both model training and testing. Strategic decisions regarding these hyperparameters, along with considerations for the ratio of the train to test data and the design of deep neural networks, are essential for improving the precision of network anomaly detection systems.

In addition to fine-tuning hyperparameters, handling imbalanced data is vital, and creating a balanced dataset using various sampling methods contributes to improved anomaly detection. Under-sampling reduces data size, posing challenges for deep learning models. At the same time, over-sampling methods generate duplicate random data, proving more effective for deep learning models to improve the anomaly detection performance in network-based anomaly detection models.

2. Literature Review

The continuous generation of data generates big data and poses challenges for traditional machine learning algorithms, requiring extensive feature engineering efforts to perform adequately. Deep learning significantly enhances detection performance in such scenarios. However, the effectiveness of network anomaly detection varies on numerous factors, with the nature of the dataset (whether balanced or unbalanced), the hyperparameter of the neural network, the amount of model train and test data, and the architecture of the neural network in the deep learning model. These elements collectively play a crucial role in successfully identifying anomalies in the network.

In their study, the researchers utilized the Bidirectional LSTM to alleviate the considerable requirements for feature reduction inherent in conventional machine learning-based anomaly detection approaches [4]. Additionally, they implemented data augmentation in data preprocessing of minor attacks user to root (U2R) and root to local (R2L) to create a well-adjusted NSL-KDD. This methodology resulted in higher anomaly detection accuracy of 90.73% and f1-scores of 89.65%.

In their study, presented an algorithm for network intrusion detection that integrated a deep hierarchical network with hybrid sampling, incorporating SMOTE to create a balanced dataset [5]. They utilized a hybrid approach that combined CNN and Bi-LSTM for anomaly detection accuracy of 83.58% on NSL-KDD and 77.16% on UNSW-NB15.

In their study, presented a method utilizing bidirectional generative adversarial networks (Bi-GAN) on the CIC-DDoS2019 and NSL-KDD datasets [6]. The Bi-GAN model exhibited strong performance, particularly on the imbalanced NSL-KDD, achieving an f1-score of 92.68% and an accuracy of 91.12%. The Bi-GAN approach was employed to enhance the performance of the NSL-KDD imbalance dataset.

In their study, implemented a new method involving auxiliary classifier generative adversarial network (ACGAN) and ACGAN with SVM to tackle data unevenness concerns by leveraging GAN to produce synthetic attack network traffic for intrusion detection systems [7]. These artificially generated attacks were merged with the existing data, resulting in an extended dataset. Research carried out on the RAWDATA, CICIDS2017, UNSW-NB15, and NSL-KDD showed that among the support vector machine, decision tree, and random forest models, the decision tree achieved a superior f1-score of 92% on the balanced NSL-KDD dataset.

In [8], researchers utilized an assorted ensemble-aided approach to binary and multi-class network anomaly detection models to tackle the challenge of uneven traffic data in network traffic-related datasets, including NSL-KDD, UNSW-NB15, and KDD99 datasets. This approach achieved a true positive rate and area under the ROC curve of 94.5% and 96.2% on the NSL-KDD, respectively.

According to their finding, the authors [9] concluded that the efficiency of the anomaly detection algorithm is improved when the number of output labels is reduced. This observation was explored across different conventional machine learning algorithms, including Naïve Bayes, J48, random forest, bayesinNet, bagging, and bayesinNet. The evaluation used three network datasets: KDD99, CICIDS2017_Thrusday, and UNSW-NB15.

In [10], the researchers observed the effectiveness of a recurrent neural network-based intrusion detection system (RNN-IDS) in multi-class and binary-class scenarios. Performance on the NSL-KDD was observed, which is affected by the number of neurons and different learning rates. Experimental outcomes illustrated that RNN-IDS is adept at constructing a classification approach with high accuracy, outperforming traditional machine learning classification methods, including random forest, artificial neural network, J48, and support vector machine in both multiclass and binary network intrusion-related datasets. In their publication [11], presented a network anomaly detection technique utilizing a convolutional autoencoder and attained a model accuracy of 96.87% on the NSL-KDD. The convolutional autoencoder methodology was utilized to simplify and determine the most significant features of the network anomaly dataset.

In [12], the authors investigated the usefulness of several autoencoders in detecting network anomalies. They compared four different types of autoencoders, including sparse autoencoders, undercomplete deep autoencoders, and denoising autoencoders, using the NSL-KDD. Sparse deep denoising autoencoder yielded a model accuracy of 89.34% compared with other models.

In [13], the authors presented a model centered around a 5-layer autoencoder (AE) tailored for network abnormality detection. The fine-tuned model designs demonstrated proficiency in attribute learning and the dimension of data reduction, resulting in improved performance metrics, including model accuracy and f1-score. The model produces the highest accuracy and f1-score of 90.61% and 92.26% on the NSL-KDD, respectively. The researchers employed the reconstruction error to determine whether the network traffic is regular or attacked.

In [14], [15], the researchers proposed a network anomaly detection approach with a combination of convolutional neural networks and bidirectional LSTM applied to the KDD99. They explored the influence of the number of nodes, the number of hidden layers and memory elements on it, and the number of epochs to improve their anomaly detection model accuracy. The performance metrics of different models, such as J48, k-nearest neighbors, NB, deep forest, RF, and convolutional neural network combined with bidirectional LSTM, were evaluated. The convolutional neural network bidirectional LSTM exhibited the ultimate model detection accuracy of 95.40%.

In [16], the researchers assessed both single-layer and four-layer LSTM models for weather forecasting, utilizing a weather-related dataset from Hang Nadim Indonesia Airport. The top model validation accuracy of 80.60%. The four hidden layers comprise 50, 90, 100, and 200 memory elements. The split ratio for testing and training dataset was used at 0.30, and the models underwent training for 500 epochs.

The researchers in [17] adopted a deep learning approach utilizing bidirectional LSTM, implemented on the UNSW-NB15 and KDDCUP99, and achieved notable outcomes with a 99% accuracy rate for both datasets. Numerous current models encounter difficulties in effectively detecting uncommon attack traffic types, notably user-to-root and remote-to-local traffic, which often demonstrate lower detection accuracy than other types of attacks. The researchers in [18] deployed an intrusion detection system based on bidirectional LSTM to address the mentioned encounters on the NSL-KDD. This anomaly detection approach, utilizing Bi-LSTM, achieved a model detection accuracy of 94.26 % for binary NSL-KDD data.

The study in [19] delved into the influence of batch size and learning rates on the performance of CNN, focusing on image classification, particularly in the context of medical images. The results indicate that a larger batch size does not necessarily lead to higher accuracy. Moreover, the choice of learning rate and optimizer significantly affects performance. The authors found that reducing the learning rate and batch size, particularly during fine-tuning, enhances the network's training effectiveness.

Diverse strategies were implemented to address the challenge of data imbalance, encompassing techniques such as data augmentation discussed in [4], application of SMOTE detailed in [5], the use of GAN technology explored in [6], [7], the assistance of Heterogeneous ensemble methods investigated in [8], and the reduction of the target class by combining smaller classes into a new category as discussed in [9]. A considerable number of research endeavors in the realm of deep learning for network anomaly detection have been scrutinized, incorporating methodologies like RNNIDS outlined in [10], CAE featured in [11], Autoencoder examined in [12], multilayer AE explored in [13], convolutional neural network combined with bidirectional LSTM hybrid methods presented in [14] and Bi-LSTM discussed in [17]-[18].

The researchers in [17] and [18] did not provide details on data pre-processing, the train-test split ratio, or adopting bidirectional LSTM hyper-parameters in their model study. Similarly, the researchers in [16] conducted weather forecasting using Bi-LSTM without specifying the hyperparameter values. In [10], there was no analysis information on epochs and the train test split ratio for the KDDTrain+ dataset. Most of the literature reviewed emphasizes enhancing model accuracy in conventional or deep learning algorithms. However, there is a notable lack of focus on deciding on hyper-parameters in deep learning approaches, determining the train test split ratio, and defining the architecture of neural networks. Some researchers do not clarify how these values are applied in their work. Consequently, our research aims to address these limitations in the network anomaly detection approaches by conducting experiments on NSL-KDD.

3. Contributions

The literature review examines a gap in the existing network intrusion detection systems during anomaly detection. The primary contribution of this research is to bridge this gap by proposing network anomaly detection models specifically tailored for imbalanced multiclass datasets.

Arbitrarily selection of hyperparameters does not yield efficient anomaly detection performance on the given dataset. This research investigates the impact of epochs, batch size, and optimizers on the efficacy of a bidirectional LSTM anomaly detection model using the multiclass NSL-KDD.

The choice of the amount of training data and testing data also influences the model's performance. A larger training dataset requires a longer training time, whereas a smaller dataset leads to quicker model training. The model's efficacy is contingent on the data size utilized for both training and testing, a factor we explored by adjusting the test train split ratio to enhance the performance of network traffic anomaly detection on the NSL-KDD.

More layers add complexity to the neural network-based model. The program execution time (program training and testing time) is large compared to small numbers of neural network layers and memory elements. The memory elements and layers in the neural network architecture influence the network anomaly detection performance. Investing layers and memory elements of neural networks improves the bidirectional LSTM on the NSL-KDD.

The careful choice of machine learning and/or deep learning algorithms significantly impacts the effectiveness of network anomaly detection. This study introduces the creation and deployment of a network traffic anomaly detection system utilizing a bidirectional LSTM-based recurrent neural network model. The developed model demonstrates a remarkable anomaly detection accuracy of 98.52 % in the network, particularly for the binary NSL-KDD.

The primary challenge when dealing with real datasets is the presence of imbalanced data. Various approaches can be employed

to address this issue. In the NIDS multiclass dataset, both under-sampling and over-sampling methods are applied to tackle data imbalance. Notably, oversampling methods proved to be more effective, achieving the highest detection accuracy of 99.83% for the multiclass NSL-KDD datasets.

4. Model Description

The proposed model consists of different steps, which are listed:

1. Data collection and modelling
2. Data cleaning and pre-processing
3. Bidirectional LSTM model preparation
4. Model training and testing
5. Evaluation model
6. Compare the model for decision

Figure 3 illustrates the schematic for the model based on Bidirectional LSTM. More elaborate explanations of the methods outlined above for the proposed model will be provided in the following sections.

4.1. Data Collection and Modelling

During this study, we employed the KDDTrain+ dataset, one of the subset data from the NSL-KDD.

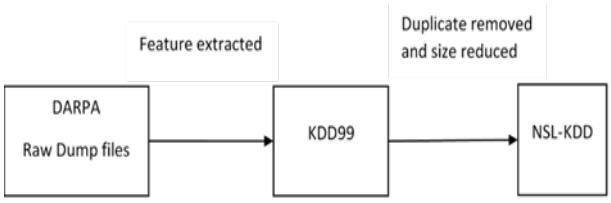


Figure 2: DARPA, KDD99, and NSL-KDD Datasets

The NSL-KDD data is derived from the DARPA KDD99 data, as depicted in Figure 2, after the removal of noise and unwanted data. This includes the complete training data from the NSL-KDD set, including features named attack_type and difficulty. It encompasses 41 attributes and covers five separate attack categories: denial of service, normal, remote_to_local, probe, and user_to_root. NSL-KDD [20] represents an enhanced version of the KDD99 network traffic anomaly data, eliminating duplicate entries in the training data and ensuring the absence of repeated records in the test data. The KDDTrain+ dataset comprises 125,973 records and includes 41 attributes. Notably, this is balanced, with 53.46% of total traffic being normal and 46.54% of total traffic entry being abnormal. We picked this data because it is balanced data between normal and abnormal traffic records within the subset, making it suitable for binary network anomaly detection data. Those numbers of attack class information from the NSL-KDD data were utilized to create the multiclass dataset for the experiment, detailed in the data pre-processing section.

4.2. Data Cleaning and pre-processing

The KDDCup99 data is widely employed in experiments related to anomaly detection in computer network traffic. It consists of network-related traffic that transfers from the virtual network environment utilized for the third knowledge discovery and data mining tools competition. The KDD99 network traffic data is a revision of the 1998 DARPA. The KDDCup99 dataset comprises three components: the "Whole" dataset, the "10% KDD," and the "Corrected KDD." The "Whole" dataset encompasses various attack traffic and one normal network traffic connection. This data involves two training data subsets: a full training data subset and a 10% training data subset. The "Whole" dataset consists of 4,898,431 individual records containing 41 attributes labeled as normal or an attack.

As indicated in reference [20], the KDD99 dataset encompasses 22 distinct attack traffic categorized into four classes: Denial of Service, Unauthorized Access to Local Privileges (U2R), Unauthorized Remote Machine Access (R2L), and Scan Network (Probe). The NSL-KDD data contains four sub-datasets, including KDDTest-21, KDDTest+, KDDTrain+_20Percent, and KDDTrain+. Notably, the KDDTrain+_20Percent and KDDTest-21 portions are sub-datasets derived from the KDDTest+ and KDDTrain+, respectively.

The KDDTrain+ dataset is designated as the training dataset, while the KDDTest+ dataset serves as the testing dataset for the machine learning model. KDDTest-21 is a subset of the test dataset that excludes the most challenging traffic records (with a score of 21), and KDDTrain+_20Percent is a subset of the training dataset, encompassing 20% of the entire training dataset. It's important to note that the traffic records found in KDDTest-21 and KDDTrain+_20Percent are already included in the test and train datasets, respectively. The NSL-KDD dataset addresses the limitations found in the KDD'99 dataset. Unlike KDD'99, NSL-KDD ensures the absence of redundant values in both the train and test datasets.

Notably, NSL-KDD is advantageous due to its smaller test and train sets, eliminating the need for random selection of a small data subset, thus making experiments more cost-effective. Each record in the NSL-KDD dataset comprises 42 features, with 41 of them corresponding to the traffic input and the final label denoted as either "normal" or "abnormal." In the KDDTrain+ contains 125,973 total network traffic records and 41 generated attributes, the data cleaning and pre-processing assigns a target label of ‘1’ for normal traffic and ‘0’ for attack traffic records, transforming the multiclass network traffic data into a binary class.

achine learning and deep learning algorithms work only for numeric values, so ‘'protocol_type,' 'service,' and 'flag' are categorical attributes transformed into numeric values, either ‘0’ or ‘1’ using one hot encoding method called dummy one hot encoding. The dataset is then normalized using the standard scalar method. Correlation-based feature reduction is also implemented where those features with a correlation factor exceeding 0.5 are preserved to reduce the features. Binary class data is employed in experiments A to E. In experiment F, the multiclass (class 5) version of the NSL-KDD dataset is utilized. Prior to training and testing the BI-LSTM model, a sampling method is applied to balance the unbalanced multiclass data. Further details on data preprocessing and model information can be found in the experimental section in the subsequent chapter.

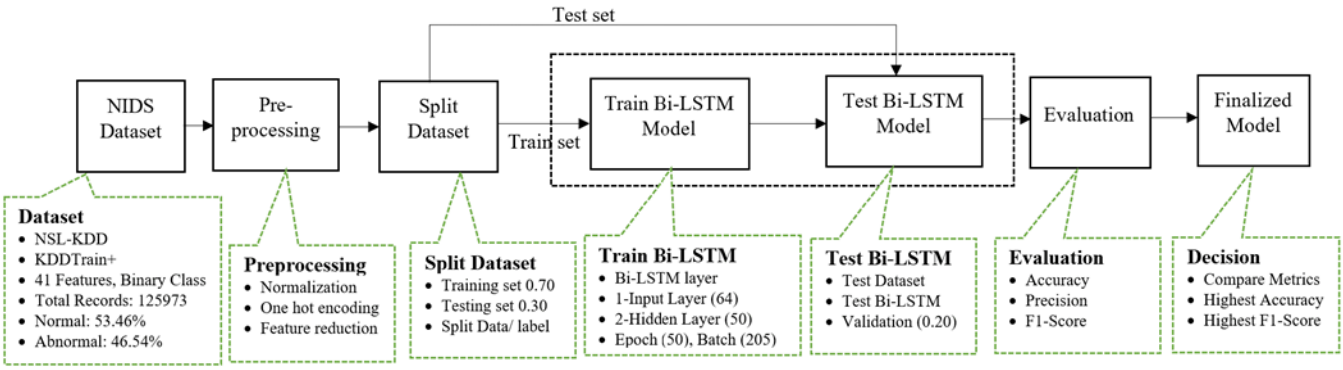


Figure 3: Bidirectional LSTM model block diagram

4.3. Train and test data preparation

The train data and test data splitting method separate the data randomly into two different subsets of the dataset. These two subsets of data contain the designed amount of data based on our selection. Since our pre-processed dataset represents just one portion of the data, we employ two separate datasets for implementing the machine learning algorithms. Researchers typically have flexibility in determining the train-test split ratio, with common choices of 80% to 20%, 60% to 40%, 70% to 30%, and 75% to 25%. We conducted experiments to determine our model's most optimal splitting ratio and found that a 70% training and 30% testing dataset ratio yielded the best performance.

4.4. Bidirectional LSTM model preparation

A recurrent neural network comprises feedback paths that analyze data sequences and patterns to make predictions. These loops enable data sharing among nodes, facilitating predictions based on accumulated information referred to as memory. RNNs have been effectively applied to address machine learning challenges, including tasks such as language preprocessing models, human voice/ speech recognition, and image processing.

The LSTM-based model resolves the challenge of vanishing gradients encountered in RNNs. The LSTM architecture comprises a memory block and three units: input gates, output gates, and forget gates. These gates function similarly to read, write, and reset functions for the cells. Due to the presence of those three gates, LSTM memory cells can effectively store and retrieve data over prolonged times, mitigating the issue of vanishing gradients.

Conventional RNNs are limited in their capacity only to consider past context information. In contrast, Bidirectional RNNs overcome this constraint by analyzing data in forward (left to right) direction and backward (right to left) directions. This involves integrating two hidden layers, with the outcomes subsequently forwarded to a shared output layer. In a conventional LSTM neural network, the output signal/data is generated directly. In contrast, a bidirectional LSTM neural network incorporates both directions (forward and backward) layers at each stage, contributing the signal to the neural network activation layer. This configuration captures data from both preceding and succeeding data, allowing the bidirectional LSTM neural network model to predict the target sequence of each element by considering finite sequences in the circumstances of both past and future elements. This is achieved by employing two consecutive LSTMs—one processing data from both directions. Traditional RNNs are constrained by their dependence solely on the previous perspective. Bidirectional LSTM defeats this limitation by examining data feed from both directions through two hidden neural network layers and then forwarding the results to a similar recurrent neural network output layer.

In a standard LSTM-based model, the model prediction is usually obtained directly via the given dataset. Conversely, the outputs from the forward layers and backward layers from each stage are combined and input into the activation layer in the bidirectional LSTM model. This resulting output encapsulates data from past and future data from the memory blocks in LSTM. The bidirectional LSTM predicts the labels or sequence from each element by leveraging finite sequences within the circumstances of preceding and following items. This process is accomplished through the sequential processing of two LSTMs—one data sequence from right to left direction and the same data sequence from left to right.

The selection of neural network architecture components, including input layers, hidden layers, output layers, layer sizes, activation functions, and dropout rates, is a critical step following data preprocessing. Hyperparameter tuning is an integral part of this research. Initially, hyperparameters are chosen randomly for experimentation, as discussed in more detail in the subsequent experimental sections. The data sampling approaches are implemented to deal with the data unevenness problem. Random oversampling and random under-sampling methods created the balanced multiclass dataset.

To initiate the random selection of the Bidirectional LSTM architecture, the neural network comprises a single input layer with 64 neurons and a dropout rate of 20%. It features two hidden layers with 50 neurons each, both employing a 20% dropout rate. The output layers consist of a single dense layer, and the choice of activation function depends on the nature of the target class size, whether binary or multi-class. Once this model is defined, it is compiled using the appropriate loss function and optimizer in preparation for training.

4.5. Evaluation Bi-LSTM model

Multiple experiments have been conducted to analyze the efficacy of the bidirectional LSTM model, revealing inconsistencies in the effectiveness of both machine learning and deep learning models. Consequently, a comprehensive analysis of the model's hyperparameters becomes imperative for performance

improvement. The selection of the optimizer, batch size, epochs, and train test splitting ratio is guided by a comparison of anomaly detection accuracy and f1-score metrics for the bidirectional LSTM model. Ultimately, the bidirectional LSTM's performance metrics are juxtaposed with previous research findings to assess its efficacy. Additionally, two distinct sampling methods, namely random under-sampling and random oversampling, were experimented with on the NSL-KDD and compared using the bidirectional LSTM model.

4.6. Compare performance for decision-making.

After conducting model testing and evaluation, the decision-making process involves selecting the most suitable model pipeline from various alternatives. During this research, multiple sets of experiments are conducted to optimize the hyperparameters for the Bi-LSTM model, aiming to enhance its performance. These hyperparameters encompass factors such as optimizers, epoch count, batch size, neural network architecture, class size selection, and methods for preprocessing raw data. This optimization process is driven by comparing performance metrics obtained from these diverse sets of experiments. Additionally, the performance metrics of the bidirectional LSTM anomaly detection models for NSL-KDD data are compared with published literature results.

5. Experiments and Results

Sets of experiments were conducted on a Windows 10 laptop with a 64-bit architecture, equipped with 16GB of random-access memory and an i7-1.99GHz processing unit. Python3.7.13, Keras2.6.0, and TensorFlow2.9.1 were utilized in this research. The investigation into train and test data split ratio, numbers of epochs, optimizers, and batch size for the bidirectional LSTM model was carried out across various experiments, as elaborated below. The intrusion detection system leverages machine and deep learning techniques for anomaly detection. Python is utilized to code network intrusion detection models, using packages such as NumPy, Pandas, Keras, imblearn, and Sci-kit-learn for developing machine learning models. Additionally, tools like WEKA, Java, C#, Visual C++, and MATLAB are commonly employed in intrusion detection.

To ensure reproducibility, seed values are configured to obtain consistent results across multiple runs on the Jupyter Notebook platform. Subsequently, the experimental results are presented in the form of plots or tables, using the Microsoft Office suite for analysis.

5.1. Experiment: Optimizers Vs. Bi-LSTM performance

During this experimentation, the bidirectional LSTM was applied to the NSL-KDD, the details of which are outlined in the preceding sections. An appropriate optimizer is essential for enhancing the network traffic anomaly detection model's training time and the overall efficacy of the model. The choice of optimizer holds significant importance as it expedites results for the ML/DL model. The choice of the optimization algorithm made by a deep learning practitioner directly impacts both the training speed and the ultimate predictive performance of their model. TensorFlow is an open-source machine-learning library containing nine optimizers: Adam, Ftrl, Adagrad, Adamax, Adadelta, SGD, RMSProp, gradient descent, and Nadam. Among them, seven optimizers were experimented with to achieve the highest performance of the model.

Table 1: Optimizer Vs. Accuracy

training data= 70%, Epochs = 50, batch size= 512				
SN	Optimizer	Accuracy %	Precision %	f1-score %
1	**Nadam**	**98.26**	**97.76**	**98.37**
2	Adam	98.24	97.66	98.35
3	RMSprop	98.19	97.56	98.31
4	Adamax	97.95	97.40	98.08
5	SGD	91.19	88.67	92.02
6	Adagrad	61.86	58.22	73.59
7	Ftrl	53.14	53.14	69.40

In this experimental task, the hyperparameter values were picked randomly, and the performance metrics and optimizers are outlined in Table 1. The structure of the bidirectional LSTM model contained 64 units, featuring two B-LSTM hidden layers having 50 units in each, along with the dense output layer. Each layer within the BLSTM model utilized an activation function called relu and 20% drop-out rate of 20%.

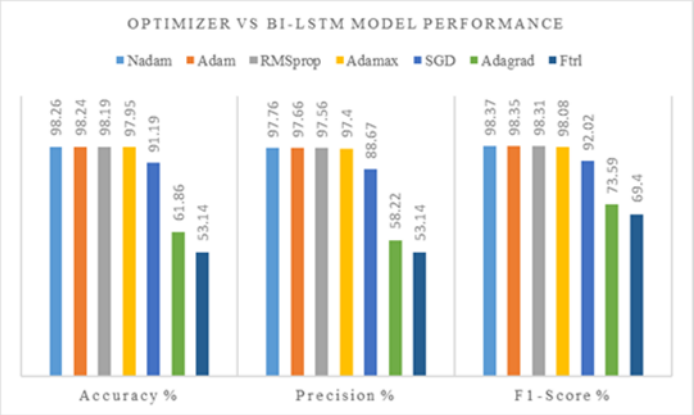


Figure 4: Optimizer Vs. Bi-LSTM performance

Observing the above results (Table 1 and Figure 4), it is determined that the Nadam optimizer is the victorious optimizer, with the winning performance metrics having an accuracy of 98.26%, precision of 97.76%, and f1-score of 98.37%. Nadam enhances the Adam algorithm by integrating Nesterov momentum, resulting in an improved performance of the Adam optimizer.

5.2. Experiment: Train test split ratio Vs. performance

In this experiment, we investigated the impact of both the train test split ratio and model performances. The process of data splitting is crucial in data science, particularly when preparing machine learning models using the available data.

The train test split methodology is utilized to calculate the efficiency of machine learning algorithms in predicting results from data that were unseen during the model training phase. Once the model gets trained, the test dataset is applied, and no fixed percentage split ratio to divide into training and test sets from the given dataset. The splitting ratio is explored to enhance the model performance by utilizing the Nadam optimizer on binary NSL-KDD data.

Table 2: Train test split ratio Vs. performance

optimizer = Nadam, Epochs = 50, Batch_size = 512			
Testing data %	accuracy %	precision %	f1-score %
10	98.15	97.55	98.29
20	98.21	97.57	98.33
30	**98.24**	**97.66**	**98.36**
40	98.18	97.52	98.30
50	98.13	97.50	98.26
60	98.10	97.52	98.24
70	98.12	97.65	98.25
80	97.82	97.39	97.97
90	97.92	97.31	98.07

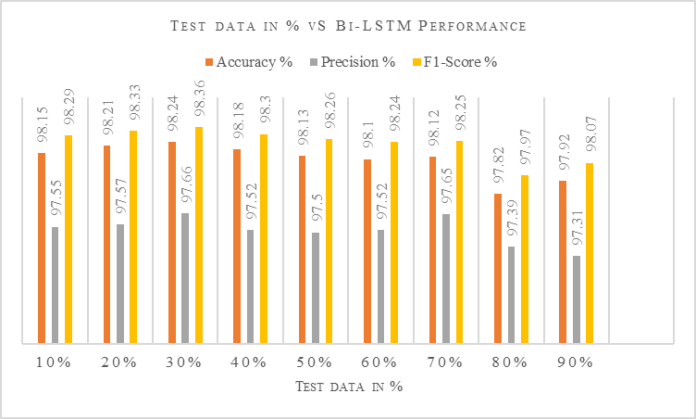


Figure 5: Test data size in % Vs. Bi-LSTM model performance

This experimental work presents the train test split ratio that achieves the optimal performance for our network traffic anomaly detection model on the NSL-KDD. The performances are tabulated in Table 2. and the plot is shown in Figure 5, where a 30% test split percentage results in the model's highest accuracy of 98.48% and f1-score of 98.57%.

5.3. Experiment: Batch size Vs. performance

This experimental work presents the train test split ratio that achieves the optimal performance for our network traffic anomaly detection model on the NSL-KDD. The performances are tabulated in Table 2. and the plot is shown in Figure 5, where a 30% test split percentage results in the model's highest accuracy of 98.48% and f1-score of 98.57%.

Table 3: Batch size Vs. model performance

Optimizer = Nadam, epochs = 105, testing data split= 0.30			
batch size	**f1-score %**	**accuracy %**	**prgm exe time (sec)**
50	**98.58**	**98.48**	2127.235
500	98.47	98.36	514.770
350	98.51	98.4	527.153
450	98.51	98.41	454.989
250	98.46	98.35	616.466
150	98.52	98.42	858.070
300	98.55	98.45	553.444
200	98.55	98.45	796.898
400	98.48	98.38	460.884
15	98.56	98.45	5671.738
100	98.56	98.46	1228.779

A smaller batch size entails the introduction of limited data samples into the Bi-LSTM anomaly detection model, necessitating a lengthier training period than a larger batch. The performance metrics and batch size are presented in Table 3. The experimented results indicate that when applying this model to the NSL-KDD, a batch of 50 produces optimal accuracy and s1-score. A larger batch of data through the model takes less training time but exhibits lower accuracy, highlighting a significant trade-off for this Bi-LSTM network traffic anomaly detection model.

5.4. Experiment: Epochs Vs. performance

In machine learning, an epoch represents one complete pass through all the training data during a model's training. During each epoch, the model is exposed to the entire dataset, and the model's parameters (weights and biases) are adjusted based on the error or loss calculated from the model's predictions compared to the actual target values.

Table 4: Epochs Vs. model performance

optimizer = Nadam, batch= 50, test data= 30% , train data = 70%			
epoch	accuracy	f1-Score	prgm exe time (sec)
175	98.48	98.58	3965.207
100	98.48	98.58	1878.803
125	98.48	98.58	2470.620
5	97.9	98.03	127.058
35	98.35	98.46	761.278
205	**98.52**	**98.62**	4103.767
50	98.38	98.48	942.129
45	98.37	98.47	1002.092
75	98.46	98.56	1465.514
150	98.48	98.58	2934.249
25	98.3	98.41	527.524
15	98.13	98.25	322.529
Accuracy and f1-score in %, prgm exe time:: program train and testing time			

In practice, the epoch is a hyperparameter set before the training begins. The choice of the epoch size depends on factors such as the model's complexity, the data size, and the model's convergence behavior during training. Selection of a small epoch may result in model underfitting, where the machine learning model hasn't learned the underlying patterns in the data. However, a large size epoch may lead the model to overfit, where the model starts memorizing the training data instead of generalizing well to unseen data. The epoch selection can be any integer value that lies between 1 to infinity. By tradition, the ML/ DL researcher selects large values of epochs.

This experiment aims to identify the optimal number of epochs that yield the highest accuracy for the Bi-LSTM model. Similar to the previous experiment, the Bi-LSTM hyperparameters were randomly selected. Longer epochs result in extended training times for the model. The random numbers of epoch values were chosen between 5 to 205, and the accuracy and f1-score were found to be highest at 205 epochs. However, it's important to note that a larger epoch value increases the training time for our model. In this experiment, a batch of 205 sizes enhances the accuracy of the Bi-LSTM network traffic anomaly detection model, achieving a detection rate of network anomalies at 98.5%.

5.5. Experiment: Model layers parameters Vs. accuracy

In our prior experiments, 5.1 to 5.4, we investigated the impact of various hyperparameters, including the optimizer, number of epochs, batch size, and the train test data split ratio. Our results

reveal that the combination of the Nadam optimizer, 205 epochs, a batch size of 50, and a train test split ratio of 70%: 30% delivers optimal performance after evaluating the model performance metrics.

Table 5: Bi-LSTM architecture Vs. accuracy

optimizer = Nadam, batch_size = 50, test data= 30%, train data=70%						
Input layer		Hidden layer 1		Hidden layer 2		acc. %
neuron	act. fn	neuron	act. fn	neuron	act. fn	
8	relu	8	relu	8	relu	97.48
4	sigmoid	4	sigmoid	4	sigmoid	97.05
16	relu	16	relu	16	relu	97.93
16	selu	16	selu	16	selu	97.97
64	**sigmoid**	**50**	**sigmoid**	**50**	**sigmoid**	**98.52**
49	sigmoid	128	sigmoid	128	sigmoid	98.18
80	relu	64	relu	64	relu	98.48
4	relu	4	relu	4	relu	97.55
act.fn::activation function, acc:: model accuracy						

This study investigated different configurations of neurons and activation functions for the neural network of the Bi-LSTM model. The dense output layer is structured to provide probabilities for distinguishing between normal and abnormal classes, rendering the softmax activation function the most appropriate selection for the binary class dataset.

This experiment evaluated diverse configurations of Bi-LSTM neurons and activation functions for input and hidden layers. Several results from the conducted experiment are outlined in Table 4. Based on the tabulated results, 64 neurons in the input layer and 50 neurons in each hidden layer of our model produce the ultimate accuracy of 98.52 % in the domain of network anomaly detection.

5.6. Experiment: Sampling Vs. performance metrics for multiclass NSL-KDD dataset

Since these data represent a refined version of the KDD99 dataset, minimal data preprocessing is required. The downloaded train data (KDDTrain+) with the target class was initially separated from the training dataset to establish the class label. Among the remaining numerical features, three categorical attributes, 'protocol_type,' 'service,' and 'flag,' are extracted. Dummy one-hot encoding methods convert categorical into numerical values, while the numerical features are normalized using standard scaling methods. Subsequently, both feature sets are merged into a unified data frame, resulting in the final data set.

The attack types on both KDD99 and NSL-KDD are presented in Table 6. The network attack traffic in these datasets is classified into 'Denial of Service,' 'Probe,' 'Remote to Local,' and 'User to Root' [21]. A denial-of-service attack prevents legitimate users from accessing resources via the network, causing a disruption in the availability of those resources. On the other hand, a probe is a scanning attack aimed at identifying vulnerabilities in a system connected to the network. This probing attack targets weaknesses and facilitates potential compromise of the system.

Table 6: Attack types and traffic information in NSL-KDD

Class	Attack Types	Data
Probe	Satan, MScan, Upsweep, Saint, Nmap, Portsweep	11656
U2R	Ps, Perl, Buffer_overflow, Sqlattack, Rootkit, Loadmodule, Xterm	52
Normal		67343
R2L	Spy, Ftp_write, Guess_Password, Imap, Phf, Multihop, Warezmaster, Xlock, Warezclient, Xsnoop, Snmpguess, Snmpgetattack, Named, Httptunnel, Sendmail	995
DoS	Back, Worm, Apache2 Neptune, Smurf, Pod, Teardrop, Udpstorm, Processtable, Land	45927
Total traffic data		12593

Likewise, the remote-to-local attack involves illegal access to a remote terminal. The user-to-root attack entails gaining privilege as a root user, with the root password obtained through various techniques such as password sniffing, brute-forcing, or social engineering.

Under-sampling is a straightforward approach and a method for addressing the class imbalance in datasets. This technique involves preserving all data within the minority class while reducing the volume of data in the majority class. It represents one of several tools available to data scientists for enhancing the accuracy of insights extracted from initially imbalanced datasets. In under-sampling, data samples from the majority class are randomly chosen and removed until a balanced distribution is achieved. This reduction in data volume can alleviate storage constraints and enhance processing efficiency. However, it's significant to note that this reduction may result in the loss of valuable information.

Conversely, oversampling is employed when the available data is insufficient in quantity. Its objective is to rectify dataset imbalance by augmenting the number of rare samples. Instead of discarding abundant samples, oversampling techniques generate new rare samples through replication, bootstrapping, or SMOTE (Synthetic Minority Over-Sampling Technique). SMOTE, which stands for synthetic minority over-sampling technique, is a specific form of oversampling that involves the synthetic generation of data points for the minority class. In this process, a random selection of k nearest neighbors is chosen to determine the appropriate oversampling level.

After preprocessing, the NSL-KDD KDDTrain+ multiclass data initially exhibits imbalanced class distributions. Various techniques can be employed to rectify this imbalance, including under-sampling, over-sampling, and hybrid sampling. Our experiment utilized an automated sampling approach combining random under-sampling and SMOTE to restructure the data for all classes based on our implemented sampling method. Random oversampling consists of randomly choosing instances from the minority class, replacing them, and incorporating them into the training dataset. On the other hand, random under-sampling entails randomly selecting instances from the majority class and removing them from the dataset.

Table 7: Bi-LSTM with random under-sampling and performance

BI-LSTM Model with Random Under-Sampling and Performance				
Epochs= 50, Batch_size= 512, Data = NSL-KDD Multiclass (5 class) _RUS				
SN	**Class**	**Precision %**	**Recall %**	**F1-Score %**
1	DoS	100	100	100
2	Probe	100	79.17	88.37
3	R2L	88.89	80	84.21
4	U2R	73.68	93.33	82.35

5	Normal	86.67	100	92.86
Average		**91.29**	**89.74**	**89.81**
Accuracy = **89.74** %				
Program exe time = 17.72 sec				

The number of new datasets generated depends on each target class's original data size. Random under-sampling reduced the NSL-KDD data to 52 instances in each of the five classes by randomly eliminating data points. Conversely, SMOTE, an oversampling technique, augmented the dataset by introducing additional data points. During this experiment, substantial data augmentation created well-balanced datasets, with each target class containing 67,343 instances.

The balanced NSL-KDD data has been partitioned into training and testing subsets to facilitate the training and evaluation of the Bidirectional LSTM model. As determined in previous experiments, the train test data split ratio of 70%:30%.

The architecture of the Bi-LSTM neural network mirrors that used in prior experiments, with the input layer containing 64 elements and both hidden layer1 and hidden layer2 comprising 50 elements. A trade-off analysis was conducted to determine the optimal combination of epochs and batch size while considering the Bi-LSTM model's performance.

Table 8: Bi-LSTM with SMOTE technique and performance

BI-LSTM Model with SMOTE and Performance				
Epochs= 50, Batch size= 512, Data = NSL-KDD Multiclass (5 class) RUS				
SN	**Class**	**Precision %**	**Recall %**	**F1-Score %**
1	DoS	99.99	99.98	99.98
2	Probe	99.99	99.98	99.98
3	R2L	99.99	99.18	99.59
4	U2R	99.18	1	99.59
5	Normal	1	99.99	1
Average		**99.83**	**99.83**	**99.83**
Accuracy = **99.83** %				
Program exe time = 770.52 sec				

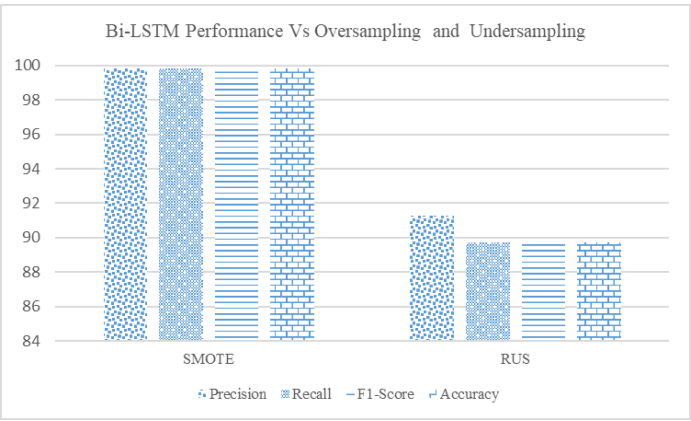


Figure 6: Bi-LSTM performance Vs. oversampling and under-sampling

In our hyperparameter tuning, we aimed to balance program execution time and model performance, as previously demonstrated. As a result, the model was trained for 50 epochs using a batch size of 512 and the Nadam optimizer with a learning rate of 0.041, as detailed in the accompanying table.

The random under-sampling methods produce the NIDS multiclass accuracy of 89.74%, average precision of 91.29 %, recall of 89.74%, and 89.91% f1-score referenced from Table 7. The program execution time is short as compared with oversampling.

Table 8. Shows the performance of the Bi-LSTM with over-sampling methods called SMOTE where the default value of K, i.e., 5, is taken during this experiment. The nearest neighbors value K defines the neighborhood of samples to generate the synthetic samples. We listed the individual class performance as well as average class performance. Figure 6 shows the visualization plot to compare the under-sampling and over-sampling performance on the NSL-KDD multiclass dataset using the Bi-LSTM model. The over-sampling (SMOTE) for the NSL-KDD multiclass dataset provides the 99.83% average precision, recall, and F1 score.

6. Conclusion

The Highest performance is achieved during network traffic anomaly detection using the bidirectional LSTM model. The combination of tunned different hyperparameters (from the above experiments) values, including epoch, optimizer, and batch size, outperformed the anomaly detection model. Determination of hyperparameters' values for the Bi-LSTM anomaly detection model on the NSL-KDD dataset highly contributes to the domain of anomaly detection using machine learning and deep learning. Similarly, we can use no fixed split ratio values for the efficient anomaly detection model. This research work determines the split ratio to produce the highest performance on anomaly detection using the Bi-LSTM model on the NSL-KDD dataset. The combination of neural network architecture memory elements plays an important role in training and testing the model during network anomaly detection. Data imbalance is another main problem to deal with during network anomaly detection. The sampling techniques either delete the data entry randomly or generate the data entry randomly. The sampling technique balances the data in the multiclass dataset. During this research work, the implementation of the random up-sampling methods outperformed the model and produced the highest performance.

We compare our results with existing research [17] to prove that our model is outperformed on the KDD-NSL multiclass dataset. The previously completed research compared their model performance in paper at 99.70% with the other previously researched model's performance, such as Artificial Neural Network (ANN) model at 95%, Decision Tree and Random Forest with 92.60%, Linear Regression, and Random Forest with 94%, Random Forest, and Bayesian Network with 93.4 %, Deep Neural Network with 97% [17]. Our proposed model pipeline for the Bi-LSTM-based network anomaly detection model delivers a higher accuracy of 99.83% is greater than the obtained model performance in research work [17]. The values of bidirectional LSTM model hyperparameters, including epochs values, optimizer, batch size, train test slit ratio, and SMOTE sampling technique for the multilayer bidirectional LSTM neuron architecture (layers, activation function, and memory units) are examined to achieve the highest anomaly detection model performance. The results from these experiments consistently demonstrate that the bidirectional LSTM model, configured with the explored parameters, significantly enhances detection accuracy and f1-score. This model can be experimented with using different network intrusion datasets. Creating a new network intrusion

dataset with the latest network attacks will be the extension of this task in the future.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The National Science Foundation (NSF) and Scholarship for Service CyberCorps (SFS CyberCorps) programs support this research study. The award information of NSF and CyberCorps are #1910868 and #2219611, respectively. This cannot be completed without the continuous support of advisors and the Electrical and Computer Engineering Departments of Prairie View A&M University

References

[1] T. Acharya, A. Annamalai, M.F. Chouikha, "Efficacy of Bidirectional LSTM Model for Network-Based Anomaly Detection," in 13th IEEE Symposium on Computer Applications and Industrial Electronics, ISCAIE 2023, Institute of Electrical and Electronics Engineers Inc.: 336–341, 2023, doi:10.1109/ISCAIE57739.2023.10165336.

[2] N. Moustafa, J. Hu, J. Slay, "A holistic review of Network Anomaly Detection Systems: A comprehensive survey," Journal of Network and Computer Applications, **128**, 33–55, 2019, doi:10.1016/j.jnca.2018.12.006.

[3] S. Samonas, D. Coss, THE CIA STRIKES BACK: REDEFINING CONFIDENTIALITY, INTEGRITY AND AVAILABILITY IN SECURITY.

[4] Y. Fu, Y. Du, Z. Cao, Q. Li, W. Xiang, "A Deep Learning Model for Network Intrusion Detection with Imbalanced Data," Electronics (Switzerland), **11**(6), 2022, doi:10.3390/electronics11060898.

[5] K. Jiang, W. Wang, A. Wang, H. Wu, "Network Intrusion Detection Combined Hybrid Sampling with Deep Hierarchical Network," IEEE Access, **8**, 32464–32476, 2020, doi:10.1109/ACCESS.2020.2973730.

[6] W. Xu, J. Jang-Jaccard, T. Liu, F. Sabrina, J. Kwak, "Improved Bidirectional GAN-Based Approach for Network Intrusion Detection Using One-Class Classifier," Computers, **11**(6), 2022, doi:10.3390/computers11060085.

[7] L. Vu, Q.U. Nguyen, "Handling Imbalanced Data in Intrusion Detection Systems using Generative Adversarial Networks," Journal of Research and Development on Information and Communication Technology, **2020**(1), 1–13, 2020, doi:10.32913/mic-ict-research.v2020.n1.894.

[8] T. Acharya, I. Khatri, A. Annamalai, M.F. Chouikha, "Efficacy of Heterogeneous Ensemble Assisted Machine Learning Model for Binary and Multi-Class Network Intrusion Detection," in 2021 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc.: 408–413, 2021, doi:10.1109/I2CACIS52118.2021.9495864.

[9] T. Acharya, I. Khatri, A. Annamalai, M.F. Chouikha, "Efficacy of Machine Learning-Based Classifiers for Binary and Multi-Class Network Intrusion Detection," in 2021 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc.: 402–407, 2021, doi:10.1109/I2CACIS52118.2021.9495877.

[10] C. Yin, Y. Zhu, J. Fei, X. He, "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," IEEE Access, **5**, 21954–21961, 2017, doi:10.1109/ACCESS.2017.2762418.

[11] Z. Chen, C.K. Yeo, B.S. Lee, C.T. Lau, "Autoencoder-based network anomaly detection," in Wireless Telecommunications Symposium, IEEE Computer Society: 1–5, 2018, doi:10.1109/WTS.2018.8363930.

[12] M. Ganesh, A. Kumar, V. Pattabiraman, "Autoencoder based network anomaly detection," in Proceedings of 2020 IEEE International Conference on Technology, Engineering, Management for Societal Impact Using Marketing, Entrepreneurship and Talent, TEMSMET 2020, Institute of Electrical and Electronics Engineers Inc., 2020, doi:10.1109/TEMSMET51618.2020.9557464.

[13] W. Xu, J. Jang-Jaccard, A. Singh, Y. Wei, F. Sabrina, "Improving Performance of Autoencoder-Based Network Anomaly Detection on NSL-KDD Dataset," IEEE Access, **9**, 140136–140146, 2021, doi:10.1109/ACCESS.2021.3116612.

[14] J. Gao, "Network Intrusion Detection Method Combining CNN and BiLSTM in Cloud Computing Environment," Computational Intelligence and Neuroscience, **2022**, 2022, doi:10.1155/2022/7272479.

[15] T. Acharya, A. Annamalai, M.F. Chouikha, "Efficacy of CNN-Bidirectional LSTM Hybrid Model for Network-Based Anomaly Detection," in 13th IEEE Symposium on Computer Applications and Industrial Electronics, ISCAIE 2023, Institute of Electrical and Electronics Engineers Inc.: 348–353, 2023, doi:10.1109/ISCAIE57739.2023.10165088.

[16] A.G. Salman, Y. Heryadi, E. Abdurahman, W. Suparta, "Single Layer & Multi-layer Long Short-Term Memory (LSTM) Model with Intermediate Variables for Weather Forecasting," in Procedia Computer Science, Elsevier B.V.: 89–98, 2018, doi:10.1016/j.procs.2018.08.153.

[17] P. TS, P. Shrinivasacharya, "Evaluating neural networks using Bi-Directional LSTM for network IDS (intrusion detection systems) in cyber security," Global Transitions Proceedings, **2**(2), 448–454, 2021, doi:10.1016/j.gltp.2021.08.017.

[18] Y. Imrana, Y. Xiang, L. Ali, Z. Abdul-Rauf, "A bidirectional LSTM deep learning approach for intrusion detection," Expert Systems with Applications, **185**, 2021, doi:10.1016/j.eswa.2021.115524.

[19] I. Kandel, M. Castelli, "The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset," ICT Express, **6**(4), 312–315, 2020, doi:10.1016/j.icte.2020.04.010.

[20] M. Tavallaee, E. Bagheri, W. Lu, A.A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009, 2009, doi:10.1109/CISDA.2009.5356528.

[21] L. Dhanabal, S.P. Shantharajah, "A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms," International Journal of Advanced Research in Computer and Communication Engineering, **4**, 2015, doi:10.17148/IJARCCE.2015.4696.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 9, No. 1, 165-174 (2024)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Mathematical Model of Wind Turbine Simulator Based Five-Phase Permanent Magnet Synchronous Generator with Nonlinear Loads and Harmonic Analysis

Peerawat Meesuk, Vijit Kinnares*

School of Engineering, King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, 10520, Thailand

ARTICLE INFO

Article history:
Received: 16 November, 2023
Revised: 28 January, 2024
Accepted: 29 January, 2024
Online: 21 February, 2024

Keywords:
Wind Turbine Simulator
Permanent Magnet Synchronous
Harmonic Effect

ABSTRACT

This paper presents mathematical model of a wind turbine simulator based five-phase permanent magnet generator supplying nonlinear load. The mathematical model of wind turbine characteristics together with available tool blocks of the five-phase permanent generator and semiconductor devices of an AC-DC converter formed as a nonlinear load is implemented on MATLAB /Simulink to investigate the harmonic effect on performance of the generator. The detailed descriptions of the proposed model are fully given. The harmonic analysis is also provided. The validity of the proposed model is verified by simulation using MATLAB /Simulink in terms of dynamic responses of rotor speed, torque and power quality of the generator. It is found that the nonlinear load significantly affects the electromagnetic torque ripple and the distortions of both voltage and current of the generator. Moreover, the proposed system offers higher nonlinear load voltage and faster response compared to a conventional three-phase permanent magnet synchronous generator system. The electromagnetic torque ripple is reduced by 88% and the total harmonic distortions of the phase voltage and the stator current are more or less 7 % and 60 % which exceed the limits of the harmonic standards.

1. Introduction

Wind energy is one of the abundant renewable resources that is getting much attention around the world since it is clean, naturally occurring and environmentally friendly energy without burning fossil. In a wind turbine power generation system, the kinetic energy generated by the wind power is converted into mechanical torque at the shaft of the wind turbine which is coupled either directly or indirectly via gear box to the generator. Then the generator converts mechanical energy harvested from the wind energy by the turbine into electrical energy through magnetic field. However, the use of a wind turbine power generation system is necessary to study many components such as potential of locally generated wind speeds, electricity standards and regulations, system performance, expense and payback on investments, etc. In order to obtain the best performance of the system in terms of maximum efficiency achievement and cost-effectiveness, several research works and development of the wind turbine power generation systems have been paid attention [1,2]. Although an induction generator (IG) has various advantages over a major counterpart permanent magnet synchronous generator (PMSG), particularly for low maintenance, high ruggedness, and low cost, it gives lower efficiency and draws reactive power resulting in poor power factor for both self-excited and grid connected applications. Performance of a three-phase self-excited induction generator (SEIG) operating as a single-phase generator supplying nonlinear loads for standalone applications was investigated in [1]. However, the electromagnetic torque ripple due to harmonic effect has not been reported yet. Normally the SEIG needs a capacitor bank for reactive compensation for both voltage buildup process and voltage regulation but it significantly affects a variation of the generator frequency which is a drawback of the SEIG. Harmonic analysis and experimental tests of a three-phase SEIG with nonlinear loads were proposed in [3]. The obtained results confirm that the harmonics associated with the nonlinear loads affect the voltage and current waveforms. A PMSG is more attractive for small scale applications due to high efficiency, high power density, better voltage regulation for standalone applications [4]. Control of grid connection of an axial flux permanent magnet generator (AFPMG) (i.e. another type of PMSG) using a single-phase inverter with multifunctionality for real power transfer and reactive and harmonic compensations with various types of nonlinear loads was implemented in both simulation and hardware by [4,5]. However, the wind turbine simulator using motor drives as a prime mover has not been proposed. An example of a wind turbine

*Corresponding Author: Vijit Kinnares, School of Engineering, King Mongkut's Institute of Technology Ladkrabang (KMITL), kkwijit@kmitl.ac.th

simulator emulating actual wind turbine characteristics implemented by low cost hardware can be found in [6] which is useful in a laboratory test for wind energy conversion research. However, mathematical model of the simulator implemented on simulation was not focused. Unlike standalone applications, for grid connected applications, a large variation of the frequency of voltage and current of the generator is not necessary which is the major advantage of a wind turbine power generation system since an ac-dc rectifier converts alternating current into a direct current form. Generally, a five-phase permanent magnet synchronous machine is introduced in motoring operation rather than as a generator for electrical propulsion due to low torque pulsations, good fault tolerance, high power density and low loss [6-8]. Although the smaller ripple of the dc output voltage in the five-phase PMSG system associated with a rectifier than that in the three-phase PMSG system was presented in [9], harmonic effect on electromagnetic torque pulsation and harmonic analysis were not reported. There are a few publications related to performance evaluation of a five-phase PMSG with nonlinear load. Moreover so far the harmonic effect on the performance of the five-phase PMSG associated with wind energy conversion has not been investigated yet With the shortcomings of the previous works mentioned earlier, therefore this work contributes to mathematical model setup applied in simulation using MATLAB/Simulink computer program of the wind simulator based five-phase PMSG with nonlinear loads and harmonic analysis for investigating harmonic effect on the generator performance.

The paper is organized as follows. A wind power generation system is introduced briefly followed by mathematical model of wind turbine characteristics in Section 2. Section 3 describes mathematical model of the five-phase PMSG presented in a rotating d-q-x-y frame modeling followed by an available tool block in the MATLAB/Simulink. Section 4 deals with the nonlinear load model and harmonic analysis. Section 5 describes the proposed Simulink model and simulation results in terms of dynamic response and steady state conditions for various performances of the generator like terminal voltage and current waveforms, torque, speed, and so on together with discussion. Finally, conclusion and suggestion are given.

2. Wind Power Generation System

The electrical power generation system for grid connected applications used in the simulation is referred to Figure 1. It consists of a wind turbine simulator, a five-phase permanent magnet synchronous generator and a nonlinear load using a diode-based rectifier and a smoothing filter capacitor. Generally, an inverter circuit is used to convert DC input into AC output for grid connection in order to transfer the electrical energy produced from the wind energy to the grid. In the simulation, the model of the inverter is not taken into account. The equivalent resistor is used for representation of the transferred power and the converter power loss.

In steady state, the mechanical power generated at the shaft of the wind turbine can be obtained as [6]

$$P_{WT} = 0.5\rho\pi R^2 v_{wind}^3 C_p(\lambda,\beta) \quad (1)$$

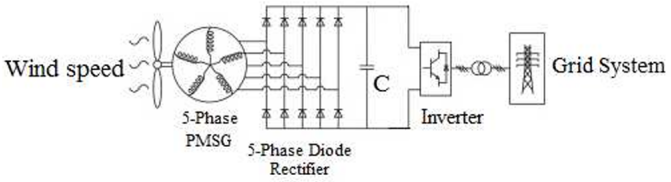


Figure 1: Overview of an electric power generation system using wind energy conversion system for grid connected applications.

where

P_{WT} is the mechanical power of the wind turbine ,

$\rho = 1.25\ kg/m^3$ which is the air density ,

R is the turbine rotor radius,

v_{wind} is the wind velocity,

C_p is the power coefficient,

β is the pitch angle, and

λ is the tip speed ratio.

The power coefficient is given by:

$$C_P(\lambda,\beta) = C_1\left(\frac{C_2}{\lambda_i} - C_3\beta - C_4 e^{\frac{-C_5}{\lambda_i}}\right) + C_6\lambda \quad (2)$$

and

$$\frac{1}{\lambda_i} = \frac{1}{\lambda+0.08\beta} - \frac{0.035}{\beta^3+1} \quad (3)$$

The coefficients C_1 to C_6 are as follows. $C_1\ is\ 0.5176$, $C_2\ is\ 116$, $C_3\ is\ 0.4$, $C_4\ is\ 5$, $C_5\ is\ 21$ and $C_6\ is\ 0.0068$.

The tip speed ratio is given by:

$$\lambda = \frac{\pi n R}{30 v_{wind}} \quad (4)$$

The shaft torque of the wind turbine is expressed as

$$T_{WT} = \frac{30 P_{WT}}{\pi n} \quad (5)$$

where T_{WT} is the mechanical torque of the wind turbine and n is the rotor speed of the wind turbine. Schematic diagram of the proposed system for the simulation on a computer program is shown in Figure 2 in accordance with Figure 1. The power transferred to the grid and the converter power loss are lumped together represented by resistance R_{eq}. Figure 3 shows the simulation model implemented on MATLAB/Simulink simulation program as a wind turbine simulator. The mathematical expressions in (1)-(5) are used to generate wind turbine characteristics providing the mechanical torque applied to the PMSG tool block model. Detailed parameter values can be found in [6]. The wind turbine characteristics are plotted by using the

wind turbine simulator as shown in Figures 4 and 5 for input mechanical torque and mechanical power against the rotor speed, respectively. As can be seen, below maximum values, the mechanical torque is proportional to the squared rotor speed whilst the mechanical power is proportional to the cube of the rotor speed at a given wind speed.

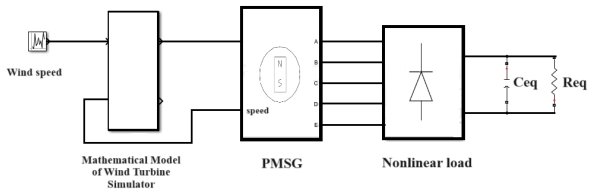


Figure 2: Schematic Diagram of wind turbine simulator based five-phase permanent magnet synchronous generator with nonlinear load.

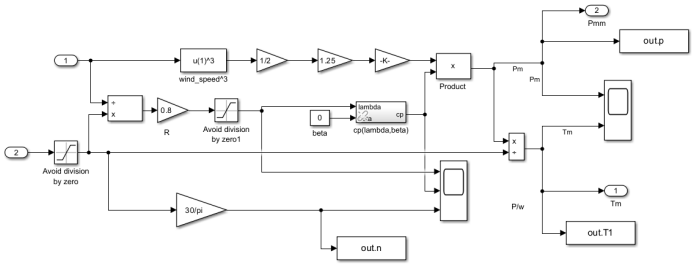


Figure 3: Mathematical model of wind simulator using MATLAB/SIMULINK

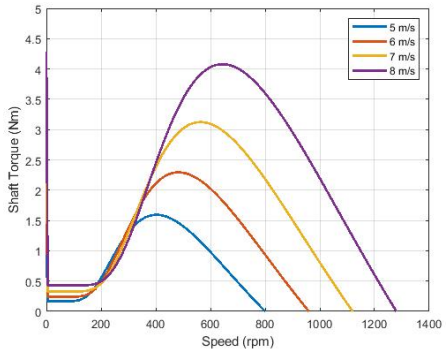


Figure 4: Applied mechanical shaft torque versus rotor speed with variation of wind speed

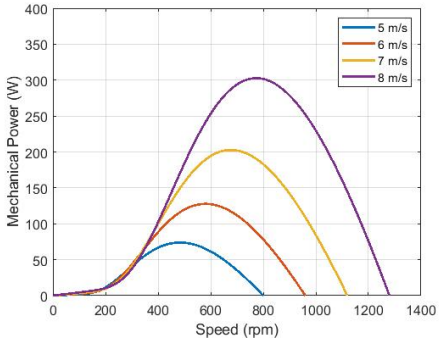


Figure 5: Applied mechanical power against rotor speed with variation of wind speed.

The results from Figures 4 and 5 are in accordance with [6]. Theses confirm the validity of the proposed wind turbine simulator in a mathematical model form.

3. Five-Phase Permanent Magnet Synchronous Generator

The stator winding for each phase is displaced by 72 electrical degrees. The 5-phase PMSG can be modeled in a rotating d-q-x-y frame for analysis of dynamic response and steady state conditions [6]. More details can be also found in [7], [8] and [10]. The mathematical equations can be expressed as follows.

$$v_d = R_s i_d + L_d \frac{di_d}{dt} - \omega L_q i_q \quad (6)$$

$$v_q = R_s i_q + L_q \frac{di_q}{dt} + \omega L_d i_d + \omega \phi_f \quad (7)$$

$$v_x = R_s i_x + L_{ls} \frac{di_x}{dt} \quad (8)$$

$$v_y = R_s i_y + L_{ls} \frac{di_y}{dt} \quad (9)$$

$$v_o = R_s i_o + L_{ls} \frac{di_o}{dt} \quad (10)$$

The relationship between the mechanical torque (T_l) and the electromechanical torque (T_{em}) can be given as

$$J \frac{d\Omega}{dt} + f_r \Omega = T_{em} - T_l \quad (11)$$

where v_d, v_q, v_x, v_y, i_d , i_q, i_x, i_y are the stator voltages and stator currents in the d-q-x-y, respectively. J is the moment of inertia, f_r is the viscous coefficient, and Ω is the rotor speed and T_l is the load torque.

The electromagnetic torque can be expressed as

$$T_{em} = \frac{5}{2} p\left(\phi_f i_q - (L_d - L_q) i_d i_q\right) \quad (12)$$

where p is the pole pair, L_d is the direct axis inductance, L_q is the quadrature axis inductance, i_q is the quadrature axis current, i_d is the direct axis current, L_{ls} is the leakage inductance and ϕ_f is the flux generated by permanent magnet in the rotor. These equations are used for calculation in the machine model. The available tool block for the PMSG model in MATLAB/SIMULINK is illustrated in Figure 6. T_m is the input mechanical torque which the positive sign is for the motoring operation and negative sign represents the generator operation. "m" is the output for measuring various variables and machine performance like stator currents, flux linkages, and mechanical outputs such as speed, electromagnetic torque, etc. The rated values of the used model of the 5-phase PMSG are output torque of 8 Nm and the rotor speed of 2000 RPM.

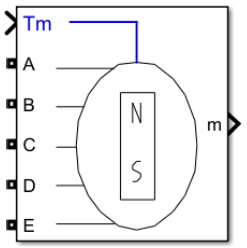


Figure 6: Tool block model of a five-phase permanent magnet synchronous generator available in MATLAB/ Simulink.

4. Nonlinear Load model and Harmonic Current Analysis

In this work, the nonlinear load is connected in parallel with the generator output consisting of a five-phase full bridge rectifier

including 10 diodes and a smoothing filter capacitor connected in parallel at the output of the rectifier. The per phase simplified model of the PMSG with nonlinear load is shown in Figure 7 which various harmonic currents produced by the nonlinear load are represented by harmonic current sources. The time domain of harmonic currents can be expressed as

$$i_s(t) = \sum_{n=1}^{\infty} I_n \cos(n\omega t - \theta_n) \quad (13)$$

where I_n is the amplitude of the source current at harmonic order n and θ_n is the phase difference angle of the current with respect to the source voltage. The nonlinear load model in [11] with Matlab/Simulink is shown in Figure 8 where R_{eq} represents the converter loss and the active power drawn by the grid system in accordance with Figure 1. The values of the smoothing filter capacitor and R_{eq} are 5000 μF and 10 Ω respectively. The diode model consists of a RC snubber circuit including R_s of 250 Ω, connected in series with C of 250 nF , forward voltage drops of 0.8 V and conduction resistance (R_{on}) of 0.001 Ω. In order to analyze the propagation of a harmonic source in the PMSG, the simplified equivalent circuit can be depicted as shown in Figure 9 which is treated in the same manner as the circuit reported in [12] for a three-phase SEIG. The transfer function of the harmonic current propagation can be expressed as

$$T(s) = \frac{I_{mh}}{I_h} \quad (14)$$

By using KCL (Kirchoff's current law), the harmonic current at order h injected into the machine yields

$$I_{mh} = \frac{R_L}{R_L + R_s + sL_s} I_h \quad (15)$$

In this study, the parameters of the PMSG are as follows. R_s is 0.0485 Ω , L_s is 8.5 mH and R_L is 10Ω . Therefore, the transfer function of the harmonic propagation is

$$T(s) = \frac{I_{mh}}{I_h} = \frac{R_L}{R_L + R_s + sL_s} = \frac{10}{0.0085s + 10.0485} \quad (16)$$

From (16) the frequency response can be illustrated in magnitude and phase forms by using the Bode Plot as shown in Figure 10. According to these results the cut-off frequency is approximately 100 kHz. It implies that the harmonic frequency below this value can affect the PMSG. Note that the harmonic currents can give the effect of a decrease in efficiency, heating and reducing the lifetime of the machine [12].

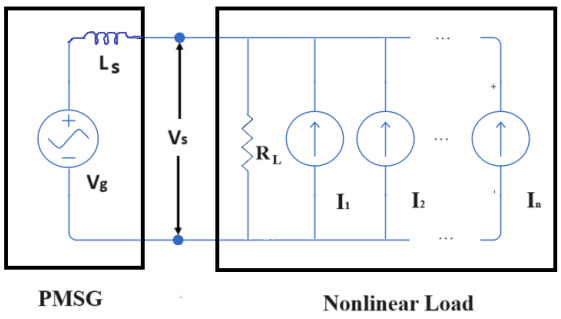


Figure 7: Simplified model of PMSG considering nonlinear load.

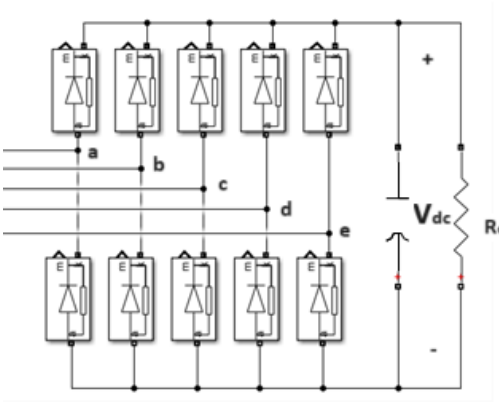


Figure 8: Model of the nonlinear load using MATLAB/ Simulink.

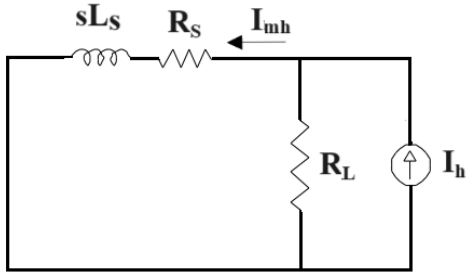


Figure 9: Simplified equivalent circuit for considering harmonic current propagation on PMSG.

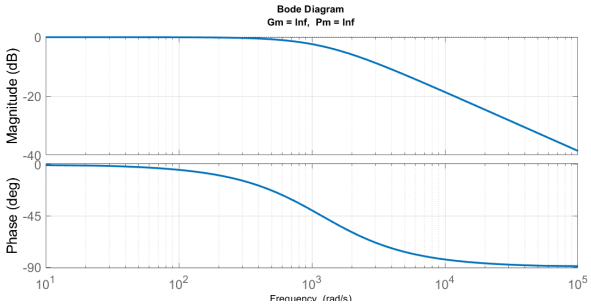


Figure 10: Frequency response for harmonic current propagation.

5. Simulation Results and Discussion

The simulation model of the wind turbine simulator based five- phase PMSG supplying nonlinear load is implemented by using MATLAB/SIMULINK as shown in Figure 11. It consists of a five-phase PMSG tool block and subsystems of the wind turbine simulator using the proposed mathematical model, and the nonlinear load. Wind speed is required for the command. The wind turbine simulator provides the input mechanical torque for the PMSG. The rotor speed of the PMSG is used to feedback to the wind simulator. Various waveforms and signals are measured by measurement units.

Figure 12 illustrates comparative waveforms between the conventional three-phase PMSG and the proposed five-phase PMSG of the machine terminal voltage and nonlinear load voltage (i.e. dc link voltage or output of the rectifier, Vdc) at the wind speed of 6 m/s during startup and steady state conditions.

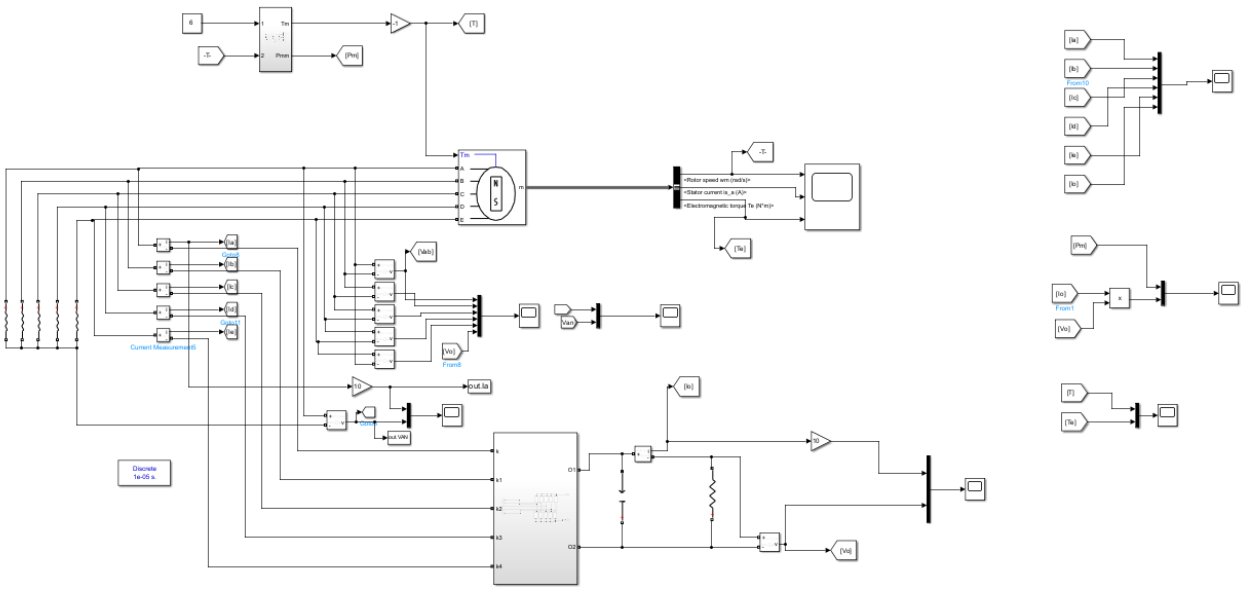
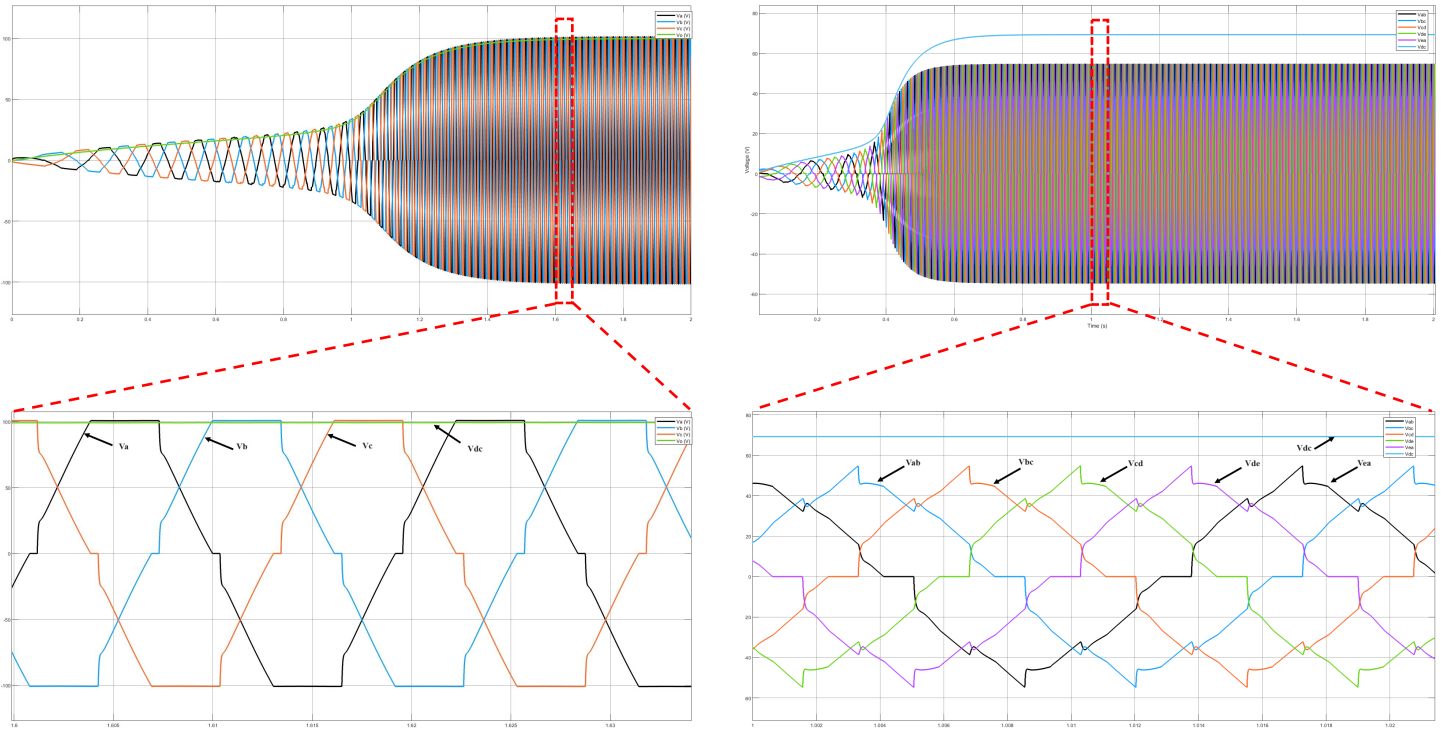


Figure 11: Simulation model of wind turbine simulator based five-phase PMSG.



a.) b.)

Figure 12: Waveforms of output voltage of rectifier and terminal voltage during startup and steady state conditions and corresponding enlarged waveforms
a.) conventional three-phase PMSG b.) proposed five-phase PMSG

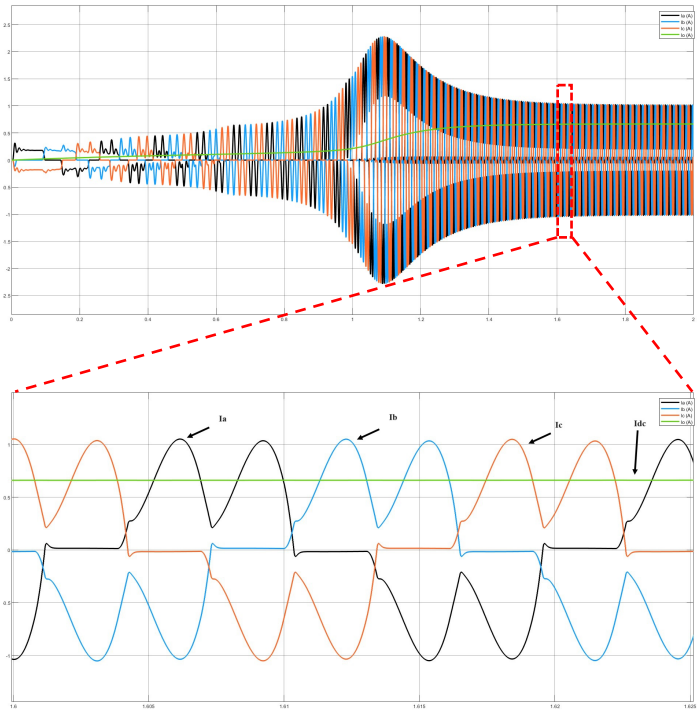
Clearly, with the same input mechanical torque for both cases the voltage and its frequency are gradually increased and then abruptly grown up until steady state. The dc output voltage of the rectifier and dynamic response for the proposed PMSG are higher and faster response than those for the conventional three-phase PMSG, respectively. The enlarged waveforms show the distortion due to the propagation of the harmonic currents associated with the nonlinear load. The dc voltage of the rectifier output is much higher than the peak value of the terminal voltage which is the advantage of the proposed five-phase PMSG compared to the conventional three-phase PMSG. Moreover, the dc output contains insignificant ripple. Apparently, enlarged waveforms of the

terminal voltages are distorted from sinusoid due to the propagation of the harmonic currents associated with the nonlinear load for both cases. The five-phase PMSG with nonlinear load offers less distortion of voltage waveforms than those for the three-phase PMSG. Figure 13 illustrates comparative waveforms between the conventional three-phase PMSG and the proposed five-phase PMSG of the corresponding stator currents and dc output current waveform during start up and steady state conditions. The dc output current of the rectifier and dynamic response for the proposed PMSG are higher and faster response than those for the conventional three-phase PMSG, respectively. Obviously, the inrush currents are present during a suddenly increase in terminal voltage for both cases. Apparently, the enlarged current waveforms are distorted due to the nonlinear load.

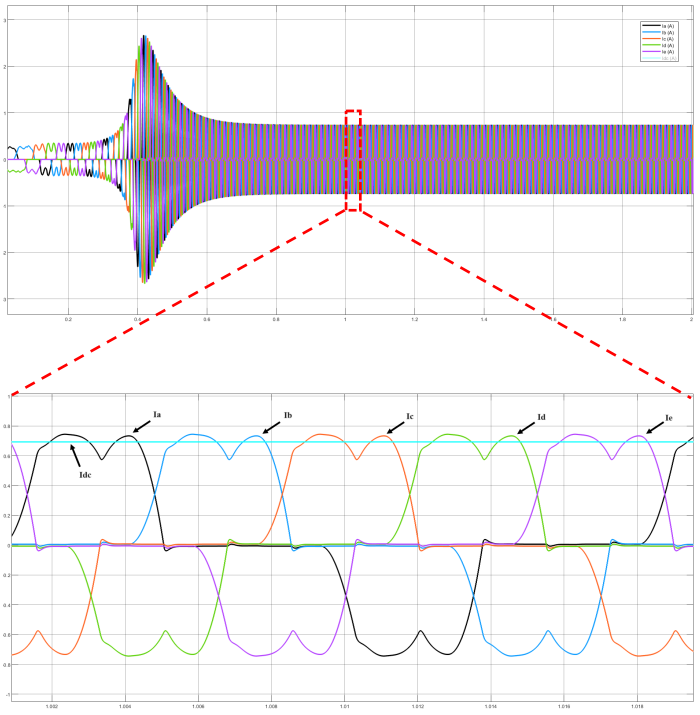
Figures 14 and 15 show dynamic response during startup of the input mechanical torque and the electromechanical torque of the PMSG for linear and nonlinear loads with the same condition for a conventional three-phase PMSG and the proposed five-phase PMSG, respectively. Noticeably, the electromagnetic torque contains higher ripple and slower response for the nonlinear load compared to the linear load due to the large value of the smoothing filter capacitor and harmonic effect. When comparing the ripple of the electromagnetic torque and dynamic response for the nonlinear load between the conventional three-phase PMSG and the proposed five-phase PMSG, obviously, the proposed five-phase PMSG has lower ripple and faster response than the three-phase PMSG. During the growth of the terminal voltage, the transient torque occurs for both linear and nonlinear loads. The peak values of the mechanic torque for both cases are almost equal. The significant difference between the input mechanical torque and the electromagnetic torque during transient response represents the large acceleration torque required for the linear load resulting in faster response (i.e. see equation (11). At the steady state, the absolute input mechanical torque is little higher than the absolute electromagnetic torque due to friction and windage components.

Figures 16 and 17 illustrate the enlarged waveforms during the steady state of the input mechanical torque and the electromagnetic torque for linear and nonlinear loads. Clearly, the electromagnetic torque has higher ripple for the nonlinear load than that for the linear load. It is found that the electromagnetic torque ripple is about 9.34% which is in good agreement with [9]. Note that the electromagnetic torque ripple for the five-phase PMSG is lower than that for the conventional three-phase PMSG reported in [9]. The peak-to-peak electromagnetic torque ripple of the proposed five-phase PMSG is reduced about 88.18%. compared to that of the conventional three-phase PMSG. In order to confirm the higher ripple of the electromagnetic torque for the conventional three-phase PMSG with nonlinear load, Figure 18 illustrates comparative harmonic spectra of the electromagnetic torque between the three-phase PMSG and the proposed PMSG by using FFT analysis tool block. Obviously, the electromagnetic torque of the conventional three-phase PMSG contains large low order harmonic components (i.e. the sixth hramonics) resulting in high pulsation.

Figure 19 shows the input mechanical power and the electrical power of the nonlinear load. During transient response in startup duration the input power is drastically higher than the output power since the input power is drawn with large amount for acceleration of the generator. Eventually, the input power is higher than the output power with a fixed value. The difference of both powers is present due to the machine losses including core loss, copper loss, and friction and windage loss, semiconductor losses including conduction and switching losses, and snubber loss in the model. As a consequence, it implies that the efficiency of the system can be determined which is not included in here. Comparison between linear and nonlinear loads for phase voltage and current waveforms are shown in Figure 20.



a.)



b.)

Figure 13: Waveforms of stator currents during start up and steady conditions and enlarged waveforms
a.) conventional three-phase PMSG b.) proposed five-phase PMSG

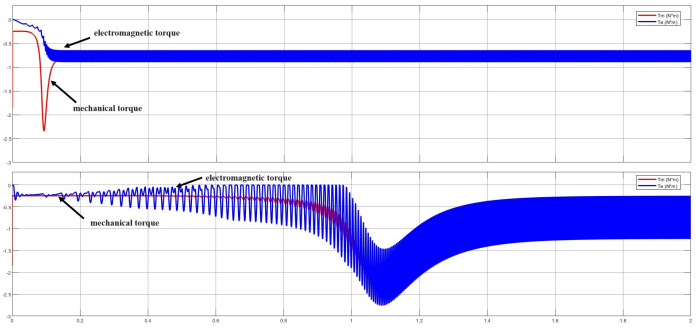


Figure 14: Input mechanical torque and electromagnetic torque of conventional three-phase PMSG a.) linear load b.) nonlinear load

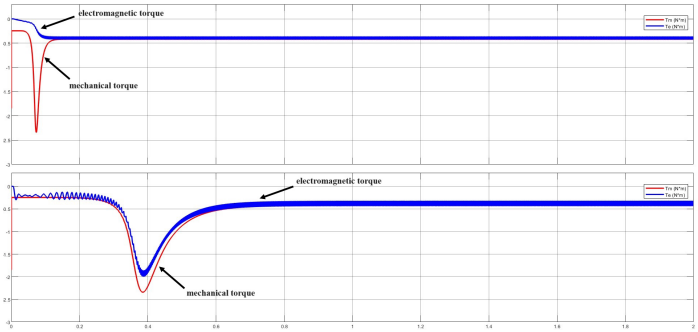


Figure 15: Input mechanical torque and electromagnetic torque of the proposed five-phase PMSG a.) linear load b.) nonlinear load

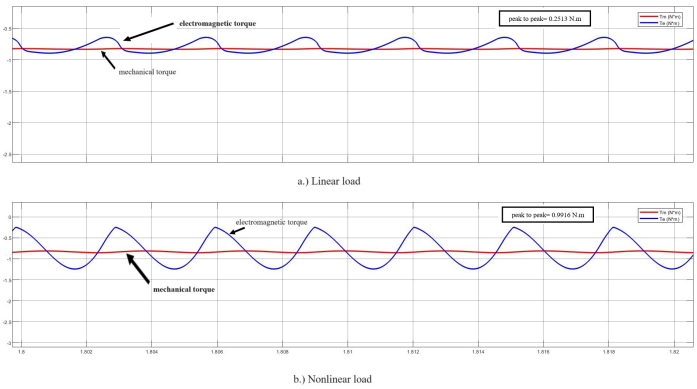


Figure 16: Instantaneous input mechanical torque (red signal) and instantaneous electromagnetic torque (blue signal) under steady state of conventional three-phase PMSG.

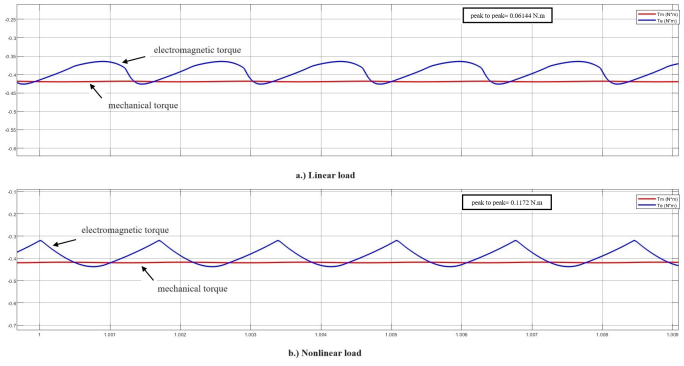


Figure 17: Instantaneous input mechanical torque (red signal) and instantaneous electromagnetic torque (blue signal) under steady state of the proposed five-phase PMSG.

Apparently, unlike for linear load without distortion, both terminal voltage and stator current waveforms of the generator for the nonlinear load are significantly distorted from sinusoidal waveforms. This is the fact that the harmonic currents associated with the nonlinear load are injected to the machine resulting in the distortion of sinusoidal waveforms.

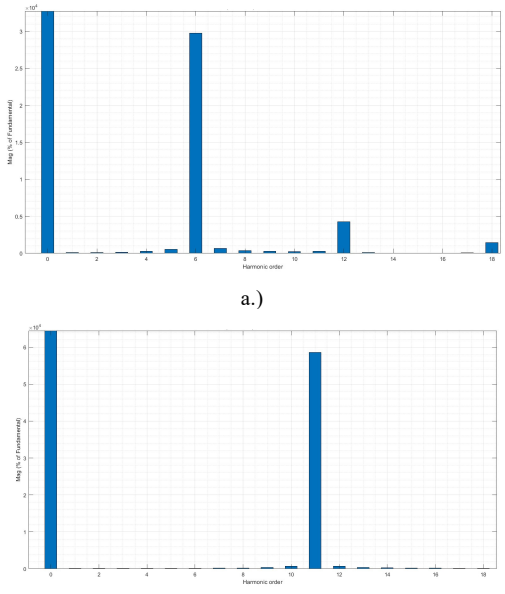


Figure 18: Electromagnetic torque spectra for nonlinear load a.) conventional three-phase PMSG b.) proposed five-phase PMSG

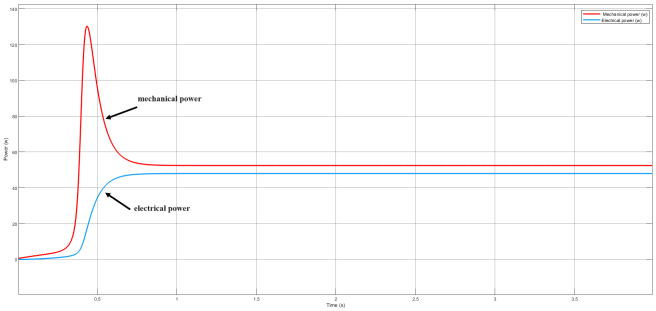


Figure 19: Transient response and steady state conditions during startup of input mechanical power and the output electrical power of the nonlinear load.

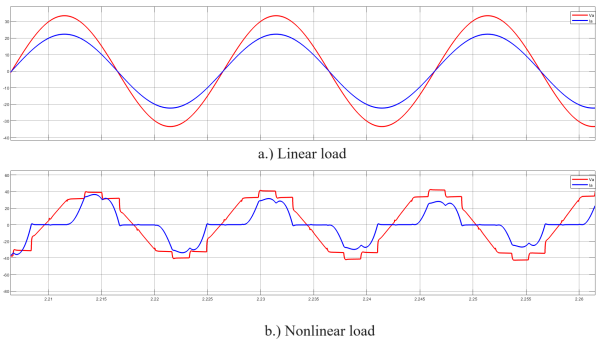
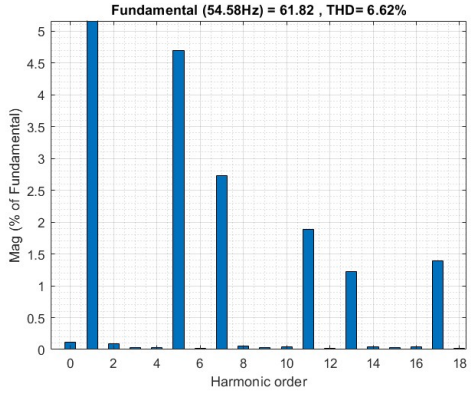
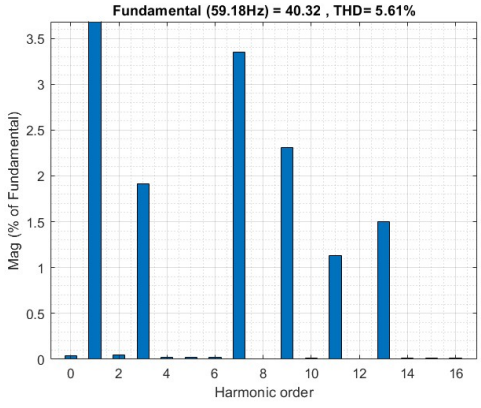


Figure 20: Terminal voltage and stator current waveforms for nonlinear load and linear load.

Figures 21 and 22 illustrate comparative harmonic spectra of both terminal voltages and stator currents between the three-phase PMSG and the proposed PMSG for the nonlinear load by using FFT analysis tool block. For the proposed PMSG case, the total harmonic distortion of the phase voltage and stator current are more or less 7 % and 60 % , respectively. These high figures are not within acceptable levels, particularly for the current case [13]. The solution for this problem could be introduced by either passive or active power filters. Note that these values depend on load level conditions. Quite clearly, unlike the conventional PMSG, the magnitudes of the third harmonics for both current and voltage are dominant. Unlike the conventional three-phase PMSG, some multiples of three of harmonics are still present which are different from a three-phase system in which the multiples of three of harmonic orders are absent. As a consequence, the harmonic distortion for the conventional three-phase PMSG is better than that for the proposed PMSG. The harmonic orders are in agreement with those reported in [14-17].



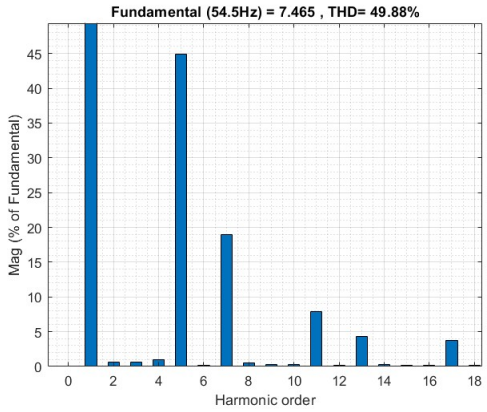
a.)



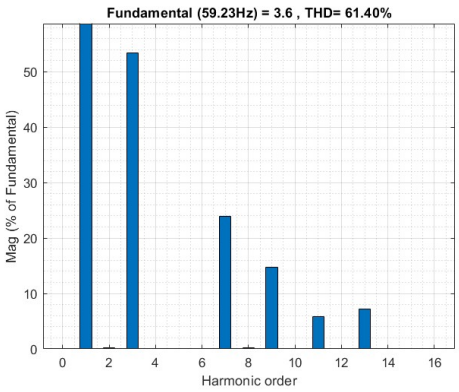
b.)

Figure 21: Terminal voltage harmonic spectra for nonlinear load
a.) conventional three-phase PMSG b.) proposed five-phase PMSG

Figure 23 show a comparison of the ripple of the dc output voltage for the nonlinear load between the conventional three-phase PMSG and the proposed five-phase PMSG. Obviously, unlike the conventional three-phase PMSG, the ripple exists twice frequency of the ac terminal voltage of the machine. As a result, the proposed five-phase PMSG has lower ripple of dc output voltage lower than three-phase PMSG. It implies that the proposed five-phase PMSG requires a smaller size of the smoothing capacitor.

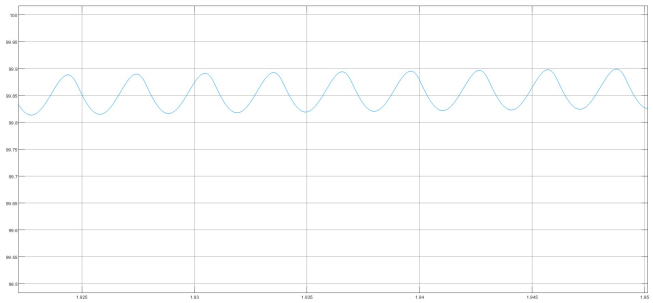


a.)

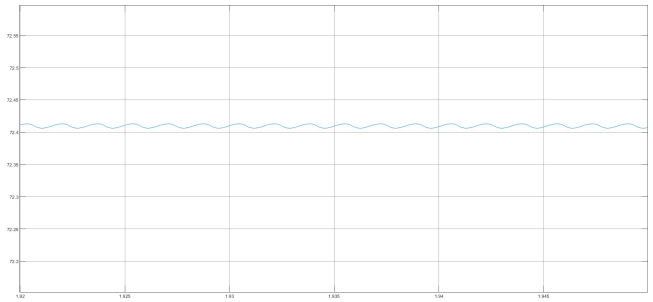


b.)

Figure 22: Stator current harmonic spectra for nonlinear load
a.) conventional three-phase PMSG b.) proposed five-phase PMSG



a.)



b.)

Figure 23: Ripple voltage of the DC output voltage with nonlinear load
a.) conventional three-phase PMSG b.) proposed five-phase PMSG

Figure 24 displays transient responses of the rotor speed, the stator phase current and the electromagnetic torque during a step change in wind speed from 5 to 8 m/s at t=2 seconds. At the beginning of the startup operation, the rotor speed, the electromagnetic torque (i.e. negative sign), and the stator current and its frequency are gradually increased. Then at about t=0.5 seconds, they are rapidly grown up until they reach to the steady state. When the wind speed is changed in a step, the high peak electromagnetic torque and high inrush stator current occur. Obviously after transient response the electromagnetic torque and stator current fall down to steady state values which are higher than the previous ones before disturbance. The rotor speed is increased slowly which looks like the response of the first order of a derivative equation in accordance with (11).

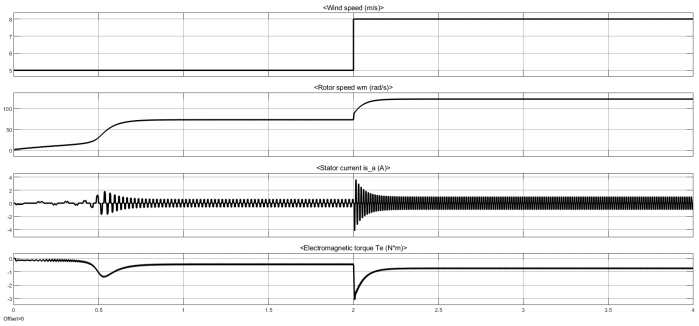


Figure 24: Step change in wind speed from 5 m/s to 8 m/s

In order to investigate the performance of the proposed system during wind speed changes according to the realistic conditions, wind speeds with random input ranging between 5 to 8 m/s, have been applied. The dynamic responses of the rotor speed, the electromagnetic torque and stator current with such random wind speeds are illustrated in Figure 25. The rotor speed response is slower than the wind speed command. The final value of the rotor speed after disturbance is promotional to the wind speed. During a step change in either an increase or a decrease in the wind speed, the high inrush current and high peak torque occur particularly during startup.

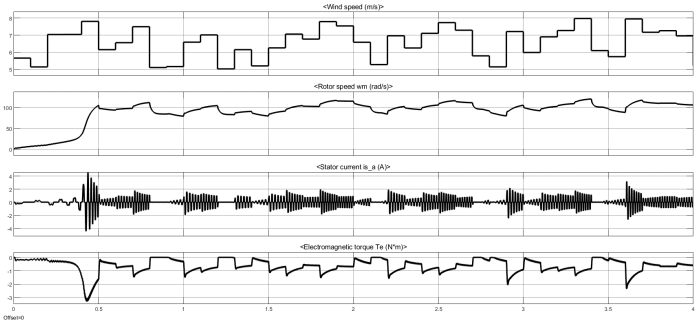


Figure 25: Dynamic response for random input wind speed for the nonlinear load.

6. Conclusion

This paper has proposed mathematical model of a wind turbine simulator based five-phase permanent synchronous generator supplying nonlinear load using MATLAB/Simulink. The related mathematical equations of the machine model are also described. The harmonic analysis for propagation and the harmonic effect on the PMSG have been given. The descriptions of Simulink model of the proposed system are fully given. The investigation of the performance and power quality of the PMSG with the nonlinear load under various conditions has been conducted. It is found that the harmonics associated with the nonlinear load adversely affects the performance and power quality of the PMSG. The nonlinear load causes degradation of the performances of the proposed PMSG in terms of higher electromagnetic torque ripple and slower response when compared to the linear load case. The proposed five-phase PMSG offers some advantages over the conventional three-phase PMSG in terms of lower electromagnetic torque, higher output dc voltage value and lower ripple of the dc output voltage. However, its disadvantage is higher harmonic distortions of both terminal voltage and stator current which are over limits of the standards. The suggestion is a requirement of harmonic solutions such as either passive or active power filters to mitigation of such harmonics.

References

[1] P. Nakorn, P. Machot, V. Kinnares, and C. Manop, “Study of Three phase Self-excited Induction Generator Operating as Single-phase Induction Generator Supplying Non-linear Load,” In 2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTICON), 806-809, 2021, DOI: 10.1109/ECTI-CON51831.2021.9454845.

[2] C. Xue, W. Song, and X. Feng, “Finite control-set model predictive current control of five - phase permanent - magnet synchronous machine based on virtual voltage vectors,” IET Electric Power Applications, **11**(5), 836-846,2017, doi.org/10.1049/iet-epa.2016.0529.

[3] T. Kamel, D. Abdelkader, and B. Said, “Vector control of five-phase permanent magnet synchronous motor drive,” in 2015 4th International Conference on Electrical Engineering (ICEE) , 1-4 , 2015, DOI: 10.1109/INTEE.2015.7416853 .

[4] P. Wannakarn, and V. Kinnares, "Single-phase grid connected axial flux permanent magnet generator system with reactive power compensation functionality," in 2012 10th International Power & Energy Conference (IPEC) ,338-341,2012, DOI: 10.1109/ASSCC.2012.6523289.

[5] P. Wannakarn, and V. Kinnares, “Single-phase grid connected axial flux permanent magnet generator system with harmonic mitigation functionality for various types of nonlinear loads,” International Review of Electrical Engineering, **13**(2), 157-164,2018, DOI: 10.1109/ASSCC.2012.6523289.

[6] N. Thodsaporn, and V. Kinnares, “Wind turbine simulator equipped with real-time monitoring and user-friendly parameter setup controlled by C2000 microcontroller,” In 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 713-716, 2020, DOI: 10.1109/ECTI-CON49241.2020.9158236.

[7] H. C. Chen, C. H. Hsu, and D. K. Chang, “Position sensorless control for five-phase permanent-magnet synchronous motors,” in 2014 IEEE/ASME International Conference on Advanced Intelligent Mechatronics 794-799, 2014, DOI: 10.1109/AIM.2014.6878176.

[8] F. Yu, X. Zhang, M. Qiao, and C. Du, “The direct torque control of multiphase permanent magnet synchronous motor based on low harmonic space vector PWM,” in 2008 IEEE International Conference on Industrial Technology, 1-5, 2008, DOI: 10.1109/ICIT.2008.4608494.

[9] N. E. A. M. Hassanain, Comparison between three-and five-phase permanent magnet generators connected to a diode bridge rectifier, Ph. D Thesis, University of Strathclyde Department of Electronics and Electrical Engineering, 2009.

[10] S. Rhaili, A. Abbou, S. Marhraoui, N. El Hichami, , and A. V. Hemeyine, “Robustness investigation of Vector Control of Five-phase PMSG based Variable-Speed Wind Turbine under faulty condition.” In 2018 Renewable Energies, Power Systems & Green Inclusive Economy (REPS-GIE) , 1-6, 2018, DOI: 10.1109/REPSGIE.2018.8488809.

[11] P. Meesuk, and V. Kinnares, "Mathematical Model of Wind Turbine Simulator Based Five Phase Permanent Magnet Synchronous Generator Supplying Non-Linear Loads," in 2023 9th International Conference on Engineering, Applied Sciences, and Technology (ICEAST), 168-171, 2023,

DOI: 10.1109/ICEAST58324.2023.10157898.

[12] R.Nazir, "Analysis of Harmonic Currents Propagation on the SelfExcited Induction Generator with Nonlinear Loads," Journal of Electrical Engineering and Technology, **9**(6), 1935-1943, 2014, http://dx.doi.org/10.5370/JEET.2014.9.6.1935.

[13] K. Wirtayasa, C. Y. Hsiao, N. C. Yang, "High-Performance Five-Phase Axial Flux Permanent Magnet Generator for Small-Scale Vertical Axis Wind Turbine," Applied sciences, 2020, **12**(7), 3632, doi.org/10.3390/app12073632.

[14] M. McCarty, T. Taufik, A. Pratama, and M. Anwari, "Harmonic analysis of input current of single-phase controlled bridge rectifier," in 2009 IEEE Symposium on Industrial Electronics & Applications, **1**, 520-524, 2009, DOI: 10.1109/ISIEA.2009.5356404.

[15] R. R. Kumar, S. K. Singh**,** R. K. Srivastava, A. S. S. Vardhan, R. M. Elavarasan, R. K. Saket, E. Hossain, "Modeling of airgap fluxes and performance analysis of five phase permanent magnet synchronous generator for wind power application," IEEE Access, **8**, 195472-195486, 2014.

[16] F. Bu, H. Liu, W. Huang, H. Xu, K. Shi, "Optimal third harmonic injection-based control for a five phase dual stator winding induction generator DC generating system," IEEE Transactions on Industrial Electronics, 2018, 65(11), 9124-9134.

[17] R. R. Kumar, A. Kumari, P. Devi, and C. Chetri, "Design and Performance Characteristics of Dual-Rotor Magnetically Decoupled Stator Five-Phase Permanent Magnet Synchronous Generator for Wind Power Applications.," 2021, In Advances in Electrical and Computer Technologies: Select Proceedings of ICAECT 2020, 1261-1274, 2020, DOI: 10.1109/ACCESS.2020.3034268.

[18] E. Ebrahimzadeh, F. Blaabjerg, X. Wang, and C. L. Bak, "Harmonic stability and resonance analysis in large PMSG-based wind power plants," IEEE Transactions on Sustainable Energy, **9**(1), 12-23, 2017, DOI: 10.1109/TSTE.2017.2712098.

[19] H. Abusannuga, and M. Özkaymak, "Performance of Vertical Axis Wind Turbine Type of Slant Straight Blades," Advances in Science, Technology and Engineering Systems Journal, **6**(4), 292–297, 2021, DOI:10.25046/aj060432.

[20] M. H. Rashid, Power electronics: circuits, devices, and applications., Pearson Education India, 2009.

[21] Ned, M., Ullmann, F., & Robbins, Power electronics., Pearson Education India, 2003.

[22] Simulink reference, from https://www.mathworks.com/help/pdf_doc/simulink/simulink_ref.pdf

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 3, 212-217 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Minimum Static VAR Compensation Capacity for Bad Voltage Drop Buses in Power Systems

Hermagasantos Zein*, Ahmad Deni Mulyadi, Achmad Mudawari

Jurusan Teknik Konversi Energi, Politeknik Negeri Bandung (Polban), Bandung, Kotak Pos 1234, Indonesia

ARTICLE INFO

Article history:
Received: 28 February, 2023
Accepted: 11 May, 2023
Online: 25 June, 2023

Keywords:
Quality
Bad voltage
Compensation
Power flow
Compensating modeling

ABSTRACT

The quality of the electric power system must be maintained properly, one of which is voltage. Under certain operating conditions, the bus voltage may drop below its minimum level, called bad voltage. A large reactive load can cause a voltage drop across the bus or the location of the bus is far from the generator stations, so that the line impedance has a large value. One technique to increase the voltage is through compensation technique. This paper presents the determination of the minimum statistical VAR compensation for increasing the voltage to the minimum safety limit. The methodology creates a compensating model for bad bus voltages. Electrical quantities (voltage, power and system losses) *are calculated through the power flow technique. The compensating capacity is increased until the voltage rise reaches its minimum security limit. The simulation results on the IEEE 9 bus system show that the voltage increases on all buses with minimal compensation on buses 5 and 8, and can save up to 1.37 MW of power*

1. Introduction

The electric power system must maintain its quality. One of its qualities is voltage, where voltage must be kept within its safety limits (minimum and maximum limits). In addition, the operation of the electric power system is highly desirable to be safe, efficient, and economical.

The existence of reactive power flow in the network will worsen the quality of service because it will cause voltage drops. An improved voltage profile by controlling the reactive power flow needs to be considered. Reactive power compensation (RPC) in power systems can increase voltage in addition to improving power factor [1]. However, determining the amount of effective compensation needs to be considered in maintaining voltage security. Maintaining voltage stability through reactive power control has been extensively discussed in [2–4]. This work succeeded in controlling the voltage at the desired level to improve the quality of the voltage to the customer. In electric power, the voltage must not drop below its minimum level. Voltage correction through compensation must consider the voltage limit. The effect of compensation can effects voltage fluctuations and it is an option to support reactive power flow into the network [5–7]. The reactive power flow in the network can be shown to decrease when it is compensated. However, the eff ect of reactive power flow on the voltage drop has not been clearly formulated.

*Corresponding Author: Hermagasantos Zein; hermaga_s@yahoo.co.id

Intelligent algorithm to control voltage has been studied in [8–10]. The [8] has developed an intelligent algorithm to control voltage. Meanwhile, controlling the flow of reactive power through the application of intelligent algorithms has been implemented in [9]-[10]. These works have successfully applied artificial methods to control the voltage on the demand side.

Whereas in [11–16], compensation techniques are used to control the voltage in the distribution system. The voltage control strategy to maintain the continuity of the distribution network [11] and for optimizing the voltage coordination [12]-[13]. Generally, the centralized voltage control is a good option. Multi-agent techniques can be used for centralized voltage control in power systems as discussed in [14]. The coordinated voltage control for fast location determination is a problem and the solution has been developed in the scheme in [15]. Meanwhile, the voltage control globally in a coordinated online system has been proposed in [16]. One of the problems in the operation of electric power is how to keep the voltage from going out of bounds. If there is a voltage below the minimum level, it must be increased to the minimum level.

Compensation technique can repair voltage drop in the medium voltage feeder [17,18] The voltage drop monitor of the load side can been implemented by using compensation [17]. The simple voltage drop calculation is presented in [18]. The methodology uses the feed current approach. While in [19–22] has studied the optimal compensation to voltage control. The [19-20] presents the

formulation to control the optimal reactive power. Repair power factor and optimal saving losses depend on the location of capacitor bank in the line feeder [21]. Power system large losses can be repaired by compensation. Optimal compensation to get minimum system losses has been clearly formulated in [22]. Considering that the cost of compensation in high-voltage networks is quite expensive, it is necessary to pay attention to the minimum compensation option so that voltage security can be maintained.

This paper proposes a method to improve voltage quality through minimum compensation in power systems using SVC devices on buses that have bad voltage. The methodology is based on compensation modeling on the observed bus, where the power supply comes from the network, generator and SVC to meet the load.. The compensating capacity is updated until the voltage level has reached the minimum security limit. Each process changes in compensation capacity, the voltage value is calculated using the Newton-Rapson technique. This technique has proven to be robust in power flow calculations in power systems.

This SVC device will act as a reactive power generator that supplies the network. The reactive power flow before the SVC device is installed will be compensated by the reactive power generated by the SVC device. In addition to increasing the voltage, this compensation technique also reduces system losses. In theory, the loss vs compensation curve is as a saddle horse. System losses will decrease towards optimal compensation, but system losses will increase if compensation is greater than optimal compensation. While the voltage tends to rise when compensation increases. The proposed method formulation has been tested with an IEEE 9-bus power system. The test results can increase the voltage to the minimum security limit (0.9 pu) on bad voltage buses (buses 5 and 8) and save system losses of 1.37 MW.

The systematics of this paper consists of 5 parts. The first session explained the state of the art regarding voltage repair through compensation techniques. The third paragraph from the end of this section explains the contribution of the work done.. The second paragraph at the end of this section describes the contribution of the work performed. The second session is to define the problem formulation. This section consists of Newton-Rapson technique, modeling of SCV and effect SCV to system losses. The third part contains procedures for solving problems in the form of a flowchart. The fourth session discussed the simulation results of the IEEE 9-bus power system. The last, fifth session, is to provide conclusions based on the findings in this work.

2. Problem Formulation

2.1. Newton-Raphson Technique

Refer to [22], the power flow problem is to solve the work network equation in (1).

$$\begin{bmatrix}\Delta P\\ \Delta Q\end{bmatrix} = \begin{bmatrix}J_1 & J_2\\ J_3 & J_4\end{bmatrix}\begin{bmatrix}\Delta\delta\\ \Delta V\end{bmatrix} \quad (1)$$

where P and Q are the active and reactive powers. δ and V are the complex voltage. $J_1, J_2, J_3,$ and J_4 are Jacobean matrices. By using the Newton-Raphson technique, the solution to the network equation through an iteration process will converge when the power balance is reached. The power balance is expressed in (2) and (3).

$$\Delta P = P^{fix} - P^{count} \leq \varepsilon \quad (2)$$

$$\Delta Q = Q^{fix} - Q^{count} \leq \varepsilon \quad (3)$$

where ε is the smallest number, it represents the degree of accuracy.

2.2. Modeling of SVC

The SVC model is shown in Figure 1. Assumed that the active and reactive power flows from the grid are P_{gd-i} and Q_{gd-i}, respectively. From the figure, the convergent power flow will satisfy (4) and (5).

$$P_{d-i} = P_{g-i} + P_{gd-i} \quad (4)$$

$$Q_{d-i} = Q_{g-i} + Q_{gd-i} + Q_{SVC-i} \quad (5)$$

The current and capacity of capacitor can be derived from Figure 1, namely:

$$I_{SC-i} = 2\pi f C_{SC-i} V_i \quad (6)$$

$$Q_{CSVC-i} = B_{SVC-i} V_i^2 \quad (7)$$

Where f is the system frequency, C_{SC-i} is capacitance of SCV and B_{SVC-i} is the susceptance of SVC with the formulation is,

$$B_{SVC-i} = 2\pi f C_{SC-i} \quad (8)$$

Changes in voltage can be caused by the SCV capacity. The value of this voltage change can be derived from (7) and yield,

$$\partial V_i = \frac{\partial Q_{CSVC-i}}{2 Q_{SVC-i}} V_i \quad (9)$$

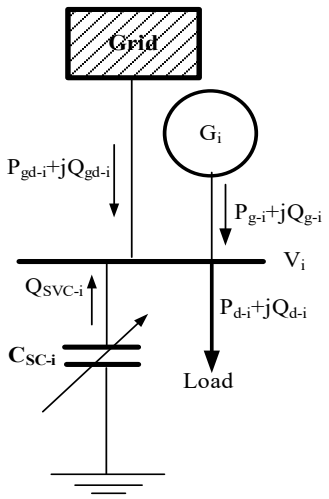


Figure 1: SCV model

While the susceptance change is,

$$\partial B_{CSVC-i} = \frac{B_{SVC-i}}{Q_{CSVC-i}} \partial Q_{CSVC-i} \quad (10)$$

From (9) and (10), the value of the voltage change is obtained,

$$\Delta V_i = \frac{V_i}{2B_{SVC-i}} \Delta B_{CSVC-i} \quad (11)$$

When the bus voltage drops below the minimum security limit, it must be increased at least to the lower limit to maintain the quality of the power flow. In this paper, the voltage increase is limited to the lower limit so that a minimum SCV capacity is required.

$$V_i^{new} = V_i^{old} + \Delta V_{i-min} = V_{i-min} \quad (12)$$

In this proposed method, the value of V_i^{new} already meets the minimum-security limit voltage (V_{i-min}), where previously (V_i^{old}) is below the minimum-security limit voltage.

From (9) and (12) it can be derived the change in the minimum SCV capacity to increase the voltage to the minimum-security limit, namely:

$$\Delta Q_{cSVC-i-min} = 2 \frac{Q_{SVC-i}}{V_i} \Delta V_{i-min} \quad (13)$$

2.3. Effect SVC to Losses

Calculation of the effect SVC to losses refer to [22]. The line representation is shown in Figure 2. SCV generates reactive power in the network so that it will have an impact on line losses.

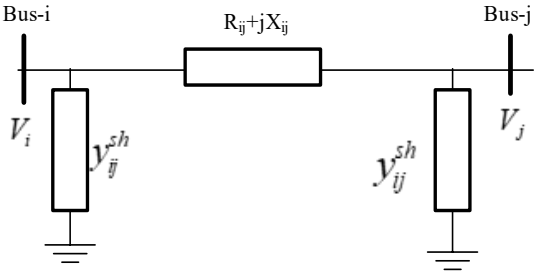


Figure 2: Phi equivalent circuit of line i-j

From Figure 2, line losses can be derived and yield,

$$Pl_{ij} = \frac{R_{ij}}{|Z_{ij}|} \left(V_i^2 - 2V_iV_jCos\delta_{ij} + V_j^2\right) \quad (14)$$

So for power system with n-buses, the total losses in power systems is,

$$Pl = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j>i}}^{n} \frac{R_{ij}}{|Z_{ij}|} \left(V_i^2 - 2V_iV_jCos\delta_{ij} + V_j^2\right) \quad (15)$$

Or,

$$Pl = \sum_{i=1}^{n_g} P_{g-i} - P_d \quad (16)$$

where P_{g-i} is the power generated by generator i, and P_d is the total load. The effect of the voltage on the bus k on the minimum system losses is shown in Figure 3. Increasing the voltage V_j in (15) will reduce losses.

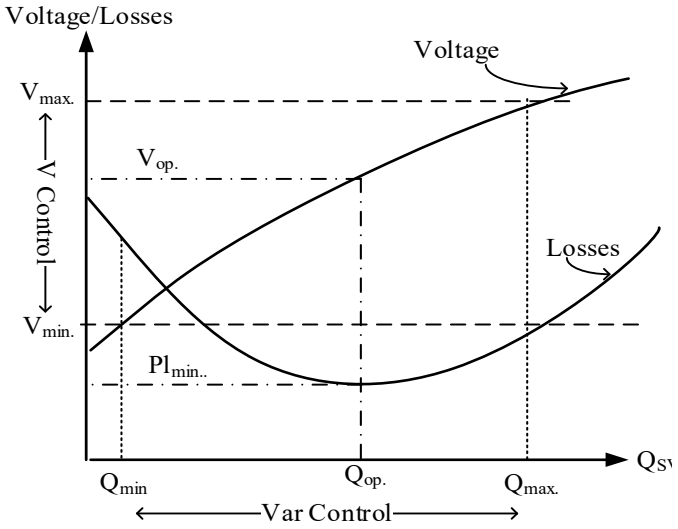


Figure 3: Bus voltage control

From Figure 3 it can be seen that the addition of SCV capacity will increase the voltage. Raising the voltage to the minimum safety limit (V_{min}) requires the SCV capacity of Q_{min}. The effect of the voltage on the bus k on the system losses can be derived from Equation 15 and yield,

$$\frac{\partial Pl}{\partial V_k} = \sum_{\substack{i=1 \\ i \neq k \\ i \in k}}^{n} \frac{R_{ik}}{|Z_{ik}|} (-2V_iCos\delta_{ik} + 2V_k) \quad (17)$$

Meanwhile based on Equation 7 for Q_{SVC} compensation on bus k can be defined,

$$X = \frac{Q_{SVC}}{B_{SVC}} = V_k^2 \quad (18)$$

Substitute V_k for X from Equation 18 to Equation 16 and yield,

$$\partial X = \frac{\partial Pl}{\sum \frac{R_{ik}}{|Z_{ik}|}(X - Cos\delta_{ik})} \quad (19)$$

Then obtained ΔX through Equations 16 and 19, namely:

$$\Delta X = \frac{\sum P_{Gi}^{old} - \sum P_{Gi}^{new}}{\sum \frac{R_{ik}}{|Z_{ik}|}\left(1 - \frac{V_i}{\sqrt{X}} Cos\partial_{ik}\right)} \quad (20)$$

Finally the value of X can be updated,

$$X^{new} = X^{old} + \Delta X \quad (21)$$

Minimum compensation is reached if,

$$X_{k-min}^{SCV} = V_{k-min}^2 \quad (22)$$

where X_{k-min}^{SVC} is the minimum compensation. V_{k-min}^2 is the lower voltage limit at bus k.

3. Flowchart of Proposed Method

The procedures for completing this work will follow the step by step diagram in Figure 4. Base case is determined by the results of calculating the power flow without SVC. Initial conditions are obtained from the base case. The voltage value is calculated from the updated compensation (X). This process is repeated until the voltage has reached the value of the minimum limit security (V_{min}). Repeat for bad voltage drop on other buses. Finally, the calculation results are obtained: the minimum SCV capacity.

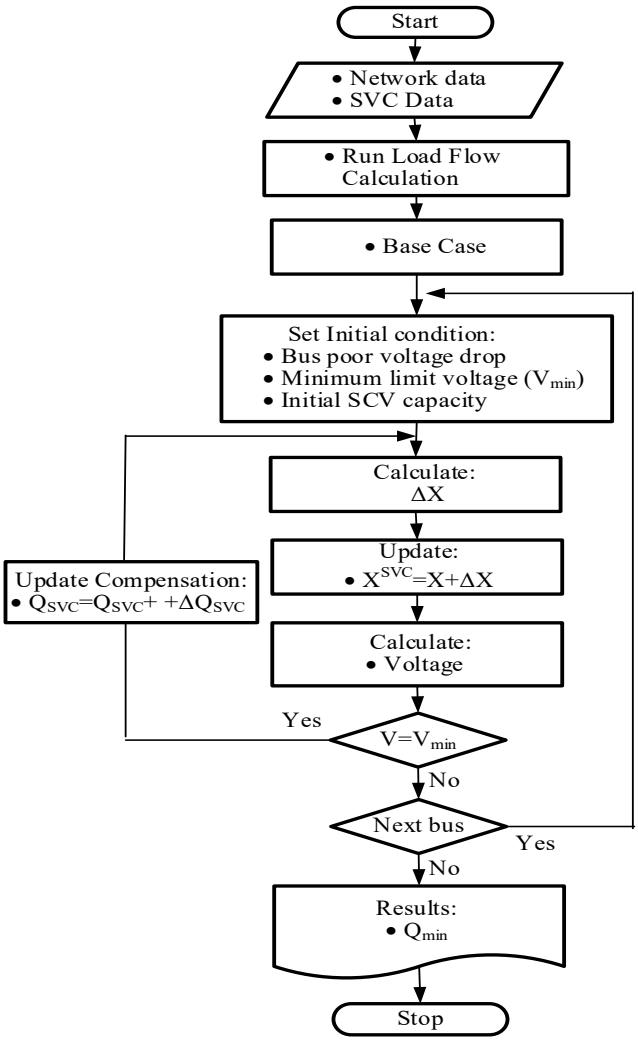


Figure 4: Flowchart of the proposed method

4. Simulation, Results and Discussion

The simulation is carried out through application software based on the flowchart in Figure 4 using the FORTRAN language. The software application is run using an Asus Core i3 laptop.

4.1. Simulation

The proposed method has been tested through simulations on electric power systems in order to validate it. This simulation uses an IEEE 9-bus system with a one-line diagram in Figure 5. The data for this system are presented in Tables 1 and 2. In particular, Table 1 is the result of load flow based on the IEEE standard and is used as a basis for calculations. While Table 2 shows the line characteristics. From the base case the voltage at buses 5 and 8 is below the minimum-security limit (0.90 pu). For this reason, minimum compensation is needed on these two buses to increase the voltage to the minimum-security limit.

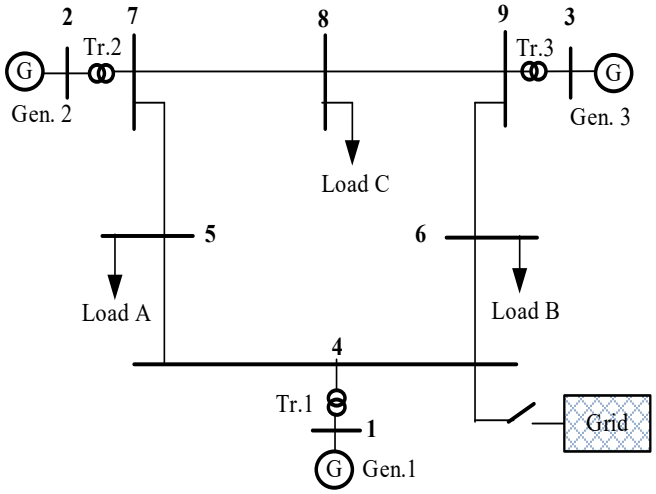


Figure 5: IEEE 9-bus power system

Table 1: Base case

No bus	V [pu]	P_g [MW]	Q_g [MVAR]	P_d [MW]	Q_d [MVAR]
1	0.950	0	0	0	0
2	1.025	163.00	130.75	0	0
3	1.025	85.00	192.92	0	0
4	0.953	0	0	0	0
5	**0.872**	**0**	**0**	**200.0**	**100.0**
6	0.909	0	0	90.0	30.0
7	1.040	391.84	203.21	0	0
8	**0.884**	**0**	**0**	**150.0**	**95.00**
9	0.916	0	0	180.0	130.0
Total		639.84	526.88	620.0	335.0

Table 2: Line characteristic

From bus	To	R(pu)	X(pu)	1/2Y(pu)
1	2	0	0.0625	1.000
2	8	0	0.0625	1.000
3	6	0	0.0586	1.000
4	5	0.0085	0.0720	0.0745
4	9	0.0320	0.6110	0.1530
5	6	0.0170	0.0920	0.0790
6	7	0.0170	0.0920	0.0790
7	8	0.0170	0.0920	0.0790
8	9	0.0320	0.6110	0.1530

4.2. Results

The comparison of the results of the base case simulation without the SCV device is shown in Table 1. This table indicates the need for compensation for buses 5, and 8 because low voltage. From the results of the simulations carried out to increase the

voltage on buses 5 and 8, the minimum compensation values are obtained as shown in Table 3.

Table 3: Minimum compensation

#	Before Compensation			After Compensation		
	V (pu)	B_{SCV} (pu)	Q (MVar)	V (pu)	B_{SCV} (pu)	Q (MVar)
Bus 5	0.872	0	0	0.9	0.1978	21.0
Bus 8	0.884	0	0	0.9	0.0942	10.0
f_d	0.771958			0.803636		
Losses	19.84 MW			18.47 MW		

Note: f_d is the system power factor.

The load flow simulation results after compensation are presented in Table 4. The gain from this compensation is that the system loss drops by 1.37 MW.

Table 4: Power flow after compensation

No bus	V [pu]	P_g [MW]	Q_g [MVAR]	P_d [MW]	Q_d [MVAR]
1	0.961	0	0	0	0
2	1.025	163.00	112.77	0	0
3	1.025	85.00	179.83	0	0
4	0.965	0	0	0	0
5	0.900	0	0	200.0	79.0
6	0.921	0	0	90.0	30.0
7	1.040	390.47	180.21	0	0
8	0.900	0	0	150.0	85.00
9	0.923	0	0	180.0	130.0
Total		638.47	472.81	620.0	324.0

4.3. Discussion

The proposed method has successfully simulated the IEEE 9 bus system, which is presented in Tables 3 and 4. Table 3 presents the results of the minimum compensation on buses 5 and 8. The previous voltages (based on the base case) were below the minimum safety limits of 0.872 and 0.884 pu, all marked in bold. In the simulation using voltage security 0.9-1.1 pu, so the voltage on the bus becomes bad. To increase the voltage on buses 5 and 8 installed SCV. The SCV capacities required to increase the voltage to the minimum safety level (0.9 pu) are 0.1978 pu and 0.0942 pu, respectively.

The effect of Compensation (see Table 3) is to increase the voltage, not only at the bus being compensated but also at other buses. For example Bus-1 has a voltage of 0.95 pu (see Table 1) before compensation increasing to 0.961 (see Table 4) after compensation. In addition, shown by Table 3, the system power factor increased from 0.803636 to 0.771958 and losses decreased from 4MW to 3 MW or losses were saved by 6.9%.

This increase in voltage needs to be watched out for when it exceeds its maximum limit. If there is an overvoltage due to compensation, the compensation capacity is stopped until one/several bus voltages reach their maximum limit. Furthermore, the load shedding operation needs to be considered to increase the voltage on buses that have the bad voltage.

Besides that, compensation has advantages because it reduces losses. The load flow simulation after compensation in Table 4 states that total power production degrease from 639.84 MW to 638.47 MW or saving 1.37 MW. When these savings are converted to costs, the value is quite large.

Comparison with other results from [23] is shown in Table 5, where compensation on bus 8. The amount of compensation based on the reference is 27.46 MVar to increase the voltage at bus 5 by 0.993 pu compared to the proposed method requiring compensation of 10.00 MVar and the voltage at bus 5 increases to 0.921 pu.

Table 5: comparison results with another method

Item	Base Case	Reference Method	Proposed Method
		SVC at Bus 8	SVC at Bus 8
Losses (MW)	5.300	4.773	4.5389
Q_{SCV} (MVar)	0	27.46	10.00
Bus	V(pu)	V(pu)	0.961
1	0.950	1.040	1.025
2	1.025	1.025	1.025
3	1.025	1.025	0.965
4	0.953	1.024	0.900
5	**0.872**	**0.993**	0.921
6	0.909	1.010	1.040
7	1.040	1.023	0.900
8	0.884	1.000	0.923
9	0.916	1.027	0.961

This proposed method should be supported by fast and robust power flow calculations. There are three power flow methods that should be considered, namely the Gauss-Seidel, Newton and Fast Decouple methods, The Gauss-Seidel method uses an elimination technique that involves all parameters and variables so that the calculations at each iteration step will process large power system data. In addition, it takes many iteration steps to reach convergence. However, this method will be effective for small power system sizes. Newton's method is called the complete power flow method because it uses a complete Jacobean matrix. This causes a large memory usage. The fact that the changes in the calculation results to reach the solution point are quite small so that many iteration steps are needed. Whereas the Fast Decoupled method works by ignoring the out diagonal sub matrix of the Jacobean matrix and the element values of the Jacobean matrix are constants through the approach $\cos\delta$=1 and $\sin\delta = 0$ for $\delta \leq 15^0$. Through the Fast Decoupled method approach, it can work very quickly with a few iteration steps. The condition needed by this method is that each line must meet r/x<1, where line impedance is z = r + jx.

The IEEE 9 bus system simulation results for the three methods are shown in Table 6. The Fast Decoupled method can work very fast, so it is very advantageous to apply it to real time calculations. However, this method does not work well for the line impedance conditions r/x>1. If the Fast decoupled method cannot work, it is necessary to consider the full method option (Newton method).

Table 6: Comparison of three load flow methods

Item	Gauss-Seidel	Newton	Fast Decoupled
Process	Iteration	Iteration	Iteration
Technique	Elimination	Full Jacobean	Ignore out diagonal of sub-Jacobean. Approximate sin and cost values for small angles.
Accuracy	0.00000001	0.00000001	0.00000001
Iteration step	28	25	9.5
Running time	3.715s	3.125s	0.5s

5. Conclusion

The method of calculating the minimum compensation in the electric power system has been presented in this paper. Formulation validation has been tested through simulation of electric power systems in section 4 of this paper.

This work uses the IEEE 9-bus power system, based on the base case there are bad voltages on buses 5 and 8. SCV capacities of 0.19 and 0.9 pu are required to increase the voltage to the minimum safety level (0.9 pu) on bus 5 and 8 respectively. The effect of compensation is to increase the voltage across all buses. It should be noted that the voltage increase at the bus does not exceed the maximum-security limit. The addition of compensating capacity must be stopped if the bus voltage reaches its maximum safety limit.

The advantage of compensation is the reduction of system losses. This is caused by a decrease in reactive flow in the network so that the current is reduced. From the simulation results obtained a power saving of 1.37 MW to increase the voltage to the minimum limit level (0.9), where the voltage on bus 5 and 8 is 0.872 pu and 0.884 pu.

References

[1] T. Gonen, Electrical Power Distribution System Engineering, McGraw-Hill, 1986.

[2] X. YE, J. LE, Y. LIU, W. ZHOU, K. LIU, "A coordinated consistency voltage stability control method of active distribution grid," Journal of Modern Power Systems and Clean Energy, **6**(1), 85–94, 2018, doi:10.1007/s40565-017-0294-z.

[3] J. Dixon, L. Moran, J. Rodriguez, R. Domke, "Reactive Power Compensation Technologies: State-of-the-Art Review," Proceedings of the IEEE, **93**(12), 2144–2164, 2005, doi:10.1109/JPROC.2005.859937.

[4] P. Kundur, Power System Stability and Control, McGraw-Hill, New York, 1994.

[5] D. Quoc Hung, Y. Mishra, "Voltage fluctuation mitigation: fast allocation and daily local control of DSTATCOMs to increase solar energy harvest," IET Renewable Power Generation, **13**(14), 2558–2568, 2019, doi:10.1049/iet-rpg.2019.0223.

[6] M. Alrashidi, "Community Battery Storage Systems Planning for Voltage Regulation in Low Voltage Distribution Systems," Applied Sciences, **12**(18), 9083, 2022, doi:10.3390/app12189083.

[7] M. Aybar-Mejía, J. Villanueva, D. Mariano-Hernández, F. Santos, A. Molina-García, "A Review of Low-Voltage Renewable Microgrids: Generation Forecasting and Demand-Side Management Strategies," Electronics, **10**(17), 2093, 2021, doi:10.3390/electronics10172093.

[8] R. Anilkumar, G. Devriese, A.K. Srivastava, "Voltage and Reactive Power Control to Maximize the Energy Savings in Power Distribution System With Wind Energy," IEEE Transactions on Industry Applications, **54**(1), 656–664, 2018, doi:10.1109/TIA.2017.2740850.

[9] K.-D. Kim, W.-K. Yu, E.-J. Lee, I.-S. Wang, H.-J. Kang, J.-H. Shin, H.-J. Lee, "Development of an Intelligent Voltage Control System for Jeju Island in Korea," Journal of International Council on Electrical Engineering, **1**(2), 156–162, 2011, doi:10.5370/JICEE.2011.1.2.156.

[10] H. Van Pham, S.N. Ahmed, Multi-Agent based Approach for Intelligent Control of Reactive Power Injection in Transmission Systems, John Wiley & Sons, Ltd, Chichester, UK: 269–282, 2018, doi:10.1002/9781119214984.ch13.

[11] Y. Zhang, A. Srivastava, "Voltage Control Strategy for Energy Storage System in Sustainable Distribution System Operation," Energies, **14**(4), 832, 2021, doi:10.3390/en14040832.

[12] A.K. Ali, M.M. Mahmoud, "Coordinated Distributed Voltage Control Methods for Standalone Microgrids," International Journal of Robotics and Control Systems, **2**(2), 262–276, 2022, doi:10.31763/ijrcs.v2i2.612.

[13] M. Hojati Tabatabaei, H. Siahkali, J. Olamaei, "A quick solution to optimal coordinated voltage-control, based on dimension-reduction of power system via Modified Ward-PV," Scientia Iranica, **0**(0), 0–0, 2021, doi:10.24200/sci.2021.56970.4999.

[14] H. Shahbazi, F. Karbalaei, "Decentralized Voltage Control of PowerSystems Using Multi-agent Systems," Journal of Modern Power Systems and Clean Energy, **8**(2), 249–259, 2020, doi:10.35833/MPCE.2018.000628.

[15] J.Y. Wen, Q.H. Wu, D.R. Turner, S.J. Cheng, J. Fitch, "Optimal Coordinated Voltage Control for Power System Voltage Stability," IEEE Transactions on Power Systems, **19**(2), 1115–1122, 2004, doi:10.1109/TPWRS.2004.825897.

[16] K. Alzaareer, M. Saad, H. Mehrjerdi, D. Asber, S. Lefebvre, "Development of New Identification Method for Global Group of Controls for Online Coordinated Voltage Control in Active Distribution Networks," IEEE Transactions on Smart Grid, **11**(5), 3921–3931, 2020, doi:10.1109/TSG.2020.2981145.

[17] C.K. Wachjoe, H. Zein, "A Method for Voltage Drop Monitoring on Load Sides in Medium Voltage Feeder," in 2020 7th International Conference on Control, Decision and Information Technologies (CoDIT), IEEE: 1215–1220, 2020, doi:10.1109/CoDIT49905.2020.9263852.

[18] R. Thakur, P. Chawla, Voltage Drop Calculations & Design of Urban Distribution Feeders.

[19] X.Y. Xu, P. Zhou, Q.Z. Huang, C.M. Deng, M.M. Shi, G. Xue, "Optimization of Voltage and Reactive Power in Regional Power Network Based on MCR," Advanced Materials Research, **986–987**, 394–399, 2014, doi:10.4028/www.scientific.net/AMR.986-987.394.

[20] S.-I. Go, S.-Y. Yun, S.-J. Ahn, J.-H. Choi, "Voltage and Reactive Power Optimization Using a Simplified Linear Equations at Distribution Networks with DG," Energies, **13**(13), 3334, 2020, doi:10.3390/en13133334.

[21] H. Zein, E. Dermawan, "Optimal Location of a Capacitor Bank in a Primary Feeder to Load up Linearly," IPTEK Journal of Proceedings Series, **0**(1), 2014, doi:10.12962/j23546026.y2014i1.370.

[22] H. Zein, S. Saodah, S. Utami, C.K. Wachjoe, "Static VAR Compensation Capacity for Bus Voltage Setting in Electric Power Systems," in 2022 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP), IEEE: 1–6, 2022, doi:10.1109/ICT-PEP57242.2022.9988871.

[23] J.V. and S.A.S. Parate, "Reactive Power Control and Transmission Reduction with Realization of SVC and TCSC," International Journal of Engineering Science and Technology (IJEST), **4**(7), 3592–3600, 2012.

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 3, 231-243 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Waterfall: Salto Collazo. High-Level Design of Tokenomics

Sergii Grybniak[1], Yevhen Leonchyk[2], Igor Mazurok[2], Oleksandr Nashyvan[1], Alisa Vorokhta[*,2]

[1]*Institute of Computer Systems, Odesa Polytechnic State University, Shevchenko av. 1, Odesa, 65044, Ukraine*

[2]*Faculty of Mathematics, Physics and Information Technologies, Odesa I.I. Mechnikov National University, Dvoryans'ka St 2, 65082, Odesa, Ukraine*

ARTICLE INFO

Article history:
Received: 28 February, 2023
Accepted: 20 May, 2023
Online: 25 June, 2023

Keywords:
Tokenomics
Incentive System
Distributed Protocol
Transaction Fee
Directed Acyclic Graph
Blockchain

ABSTRACT

This article explains the fundamental principles of the economic policy that are integrated into the decentralized public platform Waterfall. The platform has a DAG (Directed Acyclic Graph) based system architecture and is designed to develop decentralized applications and financial services. The main goal of this work is to create a favorable environment that incentivizes positive behavior from each network participant and from the system as a whole. Economic leverages ensure general equilibrium to provide an optimal data replication ratio, attack protection, and affordable transaction fees. Although this model of tokenomic is designed explicitly for the current version of the Waterfall platform named Salto Collazo, the presented approaches possess the potential to be applied across a broad spectrum of decentralized public platforms, owing to their inherent transparency and a set of tuned parameters.

1. Introduction

1.1. Problem Statement

This paper is an extension of work originally presented in the proceedings of the IEEE International Conference on Blockchain, Smart Healthcare, and Emerging Technologies [1] and presents a high-level economic design of decentralized public networks such as blockchains, blockDAGs (Directed Acyclic Graphs consisting of linked blocks), etc.

The emergence of blockchain technology has transformed modern society and businesses, primarily thanks to the attributes of its transparency, tamper-proof safety of data, and logical consistency [2]-[4]. However, its subpar network performance and exorbitant commissions make it unsuitable for many enterprise-class applications. On the other hand, DAG technology is considered the next generation of blockchain due to its optimized validation mechanism, high scalability, efficient provenance, and multiparty involvement [5]-[7]. A DAG-based architecture can provide the necessary functions to develop a variety of decentralized finance (DeFi) services [8], [9], including payment systems, non-fungible tokens (NFTs), web3 gaming [10], new solutions in logistics [11] and real estate sector [12], identity documents [13], and e-voting services [14]. The success of this approach depends on the establishment of ad-hoc economics for a decentralized storage system that is both high-speed and scalable, while also ensuring low-cost transaction fees. The primary focus of this article is on the core economic principles embedded within the Waterfall DAG-based protocol [15] and its efficacy in fulfilling these specific requirements.

1.2. An Exposition of the Waterfall Platform: A Short Survey

Waterfall is a decentralized platform designed for developing a variety of decentralized applications (Dapps). The platform features high scalability and is built on DAG technology, a distributed ledger system that differs from traditional blockchain technology in that it does not rely on a single chain of blocks. Instead, DAG technology uses a graph-like structure, enabling the platform to process transactions in parallel, without the need for central authority.

Waterfall leverages a fast finality Proof-of-Stake (PoS) [16] consensus algorithm to ensure the integrity of its transactions. The platform is built to rely on Coordinating and Sharding networks, enabling a significant volume of transactions via parallelized block generation. The Coordinating network processes transactions and communicates with the Validator network to ensure consensus on the validity of each block. Meanwhile, the Sharding network

*Corresponding Author: Alisa Vorokhta, Dvoryans'ka St 2, 65082, Odesa, Ukraine, +380951641320 & alisa.vorokhta@stud.onu.edu.ua

divides the workload across multiple nodes, improving the platform's overall scalability.

Each node in the Waterfall network is composed of 2 components: a Coordinator and a Validator, representing their respective roles within the system. The Sharding network is responsible for processing incoming transactions and ensuring their validity. Transactions are combined into blocks that refer to other blocks to form a DAG. Information about the created blocks enters the Coordination Network for structure linearization and finalization based on consensus. The blockchain of the Coordinating network records data on the finalization of blocks in all shards.

Overall, the Waterfall platform is a highly-scalable smart contract platform. DAGs in decentralized technologies facilitate scalability, addressing one of the key challenges faced by such systems, and allow for the parallel processing of transactions, which can be verified and confirmed quickly and efficiently. This feature of the platform enables it to process a high volume of transactions per second, making it a viable option for decentralized applications.

The Waterfall consensus protocol Gozalandia [17] has established a time-based system to ensure efficient and secure block production. This timeline is composed of distinct time intervals, including slots, epochs, and eras. The Coordinating network assumes responsibility for registering Validators and assigns roles such as block producers, committee members, and leaders during the commencement of each epoch. In addition, the Coordinating network maintains information regarding the validated blocks generated on the Sharding network. The honest producer is required to include links to all the known tip-blocks of the DAG to its created block. Furthermore, the Coordinating network performs ordering and finalization of the distributed ledger, enhancing the synchronization and overall network security.

The Waterfall platform incorporates and refines the most advantageous properties of Ethereum 2.0 [18], augmenting its performance and extending its functionality. The platform's inherited coin serves as a primary digital asset within the network, facilitating the transfer of transactions and governance voting, executing smart contracts, and providing for the creation of auxiliary tokens, thereby creating an ecosystem with the potential for mutually beneficial interaction among all its components. Nonetheless, the structure of the DAG platform requires new approaches to tokenomics and economic incentives, as traditional mechanisms may be incompatible or insufficient to support the platform's decentralized nature and ensure its sustainable growth. Therefore, the Waterfall team must continually research and implement new solutions to empower its users and incentivize positive behavior, thereby promoting the longevity and stability of the system.

1.3. Tokenomics Strategies: Achieving Sustainable Development in Decentralized Systems

Tokenomics is a fundamental concept in decentralized systems that leverages economic incentives to encourage desired behaviors from participants, such as users, developers, and Validators [19]. The goal is to create a self-sustaining and self-regulating ecosystem that maximizes benefits for all stakeholders. Tokenomics involves the use of native tokens or coins that serve as the currency of the decentralized network and facilitate transactions and interactions within the system. A well-designed tokenomics model can incentivize positive actions, such as network participation, governance voting, and contribution to the development of the network. It can also disincentivize malicious behavior by imposing penalties for non-compliance with network rules and regulations. Additionally, tokenomics can facilitate the effective allocation and use of existing and new resources on the network, resulting in a variety of affordable services for users. Therefore, a robust and balanced tokenomics model is crucial for the successful operation and growth of public decentralized networks [20], [21].

The concept of Waterfall tokenomics is a set of economic principles and mechanisms designed to support the growth and functionality of distributed networks, with a particular emphasis on DAG structures. Waterfall tokenomics has specific goals aimed at creating economic conditions that promote scalability, speed, security, and reliability.

One of the primary objectives of Waterfall tokenomics is to establish economic conditions that enable the network to expand to a size that provides an optimal data replication coefficient and a maximum speed of mempool synchronization. This objective is especially critical for networks that utilize edge networking, which entails a high volume of transactions that must be processed and propagated efficiently.

Another important goal of Waterfall tokenomics is to promote fast finality [22], which plays a pivotal role in facilitating payment systems and decentralized applications (Dapps). Fast finality ensures that transactions are secure and reliable by preventing any alteration or reversal once a transaction is confirmed.

It is worth noting that the DAG structure of the network presents unique challenges and vulnerabilities that differ from those of traditional blockchain architectures. Therefore, Waterfall tokenomics must provide proper protection and security in intimate collaboration with the technical network architecture.

In summary, the concept of Waterfall tokenomics is a vital economic framework for distributed networks, with an emphasis on DAG structures. Its goals are aimed at creating economic conditions that promote scalability, speed, security, and reliability, and it plays a crucial role in supporting payment systems and Dapps. Finally, it must work closely with the technical network design to ensure appropriate protection against possible attacks.

The vertices in bold are spine blocks in Figure 1. These blocks precede all other blocks in their slots after linearization. We used methods of statistical analysis and simulation to study the probability of successfully executed attacks (considered below) and the degree of damage caused by them. This allows us to evaluate the effectiveness of developed countermeasures and analyze the statistical properties of the acquired graph.

The set of rules and guidelines that govern the network economics is implemented through software that encompasses all the essential features required for its operation. These economic rules are autonomously enforced and transparently visible to the general public. This feature increases user confidence in the platform and makes it more resilient. Additionally, there must be mechanisms in place to adapt specific rules dynamically to changing situations. This approach ensures that the network's ability to sustain desirable behavior over its entire lifespan, fostering overall economic equilibrium and systematic development aligned with the network's objectives. Achieving this goal poses a considerable challenge within the scope of this endeavor.

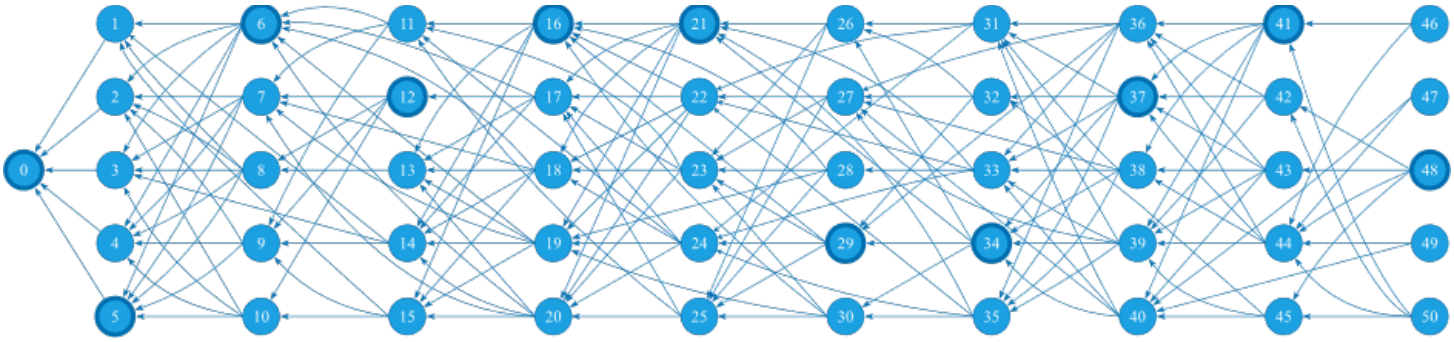


Figure 1: An example of a starting fragment of a DAG network created using simulation

2. Related Works

The evolution of economic blockchain models has been a subject of intense research in recent years. The choice of consensus protocol has important implications for the economic model of a blockchain network [23], [24]. For example, the development of Proof-of-Work (PoW) decentralized networks like Bitcoin and Ethereum 1.0 represents a significant milestone in the history of blockchain technology (as evidenced by numerous studies including [19], [25], [26]). In a PoW system, nodes compete to solve cryptographic puzzles, with the first node to solve the puzzle being rewarded with a new block and the corresponding amount of cryptocurrency. This creates an economic, incentivizing network node participation and contributes to the overall security of the system. This type of system is known for its simplicity but has also been criticized for its high energy consumption and vulnerability to 51% attacks. Early models were characterized by relatively simple economics compared to more modern PoS models, which have since emerged as a promising alternative that can potentially enhance network maintenance [27]. PoS systems use a different mechanism to achieve consensus, whereby nodes are selected to create new blocks based on their ownership of a certain amount of cryptocurrency. In such models, network participants are incentivized to hold a certain amount of cryptocurrency as collateral, and transactions are validated based on the stake they hold. Compared to PoW models, PoS models are less energy-intensive and less vulnerable to 51% attacks.

Overall, the development of economic blockchain models has been driven by the need to incentivize network participants to act in ways that contribute to the overall health and sustainability of the system. While early models like PoW and PoS have been successful in achieving this goal, research is continued on new consensus algorithms and economic models that can potentially offer even better performance and security. For example, there are also approaches to achieving consensus that have been proposed in the literature, such as the use of Byzantine Fault Tolerant (BFT) algorithms [28], DAGs [29], and Delegated Proof-of-Stake (DPoS) [30]. However, these approaches often require ad-hoc methods to incentivize network participants [31], which can make them more difficult to implement and manage.

Despite the differences among these models, they all share a common goal: incentivizing network participants to behave in a way that benefits the network as a whole. This challenge requires a deep understanding of game theory [32], [33] and the underlying properties of blockchain technology. By designing economic models that reward good behavior and penalize bad actors, these models aim to ensure the smooth and sustainable functioning of decentralized networks.

In response to modern demands, various economic models have been developed to support PoS-based systems ([34], [35], [36]). These models aim to address the challenges faced by decentralized systems, such as decreasing transaction costs, increasing transaction frequency, and reducing finalization time. Additionally, a few approaches have been proposed by incorporating economic principles into BFT-like consensus mechanisms, these systems aim to achieve a balanced internal economy that aligns with specific objectives ([37], [38], [39]).

After analyzing related works, it has been revealed that one of the primary challenges facing decentralized systems is the need to decrease transaction costs, increase transaction frequency, and reduce finalization time. This is particularly important for the mass adoption of Distributed Ledger Technology (DLT) in real-world applications. In Messari's report for 2023 [8], one can find a comprehensive examination of current prospects and significant trends in the development of tokenomics.

Therefore, our aim is to develop a tokenomics model that addresses these challenges, with a focus on limiting the increase in transaction fees while maintaining the Worker's economic viability, which includes blockchain Validators and block producers. To achieve this, we propose a flexible and transparent architecture that can be easily modified with a set of adjustable parameters. This is especially important for emerging public platforms, as they must be able to successfully meet formation requirements.

However, important to highlight that most of the already existing economic models are developed for established systems such as Ethereum 2.0 [18]. As such, there is a need for new and innovative models that take into account the specific requirements and limitations of emerging decentralized systems.

The Waterfall model incorporates and enhances the most favorable aspects of Ethereum tokenomics. Furthermore, it presents notable benefits owing to its novel consensus protocol [17] and horizontal scaling, with subnetworks necessitating economic incentives [40]:

- **Dynamic adjustment.** A mechanism for dynamic adaptation of system parameters is in place to ensure optimal network behavior, particularly by automatically adjusting the optimal number of Workers. This approach enables the system to attain self-sufficiency throughout its complete lifespan.
- **Low transaction fees.** In various scenarios, the architecture and tokenomics of the system are structured to maintain low transaction fees. The protocol scales dynamically with the increasing network load, which leads to the simultaneous publication of more blocks within the same slot and

consequently reduced transaction fees as the system expands. Thus, transaction processing during peak times remains efficient, while the number of pending transactions in the transaction pool remains low.

- **Supporting high transaction throughput.** The Waterfall platform achieves high transaction throughput via the use of system-scalable DAG-based block structures. These structures allow multiple blocks to be published simultaneously, forming a DAG that provides finality for all transactions, as long as the blocks do not conflict with each other. In recent intermediary lab tests, our protocol demonstrated the ability to process up to 3,600 transactions per second. The tokenomics model effectively supports this architecture.

In the world of public blockchain networks, the presence of corrupt and malicious nodes leads to an environment of complete distrust. To address this issue, a reputation system could be implemented to enhance the security of interactions. Several approaches have been proposed for building a reputation system based on blockchain data, as discussed in existing literature (e.g. [41], [42]).

In [43], a detailed discussion is presented on the reward distribution scheme among honest Workers for their role in block production and validation, as well as on the determination of penalties for faulty Workers, with particular attention paid to the consensus achievement mechanism of the Waterfall platform. The key principles underlying the proposed scheme are that Workers should be rewarded in proportion to the significance of their contributions, and that the penalties imposed on them should outweigh the potential benefits of any malicious actions. The tokenomics of the system is analyzed at a macro level, without delving into the specific functional roles of individual nodes. As such, the proposed approach is independent of any particular PoS-based consensus algorithm that opens up the possibility of its adaptation to a broad range of platforms for macroeconomic design, owing to a multi-parameter configuration.

3. Macroeconomic Design

This section delves into the analysis of the decentralized Waterfall platform from an economic standpoint, using a holistic approach that avoids focusing on the specific behaviors or contributions of individual participants. It covers the platform's governance framework, which includes core principles of network economic policy that shape the interactions of community groups and affect various economic indicators such as cryptocurrency rates, inflation, and deflation. The system architecture and coin circulation are visually represented in Figure 2 and are discussed in-depth later in this section.

3.1. Pre-mining Strategy

In the early stages of a network, the initial coin distribution is a critical element for ensuring system security. As such, a recommended approach is to create a set of N_{opt} Workers at the beginning, each with a fixed stake divided into nodes that share a ledger and IP address. The stake amount is denoted by s coins per Worker, and any increase in the total amount of the stake is achieved solely through the addition of new Workers. This strategy helps to ensure a secure and stable system during the network's initial phase.

Typically, the entire token supply is not immediately available at the outset. Assume α represents the proportion of the current supply (C) to the total staked amount $S_{opt} = s \cdot N_{opt}$. The value of N_{opt} should be determined as the present optimal number of Workers necessary to ensure a secure and efficient network. A restricted number of Workers in a network raises the risk of a majority holding attack. Conversely, a larger total stake results in a decrease in free funds in circulation, which can impede effective network operations. To maintain the desired level of available funds and the optimal number of Workers, it is crucial to uphold the ratio α. In other words if the current supply C is lower than the total stake multiplied by $\alpha(1 - \varepsilon_1)$, a proportional amount of coins is released. Conversely, if the current supply C exceeds the total stake multiplied by $\alpha(1 + \varepsilon_2)$, an appropriate number of Workers is added using the Foundation's funds. ε_1 and ε_2 are relative values, expressed as fractions of the base value α, which define the lower and upper limits of the range respectively. In the first scenario mentioned in the text, the Foundation has the ability to revoke its Workers if they were created using a specific protocol. This means that instead of releasing additional coins into circulation, the Foundation can choose to remove Workers from the system. To ensure that this condition is met and that the system remains in balance, a smart contract is used. This smart contract monitors the balance between the current coin supply and the number of Workers in the system. This balance is adjusted periodically to ensure that it remains within acceptable limits. Determining the optimal values of $\alpha \geq 2$, $\varepsilon_1 > 0$, and $\varepsilon_2 > 0$ to achieve effective and secure network functioning is an unresolved matter that relies on various factors. These factors include the total and current supplies, required available funds, security level, and network operating goals. We consider the initial values of parameters: $\alpha = 2.5$ and $\varepsilon_1 = \varepsilon_2 = 0.25$ as a starting point for the Waterfall case. During the initial stage of the Waterfall platform, it is expected that the foundation will control approximately 2/3 of the total number of tokens in circulation. With the current value of the parameter α, a network capture attack, also known as a majority attack [44], would inflict losses on both the attackers and the attacked parties, making such an attack economically infeasible. As the platform continues to grow and the number of nodes increases, the value of α will decrease accordingly, while maintaining a sufficiently high level to guarantee the platform's security.

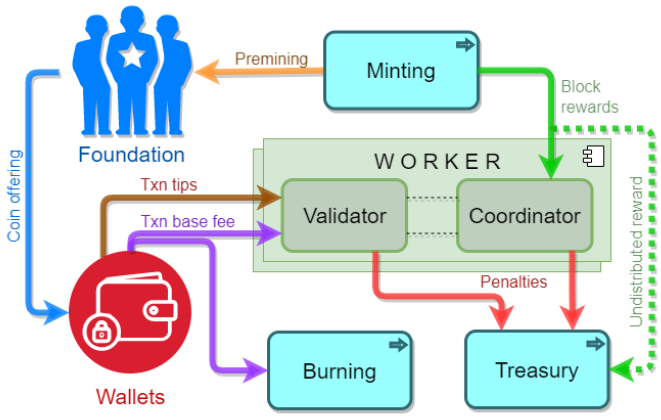


Figure 2: The circulation of coins on the platform

3.2. Understanding the Technical Aspects of Coin Mining and Its Challenges

The proposed tokenomics model incentivizes block production by offering minted rewards for every finalized block in the Coordinating network. This means that new coins will be issued to

compensate Coordinators for ensuring the required security guarantees, which is referred to as "Minimum Necessary Issuance". Earned coins are accrued to Workers every epoch. The annualized minted amount (V) is influenced by the total amount of staked coins (S):

$$V = k \cdot \sqrt{S},$$

where a coefficient k will be subsequently defined. As a result, the maximum annualized return rate (R) is determined as follows:

$$R = V / S = k / \sqrt{S}.$$

This non-linear relationship proposed in [18] implies that a decrease in the amount of staked coins leads to increased incentivization, and vice versa. This balance between the volume of minted coins and network security ensures that all initial stakes, which are initially uniform (although some may be reduced due to penalties), remain proportional, and

$$S = s \cdot N,$$

the coefficient k can be obtained through the desired value of R_{opt} at a certain total stake (S_{opt}) as well as at a certain optimal number of Workers (N_{opt}). In our case, the coefficient k is derived from a condition that the maximum annualized return rate equals R_{opt} with N_{opt} Workers:

$$k = R_{opt} \cdot \sqrt{S_{opt}} = R_{opt} \cdot \sqrt{s \cdot N_{opt}}.$$

Therefore, considering any given number of Workers, we obtain:

$$V = R_{opt} \cdot s \cdot \sqrt{N_{opt} \cdot N},$$

$$R = R_{opt} \cdot \sqrt{\frac{N_{opt}}{N}}.$$

During the initial stages of coin release, the current supply C experiences a rapid surge, necessitating the expansion of the optimal number of Workers to ensure network security. Consequently, the value of N_{opt} should also be substantially increased:

$$N_{opt} = \frac{C}{\alpha \cdot s}.$$

The coefficient k and the annual minted amount V are crucial factors that need to be adjusted to ensure the proper release of all coins. Recalibration is necessary to keep the system efficient and engaging for new Workers by increasing the rewards available for minting. However, if the number of Workers in the network exceeds a certain threshold, the platform may overpay for security, leading to inflation that could have detrimental effects on the tokenomics of the system. To prevent this, the value of α, which represents the percentage of total minted coins allocated for the development fund, should not be too low. It is important to note that the number of coins yet to be released can be approximated by subtracting the current amount in the releasing account (C) from the total coin supply. This approximation provides valuable insights into the state of the system and informs decisions about adjusting the minting process to optimize its performance. The overall rate of return (R), may be reduced due to the occurrence of faults or malicious behavior by certain Coordinators. Despite this, honest Workers who make investments at the outset can expect to receive rewards of approximately R_{opt} per year. This represents the minimum return rate that can be achieved by honest Workers. To illustrate this concept, we present Figure 3, which depicts the maximum annualized return rate that stakeholders can earn as block rewards under varying conditions of the number of Workers and different values of R_{opt} and N_{opt}. With the expansion of the Worker count, the return rate diminishes; however, the rewards could potentially be augmented by tips resulting from heightened network activity.

When evaluating the reward for a produced block in a blockchain network, it is crucial to consider various factors that contribute to the overall value of the reward. Two critical elements to consider are the annualized minted amount and the slot times. The annualized minted amount refers to the total number of tokens that will be minted during the year, assuming the network continues to operate at its current capacity. This value is essential because it determines the overall supply of tokens, and therefore impacts their value in the market. Additionally, the annualized minted amount can affect the incentive structure for network participants, such as block producers, who may receive a portion of the newly minted tokens as a reward. The slot times, on the other hand, refer to the time it takes to produce each block in the network. In most blockchain networks, a new block is produced at regular intervals, which are determined by the slot times. The shorter the slot time, the more blocks can be produced within a given timeframe, and the higher the potential rewards for block producers. To evaluate the reward for a produced block, these two factors must be taken into account. By multiplying the annualized minted amount by the slot times, we can calculate the value of each block produced in the network. This information is essential for understanding the incentives and rewards available to network participants and can inform decisions regarding the network's governance and economic policies. Overall, the evaluation of block rewards is a critical aspect of blockchain network design and operation. By considering various factors such as the annualized minted amount and slot times, we can ensure that the network is functioning efficiently and providing appropriate incentives to all participants. To determine the minted reward for a generated block, one can consider the values of the annualized minted amount and slot times. Let i represent the slot number within the Coordinating network. Then the annualized number of blocks:

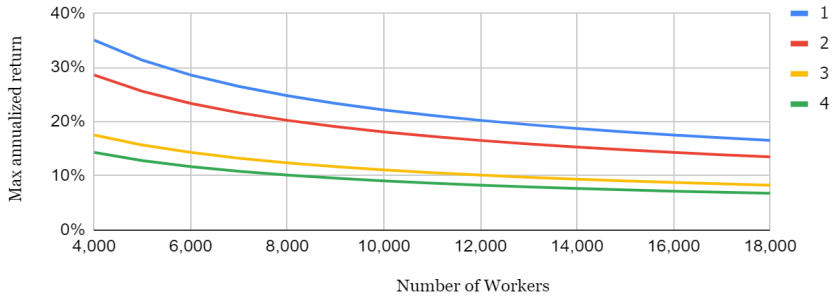


Figure 3: The maximum annual return rate of Workers (case 1 with R_{opt} = 0.20 and N_{opt} = 12,288; case 2 with R_{opt} = 0.20 and N_{opt} = 8,192; case 3 with R_{opt} = 0.10 and N_{opt} = 12,288; case 4 with R_{opt} = 0.10 and N_{opt} = 8,192)

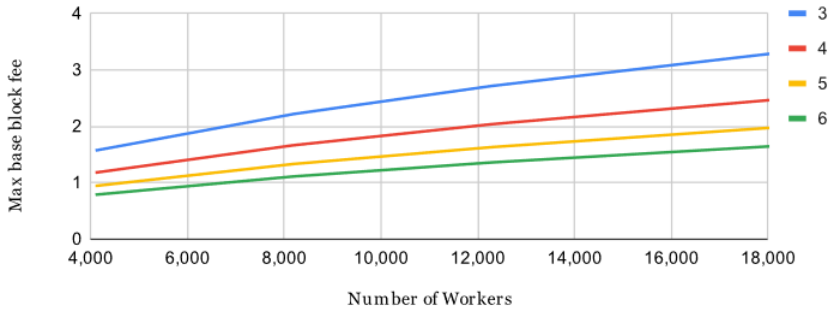


Figure 4: The maximum base block fee (W_i^s) with N_{opt} = 8,192 Workers, R_{opt} = 0.20, s = 32,000 coins, p = 1, t_i^s = 4 sec, and b_i = 3...6 blocks per slot

$$B_i^c = \frac{T}{t_i^c},$$

where t_i^c is its time in seconds, and constant $T = 60 \cdot 60 \cdot 24 \cdot 365.25$ represents the number of seconds per year.

Consequently, the reward that is minted per block in the i-th slot in the Coordinating network is:

$$W_i^c = \frac{V}{B_i^c}.$$

It is important to note that the total sum over all slots of the Coordinating network per year (Y) is equivalent to:

$$\sum_{\forall i \in Y} W_i^c = V \cdot \sum_{\forall i \in Y} \frac{t_i^c}{T} = V.$$

Further, the amount W_i^c is divided among the Committee leader and members who have created i-th block in the Coordinating network.

3.3. Understanding Base Transaction Fees

In the context of DLT, any transaction included in a block requires the payment of a base transaction fee. This fee is determined based on a mechanism similar to that used for obtaining a minted reward and depends on the annualized minted amount. Let i be the slot number of the Sharding network. Then the annualized number of blocks:

$$B_i^s = \frac{T \cdot b_i}{t_i^s}, \quad W_i^s = \frac{V}{B_i^s}.$$

where the number of blocks per slot in the Sharding network $b_i > 1$ and t_i^s is its time in seconds.

It is worth noting that if the block number and time slot remain constant throughout the year, the annual number of blocks B^s can be obtained by a simple calculation. By summing up the values of slots and blocks across the Sharding network for the entire year, we arrive at the following expression:

$$\sum_{\forall i \in Y} \sum_{j=1}^{b_i} W_i^s = V \cdot \sum_{\forall i \in Y} \frac{t_i^s}{T} = V.$$

And then, a base transaction fee in i-th slot is defined as:

$$f = \frac{G}{G_{max}} \cdot W_i^s \cdot p,$$

where G represents the required gas amount to process the transaction, G_{max} denotes the maximum permissible gas amount per block, and $p > 0$ is a price multiplier. The cumulative sum of all base transaction fees across the j-th block of the i-th slot:

$$F_{ij} = \sum_{txns} f \leq W_i^s \cdot p.$$

In the case of a regular transaction, the proportion:

$$\frac{G}{G_{max}} = \frac{1}{10{,}000}$$

ensures quite low base fees, even for quite high values of the multiplier p.

Clearly, the value of W_i^s depends on the current number of Workers as well (see Figure 4).

3.4. Coin Burning Procedure

The act of permanently removing a certain number of coins from circulation to reduce the current supply is commonly referred to as "coin burning." This process is often carried out by sending tokens to an account that can receive them but cannot withdraw them, rendering the coins useless and irretrievable. In the context of Ethereum, the Ethereum Improvement Proposal EIP-1559 has introduced a mechanism whereby base transaction fees are burned [45], effectively improving the tokenomics of the platform. However, Validators are permitted to retain the tips from transactions, providing an additional incentive for their participation. An alternative approach to coin burning involves deducting a fixed percentage of each transaction fee and sending it to an inaccessible account, thereby reducing the total supply of tokens over time. This method has been adopted by several blockchain platforms as a means of managing their token supply and improving their tokenomics [46]. Our economic model follows a similar approach, wherein the base transaction fee is divided into two portions using a burning multiplier denoted as $l \in [0; 1]$:

$$f = l \cdot f + (1 - l) \cdot f.$$

The first component is burned but the second one is left for a Validator. It is clear that l not equal to 1 increases the total Workers' rewards and reduces the burned coin amount. The value of l can be the same for all blocks and changed only by the network voting, or it depends on a particular block or the reputation of a block producer to incentivize it.

One of the possible approaches is discussed in [47]:

$$l = l_0 + (1 - l_0)(1 - q(b))$$

for each block b, where $q(b) \in [0; 1]$ is a measure of its referential structure "quality" (here 1 means the best quality), and the parameter l_0 not less than 0 signifies the minimum proportion of the transaction fees that are subjected to burning. To make this happen, the DAG topology created by honest Validators was examined and their typical behavior was described in [43]. Therefore, the function *q(b)* could be built on the basis of comparing the existing referential structure of block *b* with its theoretical reference (ideal) version. Incentivizing the encouragement of the establishment of a well-structured system of references can be highly beneficial in preserving the accuracy and safeguarding the confidentiality of data within the Sharding network. This practice serves as a means of incentivization, motivating users to adopt and adhere to a standardized method of referencing information, which in turn promotes better organization and security throughout the network.

Furthermore, in this system, all penalties incurred are effectively removed from circulation. A Worker who holds less than 50% of his/her initial stake forfeits the privilege to validate and produce blocks in the future, irrespective of any rewards he/she may have received. Any violation of the system's rules or any misbehavior on the part of a Coordinator or a Validator attracts an automatic penalty, which includes the following scenarios: (1) a Coordinator, acting as a committee member, deliberately omits a series of votes; (2) a Coordinator, acting as a committee member, issues conflicting messages or double votes; (3) a Coordinator, acting as a block producer, fails to create a block in the Coordinating network; (4) a Coordinator, acting as a block producer, generates more than 1 block within the same slot of the Coordinating network; (5) a Validator produces more than 1 block in the same slot of the Sharding network, and these blocks are subsequently confirmed in the Coordinating network; (6) a Coordinator submits an invalid proof of any of the aforementioned offenses. Penalties accumulate; for example, if three blocks are produced instead of one, the penalty doubles.

All Coordinators collectively make decisions based solely on the information provided by the coordinating ledger, without requiring any additional network consensus. If a whistleblower identifies an offense, they record proof of it in a block when it is their turn to produce one. Therefore, there is no need for an extra consensus mechanism. For instance, in case 4, a whistleblower, who detects two blocks generated in the same slot receives 50% of the penalty amount as a reward. In case 6, another Coordinator can report malicious activity by providing a reference to such a block, doubling the penalty imposed on such a leader in comparison to their potential benefits.

3.5. Treasury

In the blockchain ecosystem, a treasury refers to a reserved pool of digital assets that serve a specific purpose (e.g. [48], [49], [50]). This allows stakeholders to make decisions, usually through a decentralized governance process, on how to allocate funds. The blockchain treasury can be used for a wide range of purposes, e.g. to finance the development of new features or improvements in network protocols, to cover the costs associated with operational expenses, to fund marketing initiatives aimed at increasing awareness and mass adoption of decentralized services, to acquire other projects, for charitable activities, etc. Replenishment of the treasury can be carried out through various means. The most common methods include coin issuance and the accrual of a portion of transaction fees. Additionally, penalties and donations in the form of cryptocurrency can also be used to fund treasuries.

In the Waterfall platform, if committee members fail to vote in the *i*-th slot, they will not receive their share of the minted reward $W_i^{c,}$ and a slot leader will not receive the reward for including such missed votes in the block. Any undistributed funds are accumulated every epoch in keeping with the overall system design and can be used to replenish the treasury. In addition, a whistleblower that reports misconduct is rewarded with 50% of the penalty amount, with the other half being transferred to the treasury. If a Validator is unable to synchronize before producing their block and relies on outdated blocks, their reward may be diminished, and the resulting losses are also added to the treasury account.

The main advantage of this approach is that financing the treasury at the expense of undistributed funds and penalties does not create any additional inflationary pressure on the system. Under the consensus protocol assumptions [17], the percentage of faulty Workers does not exceed one-third of their total number. Therefore, the number of funds transferred to the treasury cannot exceed one-third of all minted coins, since at least two-thirds of Workers are well-behaved and fully rewarded. This design ensures a balanced and stable system, promoting fairness and transparency for all participants.

4. Examining Attacks on Tokenomics

Waterfall is a public decentralized and open-source peer-to-peer platform, which means that any individual or group can participate in the network, regardless of their motivations or intentions. Unfortunately, this open nature also makes the network vulnerable to bad actors who may choose to deviate from network protocols or even commit malicious acts against other platform participants.

To better understand these threats, we must first define what we mean by an "attack". In this context, an attack refers to any set of actions that results in a violation of the correct operation of the entire system or its components, leading to a deterioration in technical and economic indicators. However, not all attacks are created equally. Some may cause minor disruptions or receive insignificant benefits for a particular network participant, while others can result in irreversible violations and a complete disruption of the platform. In the case of Waterfall, we are particularly interested in attacks that are aimed at the economy of the platform and/or carried out with the help of economic leverage. These types of attacks leverage the financial incentives and structures of the platform to achieve malicious goals. They can take many forms, including, but not limited to, double-spending, Sybil, majority attacks, and other types of manipulation or exploitation.

Given the importance of economic incentives and structures in the functioning of tokenomics networks like Waterfall, it is very important to categorize these attacks based on the amount of damage or potential damage they can cause. Doing so can help us better understand the risks associated with different types of attacks and inform strategies for mitigating those risks. Ultimately, the goal is for the platform to remain secure, stable, and reliable for all involved.

Our next step is to classify attacks according to the extent of damage or potential harm they can inflict. This ranges from minor gains made by certain network participants at the expense of others, to severe violations that may lead to the complete disruption of the platform and irreversible consequences.

4.1. Typology of Node and User Behavior

In the context of the platform, users can be classified into three distinct groups based on their behavioral patterns. The first group comprises users who always abide by the network protocols without exception. The second group consists of users who seek opportunities to increase their income or decrease their expenses by flouting the rules, such as by altering the software. The third group is composed of users who engage in activities that harm the network as a whole or the other users, ultimately leading to financial losses for those involved.

Both the second and third groups of users can act either individually or collectively, forming homogeneous or mixed subgroups. For instance, users from the third group might bribe members of the second group to carry out an attack. While users who inadvertently violate the protocols due to hardware or software failures do not intentionally collude, they may still act synchronously for various reasons. Such users will also be classified under the third group, as they are not seeking to gain any benefit by violating the rules.

4.2. Reputation System

One promising indicator of reliable Worker performance is the number of earned tokens, which can be easily tracked using data from the blockchain ledger. This approach avoids the need for complex calculations that could overburden the network. Notably, we focus only on a "positive" reputation system, as a negative reputation system could be circumvented by a misbehaving Worker who could create a new account and start afresh (known as "zeroing").

It is worth highlighting that, in our scenario, each Worker starts with an equal stake. Consequently, those Workers who are more productive, particularly in block production, earn a higher income. However, the current approach can lead to an advantage for those Workers who joined the network earlier. Moreover, a Worker's productivity and efficiency may fluctuate over time. To address these concerns and ensure that node reputations are continuously updated, a depreciation mechanism can be implemented, such as a weighted moving average of rewards over several eras. This approach considers only tokens earned within a recent timeframe, ensuring that the most up-to-date information is used to determine a Worker's reputation.

The reputation system we propose can have broader applications beyond tracking Worker performance. For example, it could be utilized in peer-to-peer node interactions to prioritize communication with established nodes with a strong reputation. Moreover, we envision offering additional incentives and benefits to Workers with exceptional reputations as a means of motivating and rewarding their high level of performance.

4.3. Principles for Countering Economic Attacks

The principles for countering economic attacks on a network must be carefully considered to ensure maximum economic benefit to all users who comply with network rules. A unique Nash equilibrium [51] should be established in which it is most beneficial for all users to follow network protocols honestly and refrain from attacking the system. To achieve this, a thorough analysis of the system's functioning must be undertaken, taking into account the best rational behavior of malicious users. The behavior of users in the second group, who seek to increase their income by breaking the rules, must be rendered ineffectual in the absence of external funding for attacks.

It is also essential to estimate the cost of possible attacks by representatives of the third group, which actively disrupt the network to the detriment of individual users' economic interests. Network protocols should be designed to offset any damages caused by network disruptions with the necessary amount of funds. It should be noted that it is impossible to prevent attacks entirely in the operation of any public platform. The very concept of PoS assumes that by investing a certain amount of funds, a group of people gains control over a decentralized platform (a majority attack) [45].

To mitigate the impact of attacks on the network, protocols should be established to detect and promptly respond to any malicious activity. A mechanism for punishing attackers should also be implemented, which could involve the confiscation of tokens or other economic sanctions. Additionally, the use of staking mechanisms could provide incentives for network participants to act honestly, as they would possess a personal interest in the prosperity of the network. In summary, it is essential to build a network that is resistant to economic attacks and promotes honest behavior by all participants.

4.4. Economic Attack Vector

The Waterfall consensus protocol [17] plays a crucial role in countering attacks, including economic attacks. A consensus protocol has to have the following two features [52]:

- **Liveness:** once a valid block becomes available to network nodes, it will be appended to the distributed ledger, and all valid transactions will be accepted after a certain period of time;
- **Safety:** the consensus decision is consistent among all honest nodes.

In other words, the consensus protocol coordinates the current state of the distributed ledger to provide the necessary economic information to all network participants. Then, the directions of typical attacks and possible prevention mechanisms within the framework of high-level design are considered.

Double-spending is an attack where a certain number of coins is used multiple times [53]. A situation may arise where a transaction for spending the same amount is published in different blocks of the DAG network. To enhance security, a new PoS-oriented blockDAG linearization algorithm was presented in [17]. Here, security means that if an honest Worker proposes an order of a blockDAG part in its turn, all other honest Workers will also

approve the same block order to be further finalized. Therefore, after running the linearization and finalization procedure, only the first transaction is considered correct, and later conflicting transactions are ignored.

A common argument against PoS is that it favors the rich over time and reduces the rewards for those who start with fewer coins, the so-called **Rich-Get-Richer** problem [54]. To a greater extent, this is inherent in systems with a high financial entry threshold. Rich stakeholders receiving more income have the opportunity to increase their presence in the network by creating new nodes.

On the Waterfall platform, the number of Workers is given preference, setting a low financial entry threshold. Therefore, having earned even a relatively small amount, a new Worker can be created, thereby increasing future earnings. Additionally, there is an opportunity to generate income without staking by providing processing power for the operational platform. One can create a network node and, e.g., for a monthly fee, deploy Workers at the request of other users.

Nothing-at-stake is a security issue in PoS-based systems where nodes have no financial risks to support or generate mutually exclusive proposals (e.g. voting for different blockchain forks, producing a certain amount of blocks per slot) to gain extra benefits [55]. Such malicious behavior can disrupt the consensus, reducing system performance.

With Waterfall, as with most other public platforms, there are penalties for the misbehavior of Workers, and simultaneous validation/production of conflicting blocks/messages leads to the loss of locked coins [43]. Based on data recorded in the ledger, these penalties are levied automatically. Since data in the ledger is always the same and available, all honest Workers can make judgments based exclusively on that information without overloading the network or reducing its performance.

Maximal Extractable Value (MEV) attacks can occur on cryptocurrency platforms when a block producer is able to incorporate its own transactions and reorder users' transactions in the block to maximize its profit.

In traditional financial systems, this type of attack is only available to a narrow circle of people who have privileged/priority access to confidential information of financial institutions, since any algorithmically predictable behavior in finance is a vulnerability that can be exploited, for example, to conduct arbitrage [56]. However, in public platforms, such information is available to all users, primarily due to the openness and variety of smart contracts that implement financial instruments. Thus, there are ample opportunities for block producers to implement MEV. Strictly speaking, these attacks are not carried out directly on decentralized platforms but are seen in DeFi services deployed on their basis [9].

By definition, MEV provides a way to counteract this type of threat directly, eliminating or significantly reducing the possibility of block producers' influence on the order of transactions written to the ledger. Here, DAG-based protocols have certain advantages over blockchain-based systems [57]. In the future, we plan to implement additional MEV-resistant features into our consensus protocol.

5. Modeling Tokenomics

5.1. Inflation and Deflation

In traditional economics, the monitoring of currency issuances enables transparency in overseeing different financial aspects. Within tokenomics, an expansion and contraction in the circulation of coin supply are referred to as inflation and deflation, respectively. The distinction between minted and burned coin quantities is a significant network economic attribute, which can automatically be calculated in real-time. Therefore, the predetermined algorithm of the coin issuance should be carefully examined and simulated, considering specific transaction workloads that impact burned coin quantities as base transaction fees. We have examined several scenarios to gain a more comprehensive understanding of the economic dynamics within the Waterfall platform. Therefore, we were able to optimize and protect it.

Consider a value represented by the occupancy of *j*-th block of *i*-th slot in the Sharding network (see subsection 3.3):

$$r_{ij} = \frac{F_{ij}}{W_i^s \cdot p} \in (0; 1].$$

If $r_{ij} = r$ remains constant over a year, then the amount of burned coins (U) over a year can be calculated as follows:

$$U = \sum_{\forall i \in Y} \sum_{j=1}^{b_i} F_{ij} l \; = \sum_{\forall i \in Y} \sum_{j=1}^{b_i} r_{ij} W_i^s p \; l = rVpl.$$

If the slot time and the number of blocks are constant throughout the year, then the number of burned coins can be calculated using the following formula:

$$U = \sum_{\forall i \in Y} \sum_{j=1}^{b_i} r_{ij} \frac{V}{B^s} pl = r_0 Vpl,$$

where r_0 is the average value of r_{ij} across all blocks per year. The annualized burned amount U with a price multiplier $p = 1$ never surpasses the emitted amount V anyway, since $r_{ij} \leq 1$. It is possible to modify the value of p through network voting in the future.

The calculation of the annual inflation rate involves determining the ratio between the difference between the total coin supply and the burned coins ($V - U$) and the current coin supply, which is approximately equal to the constant α multiplied by the total coin supply $\alpha \cdot S$. Figure 5 illustrates several potential scenarios with varying degrees of block occupancy and number of Workers, all of which are based on the assumption of a constant $\alpha = 2.5$, a price multiplier $p = 1$, and a burning multiplier $l = 1$.

As more transactions take place, more coins are burned as transaction fees, which can help offset the inflationary effects resulting from the creation of new coins. Obviously, there is no inflation when $r_0 \cdot p \cdot l = 1$, and the deflation process can be observed when $r_0 \cdot p \cdot l > 1$. Thus, the parameters p and l can be used to select the scenario of the economic dynamics. For example, when $l = 1$ and $p = 2$, at the level of 50% average block occupancy, the inflation rate equals zero (Figure 6). However, with a higher block occupancy corresponding to more transactions per second), a deflationary trend can occur.

The economic model can exhibit both inflationary and deflationary tendencies, as the number of coins burned in the Sharding network depends on its workload, and is closely tied to the number of coins minted in the Coordinating network. To identify the optimal configuration of parameters for a given set of

Figure 5: The relationship of the upper limit of the inflation rate per year from the block occupancy r_0 with $p = 1$ and $l = 1$

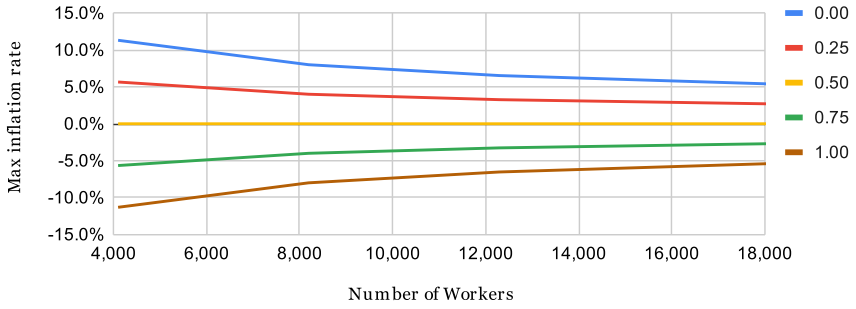


Figure 6: The relationship of the upper limit of the inflation rate per year from the block occupancy r_0 with $p = 2$ and $l = 1$

network specifications, it is possible to leverage both structural and mathematical modeling techniques, with a defined objective function tailored to the specific strategy pursued in platform development. By carefully fine-tuning the values of key parameters, it may be possible to achieve a balanced economic system that supports sustainable growth and stability over time. Ultimately, the ability to achieve the desired economic outcomes will depend on a range of factors, including the prevailing market conditions, the actions of network participants, and the overall level of demand for the platform's services.

5.2. Faulty Work Simulating

In [43], the theoretical underpinnings of the Waterfall incentive system were elaborated in detail, including the distribution of rewards and the imposition of penalties, all designed to achieve a state of general economic equilibrium. The key principle governing the calculation of penalties is that they should be significantly higher than any potential gains that could be realized through a potential attack on the system. To achieve this, a multiplier of $\lambda \geq 1$ is used to amplify the scale of the penalties. One of the core objectives of the simulation was to determine an optimal value λ that would be suitable for different scenarios of Worker misconduct. The overarching goal of this process is to ensure that faulty or idle Workers are eventually excluded from the consensus, thus preventing the accumulation of a critical mass of defective nodes that could hinder consensus. At the same time, temporary equipment failures or inadvertent shutdowns of Workers should not result in a permanent ban from the system.

The following scenarios for Worker failures were considered:

1. h_{online} hours working reliably and $h_{offline}$ hours off;
2. the shutdown duration distributed according to the normal law with mean t_{mean} and standard deviation t_{std};
3. the probability of failure when performing one or another action is $P \in (0; 1)$.

During the experiments, a fixed time interval was used, and different values were assigned to the various parameters mentioned above. After this interval, Workers were removed from participation in the work, and the topology of the resulting DAG was analyzed. Based on the findings, it was determined that a relatively high value of $\lambda = 100$ should be set for penalty rates. For instance, failing to produce two consecutive blocks or submitting conflicting messages during the voting process incurs a penalty 100 times greater than the corresponding reward. In addition to the penalty, the Worker's participation in the network is temporarily suspended for the current and next eras so that the necessary equipment settings can be made. This helps to preserve the value of the Worker's stake, as subsequent failures would cause the stake to fall below the 50% threshold, leading to permanent exclusion from the network. While the Worker is suspended, they do not take part in committee work or block production, which in turn reduces the percentage of faulty nodes in the network.

6. Implementation and Experimental Study of the Waterfall System

The development of the Waterfall system required a comprehensive approach, starting with the experimental study of its functioning. Initially, we focused on a small but increasing number of nodes, where we studied the system's behavior using simulation tools. As we gained more insights, we moved on to conduct full-scale tests in a test network with a smaller number of nodes.

Currently, we have implemented the main initial elements of our protocol and are conducting load experiments using t3.small and t3.medium AWS Servers with a 2-core CPU and 2 or 4 GB RAM respectively [58]. These tests are being conducted to evaluate the tokenomics model's ability to serve the Waterfall

Table 1: Network load test results

Number of				Loading							
				Idling		1st day			3rd day		
Servers	Nodes	Workers	Total	CPU	RAM	CPU	RAM	Spending	CPU	RAM	Spending
t3.medium	32	32	1024	1.91	0.59	2.51	1.3		2.2	1.8	
	32	16	512	2.32	0.69	2.32	1.3	3.44 USD	0.56	2.2	1.25 USD
	8	32	256	1.91	0.59	2.51	1.3				
	8	16	128	2.32	0.71	2.32	1.3		0.53	2.66	2.33 USD
	8	8	64	2.32	0.51	2.32	1.3				
t3.small	8	8	64	1.91	0.51	2.21	1.3				

platform with high transaction throughput. We are optimistic about the system's performance since it has demonstrated the desired behavior.

However, to ensure optimal performance, we continuously check the values of the system parameters and study further optimization opportunities. Our approach has been focused on caution and thoroughness to ensure that the Waterfall system is developed to the highest standards and is capable of serving its intended purpose.

The outcome of the test that measured the amount of work a server can handle (refer to Table 1) indicated that there was no meaningful connection between the CPU or RAM load and the number of Workers running on a single server. Concurrently, there is a shift in the load structure over time. At the same time, there is a shift in the load arrangement over time. By the third day of operation, the influence of software adaptation mechanisms was observable. A part of the data necessary for work was cashed in shared memory. As a consequence, the processor's burden was reduced by 75-77%. Only when the amount of available memory allowed for saving the necessary information for all the Workers running on the node was such a significant reduction in CPU load achieved. Also, an experiment was conducted to evaluate hashing of data with insufficient RAM for 32 Workers. Despite the 38% increase in memory consumption, only 12% savings in CPU resources were observed. Observations show that doubling the minimum required amount of RAM results in more than four times less load on the processor on the third day of operation. This, combined with automatic optimization of communications and the corresponding reduction in traffic, leads to an average 64% reduction in server maintenance costs at Amazon Web Services prices since the third day.

The table demonstrates that entry to the Waterfall network is more accessible compared to other popular networks of a similar nature. A lower entry threshold is made possible by less demanding software and a lower computational load, with everything else being equal.

One can notice that the test network with fewer nodes requires more costs per server. This is due to the overall cost of maintaining the TestNet infrastructure. When distributing these costs over 32 nodes, we naturally get a smaller amount compared to 8 nodes. In this regard, we make a comparative analysis of expenses only within the rows of the table, but not between different rows.

The cost of user-nodes with a transfer rate of 2 standard transactions per second amounted to 28.7 and 35.2 coins per day, on average, where the optimal number of Workers N_{opt} was equal to 8,192 and 12,288 respectively $R_{opt} = 0.2$.

The implementation of the Waterfall System provides for the possibility of deploying several autonomous Workers on each node with a common ledger and a pool of transactions. Having as many Workers as possible in one node provides an economic benefit, but it negatively affects network decentralization. Their number must have an upper boundary, taking into account technical limitations. A large number of Workers deployed on a single device may not have adequate time to create and process blocks. This leads to a shortfall in rewards and the imposition of penalties.

7. Conclusion

The multishard DAG-based platform Waterfall is a sophisticated yet efficient amalgamation of diverse pre-existing and unique solutions. Tokenomics serves as the tool to maintain coherence among the distinct components and optimize their operations. Although the presented tokenomics framework is tailored for the Waterfall platform, the methods, and approaches outlined in the document could be valuable for creating economic support systems for other decentralized networks.

The conclusions drawn from the article suggest that the general model designed for Waterfall tokenomics can bring about an economic balance that caters to the concerns of all participants within the platform. This equilibrium is achieved while keeping transaction fees at a low level, which is essential for implementing DLTs and smart contracts into applications designed for enterprise-level use. The advantages of this model become apparent when dealing with a high volume of transactions, making the platform more reliable and promoting digital transformation through transparency and trust.

By offering an affordable transaction fee, the Waterfall platform can be utilized for a broad spectrum of services, such as DeFi, IoT, Web3 gaming, digital identity, medical screening systems, and peer-to-peer energy trading, etc. The foundational principles of Waterfall tokenomics align with its DAG-based architecture design, and dynamically adapting macroeconomic mechanisms ensure that the platform performs optimally and sustainably in rapidly changing situations.

Both inflationary and deflationary tendencies can be observed in the economic model depending on initial parameters and further network management. A few possible scenarios of the system evaluation were mathematically simulated and experimentally studied in the testnet.

Future work on Waterfall tokenomics will focus on researching the economic aspects of interactions between various homomorphic Sharding networks, forming transaction fees based on the external exchange rate, voting for economic parameters, and

developing the mechanics required for implementing different incentivization strategies. Additionally, the economic issues of on/off-boarding process will also be explored, and simulations will be conducted to evaluate the effectiveness of various strategies. Overall, the article highlights the potential benefits of Waterfall tokenomics in promoting decentralized technologies and creating a more transparent and trustworthy digital environment.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] S. Grybniak, Y. Leonchyk, R. Masalskyi, I. Mazurok, O. Nashyvan, "Waterfall: Salto Collazo. Tokenomics," 2022 IEEE International Conference on Blockchain, Smart Healthcare and Emerging Technologies (SmartBlock4Health), Bucharest, Romania, 2022, 2022, doi: 10.1109/SmartBlock4Health56071.2022.10034521

[2] M. N. M. Bhutta, A. A. Khwaja, A. Nadeem, H. F. Ahmad, M. K. Khan, M. A. Hanif, M. Alshamari, Y. Cao, "A survey on blockchain technology: evolution, architecture and security," IEEE Access, **9**, 61048-61073, 2021, doi: 10.3390/s22145274.

[3] A. T. Sherman, F. Javani, H. Zhang, E. Golaszewski, "On the origins and variations of blockchain technologies," IEEE Security & Privacy, **17**(1), 72-77, 2019, doi: 10.1109/MSEC.2019.2893730.

[4] S. Grybniak, Y. Leonchyk, R. Masalskyi, Igor Mazurok, Oleksandr Nashyvan, Ruslan Shanin "Decentralized platforms: goals, challenges, and solutions," 2022 IEEE 7th Forum on Research and Technologies for Society and Industry Innovation (RTSI), 62-67, doi: 10.1109/RTSI55261.2022.9905225.

[5] H. Pervez, M. Muneeb, M. U. Irfan, I. U. Haq, "A comparative analysis of DAG-based blockchain architectures," 12th International Conference on Open Source Systems and Technologies (ICOSST), 27-34, 2018, doi: 10.1109/ICOSST.2018.8632193.

[6] F. M. Benčić, I. P. Žarko, "Distributed ledger technology: blockchain compared to directed acyclic graph," IEEE 38th International Conference on Distributed Computing Systems, 1569-1570, 2018, doi: 10.1016/j.technovation.2023.102711.

[7] Q. Wang, J. Yu, S. Chen, Y. Xiang, "SoK: diving into DAG-based blockchain systems," preprint arXiv:2012.06128, 2020, doi: 10.48550/arXiv.2012.06128.

[8] R. Selkis, A Messari report: Crypto theses for 2023, 2023, https://messari.io/crypto-theses-for-2023.

[9] S. M. Werner, D. Perez, L. Gudgeon, A. Klages-Mundt, D. Harz, W. J. Knottenbelt, "Sok: decentralized finance (DeFi)," preprint arXiv:2101.08778, 2021.

[10] L. Dam, What is Web3 Gaming? How is it different from Traditional Gaming? 2023, https://ekoios.vn/web3-games-and-traditional-games-comparison.

[11] E. Tijan, S. Aksentijević, K. Ivanić, M. Jardas, "Blockchain technology implementation in logistics," Sustainability, **11(4)**, 1185, 2019, doi: 10.3390/su11041185.

[12] R. M. Garcia-Teruel, "Legal challenges and opportunities of blockchain technology in the real estate sector," Journal of Property, Planning and Environmental Law, 2020, doi: 10.1108/JPPEL-07-2019-0039.

[13] J. Lee, "BIDaaS: Blockchain based ID as a service," IEEE Access, **6**, 2274-2278, 2018, doi: 10.1109/ACCESS.2017.2782733.

[14] N. Kshetri, J. Voas, "Blockchain-enabled e-voting," IEEE Software, **35**(4), 95-99, 2018, doi: 10.1109/MS.2018.2801546.

[15] S. Grybniak, D. Dmytryshyn, Y. Leonchyk, I. Mazurok, O. Nashyvan, R. Shanin, "Waterfall: a scalable distributed ledger technology," in 2022 IEEE 1st Global Emerging Technology Blockchain Forum: Blockchain & Beyond (iGETblockchain), 1-6, 2022, doi: 10.1109/iGETblockchain56591.2022.10087112.

[16] C. T. Nguyen, D. Thai Hoang, D. N. Nguyen, D. Niyato, H. T. Nguye, E. Dutkiewicz, "Proof-of-stake consensus mechanisms for future blockchain networks: fundamentals, applications and opportunities," IEEE Access, **7**, 85727-85745, 2019, doi: 10.1109/ACCESS.2019.2925010.

[17] S. Grybniak, D. Dmytryshyn, Y. Leonchyk, I. Mazurok, O. Nashyvan, R. Shanin, "Waterfall: Gozalandia. Distributed protocol with fast finality and proven safety and liveness," IET Blockchain, 1–12, 2023, doi: 10.1049/blc2.12023.

[18] K. Lau, Ethereum 2.0. An introduction, Crypto.com, 25 p, 2020.

[19] S. Au, T. Power, Tokenomics: the crypto shift of blockchains, ICOs, and tokens, Packt Publishing Ltd, 2018.

[20] L. W. Cong, Y. Li, N. Wang, "Tokenomics: dynamic adoption and valuation," The Review of Financial Studies, **34**(3), 1105-1155, 2021, doi: 10.1093/rfs/hhaa089.

[21] G. A. Pierro, H. Rocha, "The influence factors on Ethereum transaction fees," IEEE/ACM 2nd International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB), 24-31, 2019, doi: 10.1016/j.bcra.2022.100074.

[22] E. Anceaume, A. D. Pozzo, T. Rieutord, S. Tucci-Piergiovanni, "On finality in blockchains," preprint arXiv:2012.10172, 2020, doi: 10.4230/LIPIcs.OPODIS.2021.6.

[23] P. Freni, E. Ferro, R. Moncada, "Tokenomics and blockchain tokens: a design-oriented morphological framework," Blockchain: Research and Applications, **3**(1), 2022, doi: 10.1016/j.bcra.2022.100069.

[24] O. Letychevskyi, "Creation of a self-sustaining token economy," The Journal of The British Blockchain Association, **5**(1), 2022, doi: 10.31585/jbba-5-1-(4)2022.

[25] J. A. Kroll, I. C. Davey, E. W. Felten, "The economics of Bitcoin mining or, Bitcoin in the presence of adversaries," 12th Workshop on the Economics of Information Security, 2013, doi: 10.36484/liberal.662625.

[26] S. Davidson, P. de Filippi, J. Potts, "Economics of blockchain," Public Choice Conference, 2016, doi: 10.2139/ssrn.2744751.

[27] L. W. Cong, Y. Li, N. Wang, "Blockchain and tokenomics," COJ Reviews & Research, **1**(1), 2018, doi: 10.1093/rfs/hhz007.

[28] A. Yakovenko, Solana: a new architecture for a high performance blockchain v0.8.13, Whitepaper, 2018, https://solana.com/solana-whitepaper.pdf.

[29] W. Silvano, R. Marcelino, "Iota Tangle: a cryptocurrency to communicate Internet-of-Things data," Future Generation Computer Systems, **112**, 307-319, 2020, doi: 10.1016/j.future.2020.05.047.

[30] J. J Reuben, A. Joshua, "Blinkchain - a regulation friendly proof-of-speed blockchain v0.1," available at SSRN 4267038, 2022.

[31] S. Bano, A. Sonnino, M. Al-Bassam, S. Azouvi, P. McCorry, S. Meiklejohn, G. Danezis, "SoK: consensus in the age of blockchains," 1st ACM Conference on Advances in Financial Technologies, 183-198, 2019, https://arxiv.org/pdf/1711.03936.pdf.

[32] Z. Liu, N. C. Luong, W. Wang, D. Niyato, P. Wang, Y. Liang, D. I. Kim, "A survey on applications of game theory in blockchain," preprint arXiv:1902.10865, 2019, doi: 10.13868/j.cnki.jcr.000287.

[33] K. Iyer, C. Dannen, "Crypto-economics and game theory. Building Games with Ethereum Smart Contracts," Apress Berkeley, 129–141, 2018, doi: 10.1007/978-1-4842-3492-1_6.

[34] Z. Chang, W. Guo, X. Guo, Z. Zhou, T. Ristaniemi, "Incentive mechanism for edge-computing-based blockchain," IEEE Transactions on Industrial Informatics, **16** (11), 7105–7114, 2020, doi: 10.1109/TII.2022.3163550.

[35] S. Motepalli, H. Jacobsen, "Reward mechanism for blockchains using evolutionary game theory," IEEE 3rd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS), 217-224, 2021, doi: 10.1109/BRAINS52497.2021.9569791.

[36] A. Salau, R. Dantu, K. Morozov, S. Badruddoja, K. Upadhyay, "Making blockchain validators honest," 2022 Fourth International Conference on Blockchain Computing and Applications (BCCA), 267-273, IEEE, 2022, September, doi: 10.1109/BCCA55292.2022.9921952.

[37] I. Mazurok, V. Pienko, Y. Leonchyk, "Empowering fault-tolerant consensus algorithm by economic leverages," ICTERI Workshops, 465-472, 2019, doi: 10.1049/blc2.12023.

[38] Y. Amoussou-Guenou, A. Pozzo, M. Potop-Butucaru, S. Tucci-Piergiovanni, "Correctness and fairness of tendermint-core blockchains", preprint arXiv:1805.08429, 2018, doi: 10.4230/LIPIcs.OPODIS.2018.16.

[39] P. Chafe, A. Mashatan, A. Munro, B. Goncalves, D. Cameron, J. Xu, Dandelion network whitepaper, 22 p, 2022, https://dandelionnet.io/wp-content/uploads/2022/09/dandelion_whitepaper.pdf.

[40] O. Antonenko, S. Grybniak, D. Guzey, O. Nashyvan, R. Shanin, "Subnetworks in BlockDAG," 2022 IEEE 1st Global Emerging Technology Blockchain Forum: Blockchain & Beyond (iGETblockchain), 1-6, doi: 10.1109/iGETblockchain56591.2022.10087101.

[41] R. Dennis, G. Owen, "Rep on the block: a next generation reputation system based on the blockchain," IEEE 10th International Conference for Internet Technology and Secured Transactions (ICITST), 131-138, 2015, doi: 0.1109/ICITST.2015.7412073.

[42] Z. Zhou, M. Wang, C. Yang, Z. Fu, X. Sun, Q. Wu, "Blockchain-based decentralized reputation system in E-commerce environment," Future Generation Computer Systems, **124**, 155-167, 2021, doi: 10.1016/j.future.2021.05.035.

[43] I. Mazurok, Y. Leonchyk, S. Grybniak, O. Nashyvan, R. Masalskyi, "An incentive system for decentralized DAG-based platforms," Applied Aspects of Information Technology, **5**(3), 196–207, 2022, doi: 10.15276/aait.05.2022.13.
[44] W. Li, S. Andreina, J. Bohli, G. Karame, "Securing proof-of-stake blockchain protocols," Data privacy management, cryptocurrencies and blockchain technology, 297-315, Springer, Cham, 2017, doi: 10.1007/978-3-319-67816-0_17.
[45] T. Roughgarden, Transaction fee mechanism design for the Ethereum blockchain: an economic analysis of EIP-1559, 2020, doi: https://timroughgarden.org/papers/eip1559.pdf.
[46] Solana Foundation, Transaction fees, 2022, https://docs.solana.com/ru/transaction_fees.
[47] R. Masalskyi, "DAG Distributed Ledger Modeling," The 1st Student Sci. Conf. of Joint Res, Cooperation between Odesa I.I. Mechnikov National University and Huaiyin Institute of Technology, 171–175, 2022.
[48] Polkadot, Understanding the Polkadot treasury, Medium, 2021, https://polkassembly.medium.com/understanding-the-polkadot-treasury-816821ffc589.
[49] Cardano, Monetary policy, Cardano docs, https://docs.cardano.org/explore-cardano/monetary-policy.
[50] NEAR, Ecosystem treasury DAO, 2021, https://gov.near.org/t/near-ecosystem-treasury-dao/2946
[51] J. F. Nash Jr., "Equilibrium points in n-person games," Proceedings of the national academy of sciences, **36**(1), 48-49, 1950, doi: 10.1515/9781400884087-007.
[52] L. Lamport, "Proving the correctness of multiprocess programs," IEEE Transactions on Software Engineering, SE-**3**(2), 125–143, 1977, doi: 10.1109/TSE.1977.229904.
[53] A. Begum, A. Tareq, M. Sultana, M. Sohel, T. Rahman, A. Sarower, "Blockchain attacks analysis and a model to solve double spending attack," International Journal of Machine Learning and Computing, **10**(2), 352-357, 2020, doi: 10.18178/ijmlc.2020.10.2.942.
[54] Y. Huang, J. Tang, Q. Cong, A. Lim, J. Xu, "Do the rich get richer? Fairness analysis for blockchain incentives," International Conference on Management of Data, 790-803, 2021, doi: 10.1145/3448016.3457285.
[55] D. Rose, E. Machery, S. Stich, M. Alai, A. Angelucci, R. Berniūnas, E. Buchtel, A. Chatterjee, H. Cheon, I. Cho et al., "Nothing at stake in knowledge," Noûs, 224–247, 2019, doi: 10.1111/nous.12211.
[56] K. Qin, L. Zhou, A. Gervais, "Quantifying blockchain extractable value: How dark is the forest?" 2022 IEEE Symposium on Security and Privacy (SP), 198-214, 2022, May, doi: 10.1109/SP46214.2022.9833734.
[57] D. Malkhi, P. Szalachowski, "Maximal extractable value (MEV). Protection on a DAG," preprint arXiv:2208.00940, 2022, doi: 10.48550/arXiv.2208.00940.
[58] Waterfall, Waterfall testnet 3. Test 9, 2022, https://waterfall.foundation/testnet_3_test_9.

Advances in Science, Technology and Engineering Systems Journal Vol. 8, No. 3, 244-249 (2023)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Simulation of Obstacle Detection Based on Optical Flow Images for Avoidance Control of Mobile Robots

Mai Ngoc Anh*

Le Quy Don Technical University, Hanoi, Vietnam

ARTICLE INFO

Article history:
Received: 20 April, 2023
Accepted: 14 June, 2023
Online: 25 June, 2023

Keywords:
Noise-free optical flow
Obstacle detection
Mobile robot control

ABSTRACT

The article presents a simulation of obstacle detection based on noise-free optical flow images for motion control of mobile robots. The detection of hazardous areas in optical flow images is accomplished by dividing multiple layers of optical flow vectors into equal parts. Based on the results of calculating the average magnitude of the vectors in the divided parts and using a solution of comparing these average magnitudes with each other, the robot can figure out obstacle position to avoid and guide the robot to a safe direction.The experiments are simulated on Matlab program to test the performance of the system. The simulated office environment with many obstacles randomly arranged along the corridors is used to test the ability to recognize obstacles to avoid. Simulation results related to different obstacle density scenarios are analyzed to demonstrate the stability of obstacle detection from the noise-free optical flow images.

1. Introduction

There are many solutions for reconstructing 3D structure based on optical flow vectors of 2D images taken by camera(s) mounted on mobile robot. However, the proposed solutions are not be feasible if the optical flow image contains noise vectors. One of the approaches to remove noise in optical flow vectors introduced at the conference ICECCME-2022 [1] makes it possible to study optical flow images on unstable light conditions based on an optical flow noise filter with dynamic thresholds computed by relatively comparing between the average magnitude value of optical flow vectors and an outlier magnitude value. After removing noise vectors, optical flow imaging can be used for mobile robots to observe the working environment and perform complex tasks. Derived from noise-free optical flow images, this article is the next extension of [1] in aspect of experimental simulation of obstacle detection from the noise-free optical flow images to help mobile robots move safely in a simulated indoor office environment.

Over the past decades, mobile robots have been widely used in many production areas such as industry [2] and agriculture [3], in service activities such as search and rescue [4] and in offices [5]. Mobile robots have proven to be effective in supporting humans through their ability to move flexibly and safely in various ways such as by wheels [6], crawlers [7], omnidirectional wheels [8]. Many kinds of sensors such as infrared [9], ultra-sonic [10] and laser [11] have been applied to help mobile robots recognize the surrounding context and make corresponding decisions according to the set requirements. Among the studies of the sensor-based recognition on mobile robots, image-based recognition solutions have been studied a lot thanks to their ability to recognize many objects with different shapes, sizes and distances [12].

Based on the acquired images, the robot performs 2D or 3D scene reconstruction to identify objects to distinguish between a target (to approach) and an obstacle (to avoid). However, image processing often has to deal with the difficult problems of processing time and noise in captured images [13]. The optical flow vector-based identification solution is researched on the idea of biomimetic mimicry of bees in observing, acquiring and processing dynamic image information [14]. Technically, classical techniques of optical flow recognition usually use the method of comparing the difference in vector density of the left half with that of the right half to determine the direction of the obstacle [15].

In this study, the obstacle detection solution based on optical flow images filtered out noise vectors will be tested on a simulated environment with some situations related to different density of obstacles. The movement trajectory of the robot is automatically recorded to analyze and evaluate the safe movement when passing through the areas arranged many obstacles randomly.

The next content is organized as follows: Firstly, the optical flow processing is introduced. Then, the solution of locating obstacle from noise-free optical flow images is analyzed. After that, the simulation of optical flow-based mobile robot moving and

*Corresponding Author: Mai Ngoc Anh, maingocanh.atc@mta.edu.vn

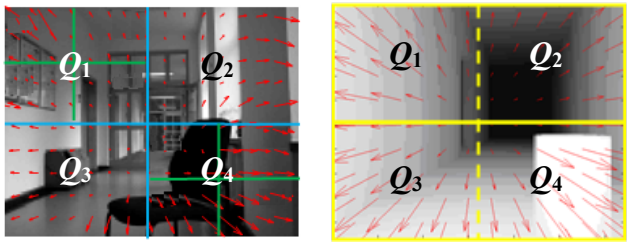
avoiding obstacles in a simulated office is depicted and analyzed. Finally, the conclusion are drawn.

2. Optical flow processing

The processing of real optical flow vectors is shown in Figure 1a and that of simulation noise-free optical flow vectors is illustrated in Figure 1b. The multi-layer segmentation of the optical flow vectors depicted in Figure 1 can be explained as the following schedule:

In the first processing stage, the optical flow image is divided into four equal parts Q_1, Q_2, Q_3 and Q_4. Mathematically, the average magnitude matrix of optical flow vectors in the first processing stage is formulated as follows:

$$\mathbf{Q}^1 = \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix} \quad (1)$$



a) b)

Figure 1: Segmentation of real (a) and simualtion (b) noise-free optical flow images.

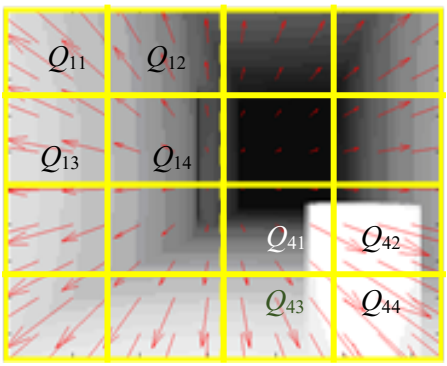


Figure 2: Segmentation in the second processing stage.

Figure 2 illustrates the second stage of multi-layer segmentation of optical flow vectors. In this stage, the four first processing parts are divided into four smaller parts Q_{1i}, Q_{2i}, Q_{3i} and Q_{4i}, where i = 1…4. The average intensity matrix of the optical flow vectors in the second processing stage has the following form:

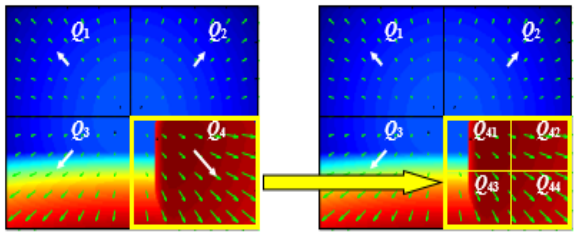
$$\mathbf{Q}^2 = \begin{bmatrix} Q_{11} & Q_{12} & Q_{21} & Q_{22} \\ Q_{13} & Q_{14} & Q_{23} & Q_{24} \\ Q_{31} & Q_{32} & Q_{41} & Q_{42} \\ Q_{33} & Q_{34} & Q_{43} & Q_{44} \end{bmatrix} \quad (2)$$

By the similar way, in the nth processing stage, the four parts of the (n-1)th processing stage are divided into four smaller parts of the deeper layer of the processing, that means $\mathbf{Q}^1$, $\mathbf{Q}^2$, ... and $\mathbf{Q}^n$ are matrices of average magnitude of optical flow vectors in processing stage 1, 2,... and n, respectively.

3. Locating obstacle from noise-free optical flow images

To detect obstacle position during robot movement, various depth-based recognition methods [12] and [16] are implemented by processing images captured by a single camera mounted on the mobile robot. Similarly, optical flow vector images are analysed after multi-layer segmentation to locate the obstruction areas and obtain depth information through qualitative calculations. In other words, the basic principle of determining the location of obstacles is to identify a part of the image having the average Q_i bigger than the others in the magnitude matrix of optical flow vectors.

Figure 3 illustrates the way for locating the obstacle based on the multi-layer segmentation of optical flow vectors. It is easy to see in Figure 3a that in the first processing stage, part Q_4 has the biggest magnitude due to covering the obstacle. In the second processing stage shown in Figure 3b, parts Q_{42} and Q_{44} have the biggest magnitudes because the obstacle fills all these parts compared to half filling in parts Q_{41} and Q_{42}.



a) b)

Figure 3: Locating obstacle from optical flow image: a) layer-1; b) layer-2.

As shown in Figure 4, an optical flow vector $\vec{r}_{ij}$ contains two components projected onto axes OU and OV as follows:

$$\vec{r}_{ij} = u_{ij} \cdot \vec{e}_U + v_{ij} \cdot \vec{e}_V \quad (3)$$

where $\vec{e}_U$ and $\vec{e}_V$ are unit vectors, u_{ij} and v_{ij} are amplitudes of a vector $\vec{r}_{ij}$ projected on axes OU and OV, respectively.

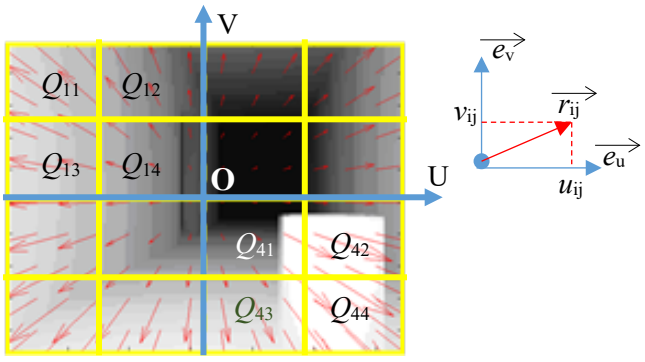


Figure 4: Projecting an optical flow vector $\vec{r}_{ij}$ onto axes OU and OV

Let consider Q_{44} containing 16 optical flow vectors arranged in four rows and four columns by the following matrixes:

$$\mathbf{u}_{44} = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{32} & u_{33} & u_{34} \\ u_{41} & u_{42} & u_{41} & u_{44} \end{bmatrix} \quad (4)$$

$$\mathbf{v}_{44} = \begin{bmatrix} v_{11} & v_{12} & v_{13} & v_{14} \\ v_{21} & v_{22} & v_{23} & v_{24} \\ v_{31} & v_{32} & v_{33} & v_{34} \\ v_{41} & v_{42} & v_{41} & v_{44} \end{bmatrix} \quad (5)$$

By averaging all amplitudes u_{ij} and v_{ij} in each column, we have the average value μ_{Uj} and μ_{Vj} as follows:

$$\mu_{Uj} = \frac{1}{n}\sum_{i=1}^{n} u_{ij} \quad (6)$$

$$\mu_{Vj} = \frac{1}{n}\sum_{i=1}^{n} v_{ij} \quad (7)$$

where n is number of components at column j. As the illustration in Figure 4, $n = 4$ and $j = 1\ldots4$.

From equation (4) and (6), let compare the average value μ_{Uj} of all u_{ij} in one column and replace the biggest average by 1, we have the matrix of obstacle direction as follows:

$$\mathbf{obstacle}_{U44} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (8)$$

Similarly, from equation (5) and (7), let compare the average value μ_{Vj} of all v_{ij} in one row and replace the biggest average by 1, we have the following matrix of obstacle direction as follows:

$$\mathbf{obstacle}_{V44} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad (9)$$

Based on the equation (8) and (9), the mobile robot can locate the most dangerous area in the noise-free optical flow image.

4. Simulation

4.1. Simulation setup

The mobile robot control system with a camera is simulated on the Matlab program and organized into functional modules as illustrated in Figure 5. The two coordinate systems including the mobile robot's coordinate system (body frame) related to the motion parameters of the simulated mobile robot and the office's coordinate system (inertial frame) related to the office environment are shown in Figure 6.

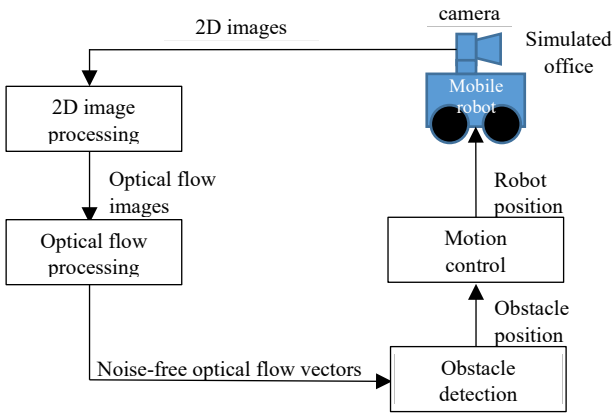


Figure 5: Simulated control system of mobile robot

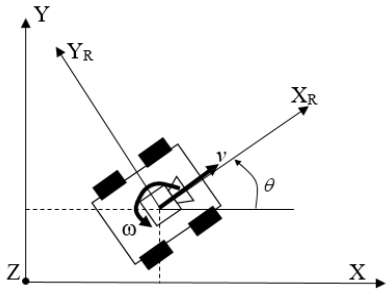


Figure 6: Two coordinate systems and motion parameters of mobile robot

The motion in the simulated mobile robot's coordinate system $(OXYZ)_R$ can be defined from the office's coordinate system (OXYZ) as follows:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{bmatrix} = \mathbf{R}_z(\theta) \cdot \begin{bmatrix} v \\ \omega \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} v \\ \omega \end{bmatrix} \quad (10)$$

where

- v, ω, and θ are the linear velocity, angular velocity, and rotational angle of the robot's coordinate system related to the office's coordinate system, respectively.

- $\dot{x}$, $\dot{y}$, and $\dot{\theta}$ are the linear velocities and the rotational angular velocity, respectively because the variable x, y and θ are positions and rotational angle of the robot , respectively.

- $\mathbf{R}_z$ is the rotation matrix on the Z-axis.

The simulated robot has a width of 50 cm and a length of 60 cm. The office is simulated as shown in Figure 7 to test the robot's ability to move through the recorded camera images. The top view of the simulated office in Figure 7a shows that there are four rooms in the central area and some doors connecting the rooms to the corridor surrounding the office. The side view of the office in

Figure 7b provides door (entrance) and ceiling height information. The width of the entrance is 130 cm and the height of the ceiling is 350 cm. A typical camera image is extracted to visualize the robot's view of the work area as the illustration in Figure 7c.

The main objective of the experiments is to test robot's ability to avoid obstacles randomly arranged along the road in different situations based on the result of noise-free optical flow recognition.

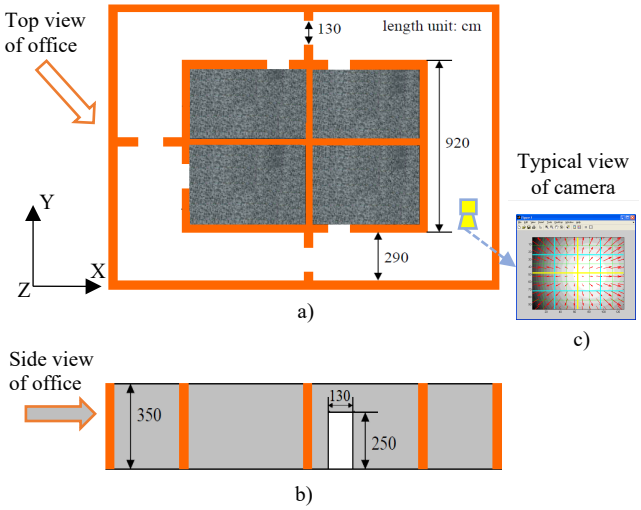


Figure 7: Simulated office for testing obstcle avoidance of mobile robot

4.2. Simulation results

The simulation is performed by the Matlab program with the computer speed 3.40GHz of Intel(R) Pentium(R) 4 CPU and the image processing is carried out by the GPU NVIDIA GeForce GTX 260 1.24GHz with segment size 32 x 32 pixels.

a) Simulated environment image without obstacles

Before testing the robot's ability of obstacle detection, we should review the typical contexts without obstacle in the simulated office situations related to walls, corners, doors and corridors.

Optical flow images are shown in Figure 8 when the robot copes with the walls: the straight view of the camera to a wall (Figure 8a) and the slanted view of the camera to the wall (Figure 8b).

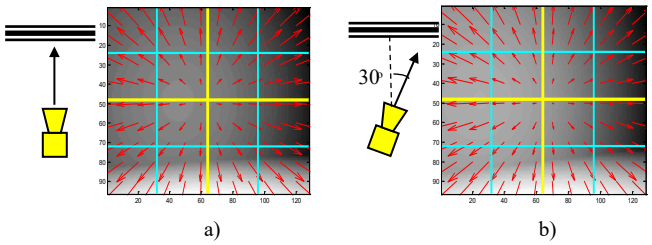


Figure 8: Optical flow images of wall: a) straight view, b) slanted view

Optical flow images are depicted in Figure 9 as the robot approaches the corners of the office: straight view of the corner center (Figure 9a) and slanted view of the corner center (Figure 9b).

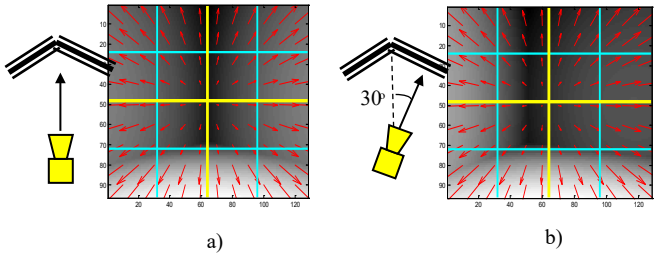


Figure 9: Optical flow images of corner: a) straight view, b) slanted view

Optical flow images are illustrated in Figure 10 as the robot approaches the doors of the office: straight view of the door center (Figure 10a) and slanted view of the door center (Figure 10b).

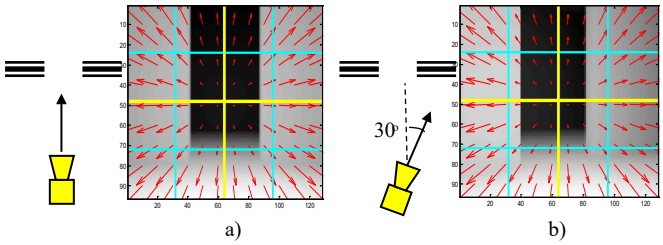


Figure 10: Optical flow images of door: a) straight view, b) slanted view

Optical flow images are sketched in Figure 11 during the robot following the corridors of the office without any obstacle: straight view of the corridor (Figure 11a) and slanted view of the corridor (Figure 11b).

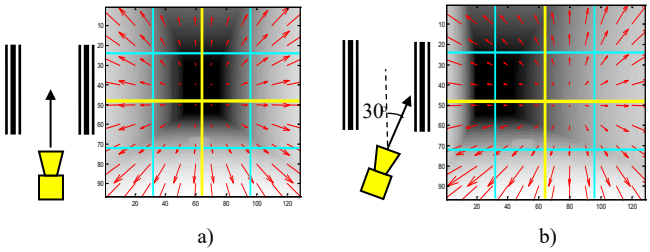


Figure 11: Optical flow images of corridor: a) straight view, b) slanted view

The statistical data of optical flow images are shown by colorful 3D histograms in Figure 12 for different distances (right horizontal axis) including near (blue), middle (yellow) and distal (red) for different objects including walls (A1), corners (A2), doors (A3) and corridors (A4) in 4 depth stages of optical flow vector processing (vertical axis) to evaluate tilt angles between the camera and the observed object by measuring the ratio of the mean value between the left optical quadrant and the right optical quadrant (nearest horizontal axis).

It is easy to see that the optical flow vectors in these situations are distributed differently between images despite the absence of obstructions. However, in the cases of close distances to the objects, the robot still has to avoid the nearest object as if avoiding

an obstacle. Based on the ratio between the left parts and the right part, the principle for avoiding obstacle is that the robot travels to the direction containing the smaller average value between left and right optical flow parts.

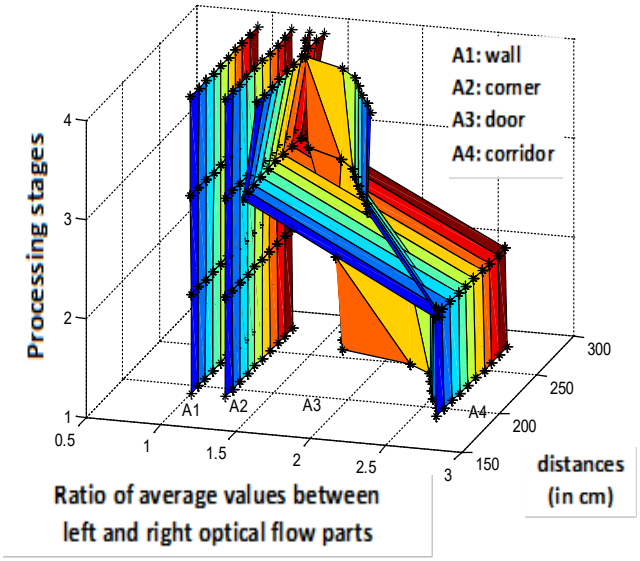


Figure 12: Ratio of average values between left and right optical flow parts.

b) Simulation test with low obstacle density

In the first test, the robot moved along a corridor containing several obstacles randomly arranged along the path illustrated in Figure 13. On the moving trajectory, the robot has to detect obstacles through the noise-free optical flow vectors.

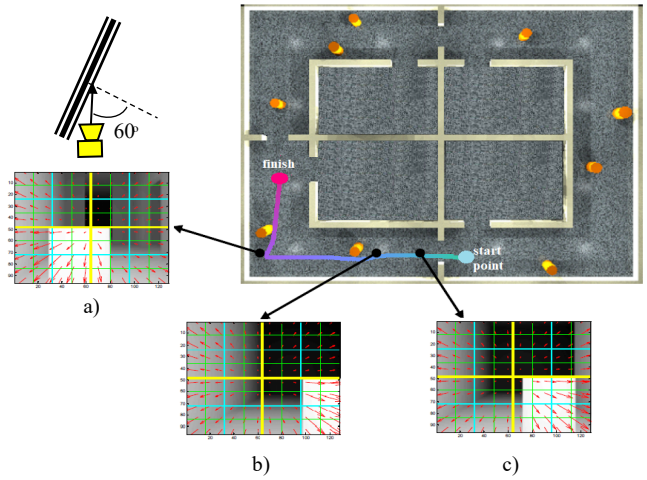


Figure 13: Robot avoids single obstacles

The robot movement is automatically recorded and depicted by a colorful truth-ground line. Along this line, several images are extracted to demonstrate that the robot has moved from the right to the left of the corridor and avoided a close front left obstacle (Figure 13a), a close right obstacle (Figure 13b), and a far front right obstacle (Figure 13c). The line with the three extracted optical flow images show the obstacle positions detected from the optical flow vectors and the robot succeeded in moving safely by avoiding the two detected obstacles.

In the second test, the robot travelled along a different part of the corridor. It has to avoid not only a single obstacle but also a pair of obstacles on the corridor. The robot movement trajectory is automatically recorded by a colorful truth-ground line as shown in Figure 14. Along the line, several images are taken out to prove that the robot has moved from the right to the left of the corridor. The robot avoided a close front obstacle (Figure 14a), a pair of obstacles (Figure 14b) in far distance, and a pair of obstacle and wall (Figure 14c).

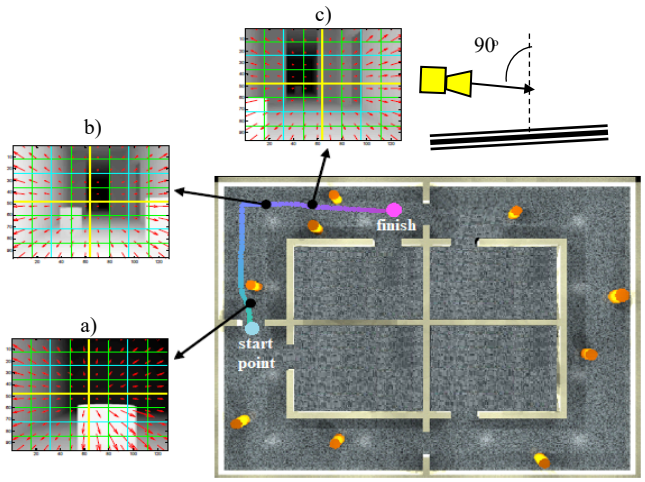


Figure 14: Robot avoids a pair of obstacles

The successful movements of the robot in the two first tests verify the robot's ability of obstacle detection based on the optical flow vectors taken from the simulated images.

c) Simulation test with high obstacle density

In the third test, the robot performed the more difficult task of safe movement through the corridor containing much more obstacles randomly arranged in a cramped area.

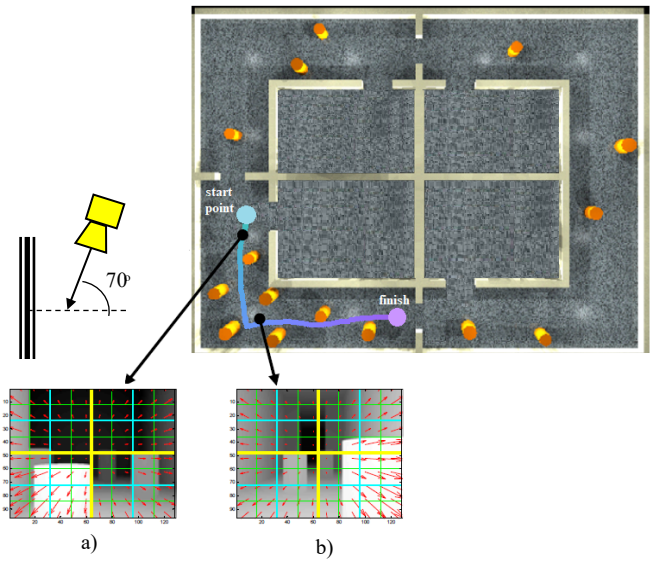


Figure 15: Robot avoids many obstacles in a cramped area

Technically, the solution of dividing the optical flow vector into multiple layers allows the robot system to compare the average magnitude of a part with other parts in the same layer. That means it can detect multiple obstacles at the same time and estimate the distance through the magnitude of the average vector. As a result, the robot can choose the appropriate (possibly suboptimal) movement direction to avoid obstacles.

The moving trajectory of the robot through the narrow area with many obstacles is automatically recorded by the colored line illustrated in Figure 15. Along this line, two images are extracted to represent the movement trajectory of the mobile robot after avoiding the closest obstacle on the left (Figure 15a), passing the area with high obstacle density, and avoiding another close obstacle on the right (Figure 15b) before leaving the danger area successfully.

5. Discussion and Conclusion

The above simulation results not only show the stability of optical flow perception in different situations, but also show that the obstacle position detection based on optical flow is independent of the robot's speed. Thanks to the simple calculation method based on the average vector comparison method, the robot does not need large memory to store complex database of obstacle characteristics such as shape and size.

In other words, the simulation results of obstacle detection based on optical flow free-noise images for mobile robot demonstrate that the recognition of dangerous area in the optical flow image is able to perform through the multi-layer processing of optical flow vectors in to many equal parts. By simply comparing the average magnitude values of the divided parts, the mobile robot can figure out obstacle positions to guide the robot follow a suitable safe way.

The limitation of the simulation results is that the mobile robot copes with static obstacles, but not dynamic obstacles. This indicates that in the next phase, the research team need to build an actual mobile robot to test in real environments with dynamic obstacles to more accurately assess the stability of the optical flow-based recognition method.

Furthermore, in the future the real tests will be performed in a complex indoor environment with unstable lighting conditions to estimate the ability of optical flow-based recognition.

Conflict of Interest

The author declares no conflict of interest.

References

[1] N. Anh Mai, "Multi-layer segmentation solution to filter the noise of optical flow vectors to assist robots in object recognition inside buildings," International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME 2022, (November), 16–18, 2022, doi:10.1109/ICECCME55909.2022.9988355.

[2] A. Markis, M. Papa, D. Kaselautzke, M. Rathmair, V. Sattinger, M. Brandstotter, "Safety of Mobile Robot Systems in Industrial Applications," Proceedings of the ARW & OAGM Workshop 2019, 26–31, 2019, doi:10.3217/978-3-85125-663-5-04.

[3] S.S.H. Hajjaj, K.S.M. Sahari, "Review of research in the area of agriculture mobile robots," Lecture Notes in Electrical Engineering, **291 LNEE**(January 2018), 107–118, 2014, doi:10.1007/978-981-4585-42-2_13.

[4] N. Ruangpayoongsak, H. Roth, J. Chudoba, "Mobile robots for search and rescue," Proceedings of the 2005 IEEE International Workshop on Safety, Security and Rescue Robotics, **2005**(July), 212–217, 2005, doi:10.1109/SSRR.2005.1501265.

[5] W. Chung, G. Kim, M. Kim, C. Lee, "Integrated Navigation of the Mobile Service Robot in Office Environments," International Conference on Control, Automation, and Systems (ICCAS), 2033–2038, 2003.

[6] M. Trojnacki, P. Dąbek, "Mechanical properties of modern wheeled mobile robots," Journal of Automation, Mobile Robotics and Intelligent Systems, **13**(3), 3–13, 2019, doi:10.14313/JAMRIS/3-2019/21.

[7] A. Dobrokvashina, R. Lavrenov, E. Magid, Y. Bai, M. Svinin, R. Meshcheryakov, "Servosila Engineer Crawler Robot Modelling in Webots Simulator," International Journal of Mechanical Engineering and Robotics Research, **11**(6), 417–421, 2022, doi:10.18178/ijmerr.11.6.417-421.

[8] K. Kanjanawanishkul, "Omnidirectional wheeled mobile robots: Wheel types and practical applications," International Journal of Advanced Mechatronic Systems, **6**(6), 289–302, 2015, doi:10.1504/IJAMECHS.2015.074788.

[9] M. Al-Mallah, M. Ali, M. Al-Khawaldeh, "Obstacles Avoidance for Mobile Robot Using Type-2 Fuzzy Logic Controller," Robotics, **11**(6), 2022, doi:10.3390/robotics11060130.

[10] S. Adarsh, S.M. Kaleemuddin, D. Bose, K.I. Ramachandran, "Performance comparison of Infrared and Ultrasonic sensors for obstacles of different materials in vehicle/ robot navigation applications," IOP Conference Series: Materials Science and Engineering, **149**(1), 2016, doi:10.1088/1757-899X/149/1/012141.

[11] J. Cai, T. Matsumaru, "Human detecting and following mobile robot using a laser range sensor," Journal of Robotics and Mechatronics, **26**(6), 718–734, 2014, doi:10.20965/jrm.2014.p0718.

[12] T.A.Q. Tawiah, "A review of algorithms and techniques for image-based recognition and inference in mobile robotic systems," International Journal of Advanced Robotic Systems, **17**(6), 1–25, 2020, doi:10.1177/1729881420972278.

[13] Y. Zhu, C. Huang, "An Improved Median Filtering Algorithm for Image Noise Reduction," Physics Procedia, **25**, 609–616, 2012, doi:10.1016/j.phpro.2012.03.133.

[14] M. Egelhaaf, "Optic flow based spatial vision in insects," Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, (0123456789), 2023, doi:10.1007/s00359-022-01610-w.

[15] Y. Zhang, R. Huang, W. Nörenberg, A.B. Arrenberg, "A robust receptive field code for optic flow detection and decomposition during self-motion," Current Biology, **32**(11), 2505-2516.e8, 2022, doi:10.1016/j.cub.2022.04.048.

[16] Z. El Kadmiri, O. El Kadmiri, L. Masmoudi, "Depth estimation for mobile robot using single omnidirectional camera system," Journal of Theoretical and Applied Information Technology, **44**(1), 29–34, 2012.