Editorial

Advances in Science, Technology and Engineering Systems Journal (ASTESJ) is an online-only journal dedicated to publishing significant advances covering all aspects of technology relevant to the physical science and engineering communities. The journal regularly publishes articles covering specific topics of interest.

Current Issue features key papers related to multidisciplinary domains involving complex system stemming from numerous disciplines; this is exactly how this journal differs from other interdisciplinary and multidisciplinary engineering journals. This issue contains 42 accepted papers in electrical, computing and automation domains.

Editor-in-chief
Prof. Passerini Kazmersk

CONTENTS

Traditional and Deep Learning Approaches for Sentiment Analysis: A Survey

Fatima-Ezzahra Lagrari*, Youssfi Elkettani

Department of Mathematics, MISC Laboratory, Ibn Tofail University, Kenitra, 14000, Morocco

A R T I C L E I N F O	A B S T R A C T
	Presently, individuals generate tremendous volumes of information on the internet. As a result, sentiment analysis is a critical tool for automating a deep understanding of user-generated information. Of late, deep learning algorithms have shown endless promises for a variety of sentiment analysis. The purpose of sentiment analysis is to categorize different descriptions as good, bad, or impartial based on context data. Numerous studies have been concentrated on sentiment analysis in addition to the ability to examine thoughts, views, and reactions. In this paper, we review classical and deep learning approaches that have been applied to various sentiment analysis tasks and their evolution over last years and provide performance analysis of different sentiment analysis models on particular datasets. In the end, we will highlight current challenges and suggested solutions that can be considered in future work to achieve better performance.

1. Introduction

Nowadays, the number of web 2.0 resources where every one of us is comfortable sharing our views and suggestions is growing exponentially such as e-commerce websites and social media. As a result of this growth, massive volumes of data are generated. Hence, sentiment analysis was developed to extract relevant information from this large data size automatically. Sentiment analysis is a subfield of natural language processing (NLP) which gets considerable attention from researchers, industry, and government organizations worldwide. Sentiment Analysis, also referred as view mining, which is a procedure that assists in defining the emotional tone of a collection of words; this analysis enables us to determine the impression, viewpoint or conveyed emotion. According to Liu's concept, an opinion consists of a target, an object or a feature of an entity, and a mood (positive, negative, or neutral) [1].

Additionally, it may include the originator of the view and the day and time at which the opinion was given [2]. Sentiments are categorized into numerous groups based on the classification purpose. Authors in [3] and [4] divided tweets into favorable, unfavorable, and neutral tweets. Sites like Twitter media platforms where only a vast number of tweets are posted, many of which include helpful information. As stated, [5] active Arabic users sent over 20 million tweets each day in March 2014. Each minute, millions of tweets are posted, many of them in the original

languages. Health-related issues are frequently trending on Twitter.

Numerous approaches to sentiment analysis have been introduced. Traditional techniques are categorized into two types: lexicon-based and machine learning-based. Lexicon-based methods, in which the alignment of a document is determined founded on the sentiment polarity of phrases also words in the sentiment [6], and machine learning methods, in which models are constructed from labeled examples of readers or sentences, for instance, The paper in [7] categorize text documents containing film reviews using machine learning techniques to evaluate the statement's feelings.

While traditional methods have achieved a high degree of precision, the feature engineering required is time-consuming and complex. Today's massive amount of user-generated data necessitates complete knowledge and reliable ways for its management. Thus, deep learning methods facilitate the development of computational models under their ability to learn from data sets without requiring a manual attribute selection process. Recent research indicates that deep learning techniques outperform standard machine learning techniques in sentiment analysis [8].

2. Levels of sentiment analysis

Sentiment Analysis can be conducted at three granularity levels: the document level, sentence level, and aspect level [9]. Its application is a highly demanding area of investigate that entails

*Corresponding Author: Fatima-ezzahra Lagrari.
Fatima.ezzahra.lagrari@gmail.com

a numerous of complex tasks. Subjectivity identification, sentiment analysis, lexicon construction, component extraction, aspect sentiment analysis, and viewpoint fraud detection are the most investigated tasks [10]. Sentiment analysis techniques are confronted with numerous issues relating to traditional sentiment analysis, autonomous natural language processing, and also new and challenging hurdles in this novel communication situation. This difficulty is mainly due to variant signals' properties, which are typically brief but dense with semantic information [11], unstructured, thrashed in variant languages, and involve clutter. Amongst the features discussed previously, natural language text study is a critical component on which the sentiment analysis community bases its work.

Individuals and organizations/governments alike are incredibly interested in sentiment analysis. Before purchasing a product, we examine the ratings and reviews on the internet, while businesses utilize these reviews to understand their consumers' preferences better and improve their products. Numerous attempts have been made to incorporate sentiment analysis into customer feedback. Governments use the same technologies to gauge public sentiment, for example, in response to new tactics or during polls.

3. Sentiment analysis approaches

Before As mentioned above, traditional approaches were the first ones that have merged to deal with sentiment analysis. Globally we have two categories in traditional approaches: the lexicon based and machine learning approaches.

3.1. Lexicon based approaches

Before, As indicated previously, traditional methodologies were the first to converge on sentiment analysis. Globally, traditional methods fall into two categories: lexicon-based approaches and machine learning approaches. Lexicographical techniques classify a given text into its appropriate class by utilizing precompiled sentiment lexicons with various words and their orientation (positive, negative, etc.) [12]. The terms contained in a particular document or sentence can be retrieved using the bag of words technique and classified using a lexicon. A merging method such as the average of each class established can consolidate the data and form a judgment about the overall sentiment of the phrase.

The majority of ways for producing sentiment lexicons are either dictionary-based or corpus-based: the first category relies on a lexicon in which words are labeled with their past polarity (seed words). WordNet is an example of a lexicon in this area, as is WordNet-Affect [13], which associates adaptive synsets (a collection of synonyms) with inspirational ideas. SentiNet [14] annotates each WordNet synset based on three numerical sentiment scores (positivity, negativity, and neutrality). At the same time, MPQA [14] offers a list of words, together with their PO Stagging, labeled with polarity (positive, negative, and neutral) and intensity (strong, weak). The sensations of a new document / phrase can be examined by comparing synonyms of names, verbs, adjectives and adverbs to the seed words previously labeled in the vocabulary. The second category mines a domain corpus for sentiment terms and their polarity. For instance, considering the co-occurrence likelihood of positive and negative phrases from search engines [15]. The corpus-based approaches address

domain independence and provide a more accurate sentiment analysis [16]. The authors of [17] began by establishing a similarity between each pair of words in the corpus by generating a word graph with nodes representing words and reinforced edges representing similarities. The unidentified words are then automatically formed lexicons by attributing their polarity based on the graph. Certain lexicons feature non-words as well, such as emoticons [18]. SentiStrength is a hybrid sentiment analysis software suggested by [19] for extracting sentiment strength from the informal English text. Between 1 and 5, a score for negative or positive sentiment is assigned. It is built on a manual sentiment word list that was improved using a training algorithm, followed by a spelling correction algorithm, a booster and negation word list, and popular emoticons.

It is worth noting that developing or even maintaining a sentiment lexicon is a challenging effort in the modern era due to the rise of current data that is unstructured mainly and user-generated, such as social media. Rather than relying solely on the sentiment values included in the lexicon, the background and other characteristics of the words must be considered; this is accomplished through a process known as feature engineering. Bag of Words is one of the NLP's preparation techniques. It separates the different words from a text [20]. This technique will be unable to distinguish between "I like you," where "like" is a positive verb, and "I am like you," where "like" is a neutral proposition. To address this, Part of Speech POS enables the assignment of a word in a corpus to the appropriate part of the speech tag (verbs, nouns, adjectives, and adverbs) depending on its context and definition. Adjectives and adverbs are mainly employed to describe entities that exhibit emotional characteristics.

Apart from using individual words (unigrams) as features, classic text classification algorithms weight the words within a sentence using Word Positions and Frequency. For instance, a movie review may begin with an overall sentiment statement, move to a conversation, and close with a summary of the author's opinions [1], indicating that position can be used to change sentiment weights. The term frequency can be used to highlight terms that may be significant because they appear in many publications. For example, the TF-IDF (Term Frequency - Inverse Document Frequency) approach [21] enables this to be calculated by comparing the frequency of occurrence of a given term in a document to the number of documents containing that word. Another strategy is to hunt for the most uncommon words rather than the most common ones. In [7] , the authors optimize their system's performance by referring to it as "presence" rather than "frequency." Rather than weighing all words equally, some approaches favor terms that reflect or imply thoughts and opinions, such as excellent, great, wrong, incorrect, natural, like, agree, hate, trust, conceive, and so forth [22].

Another factor to consider when interpreting the context of a word is negation. To ascertain the negation's scope and flip the polarity of terms affected by a negation [23]. Negation phrases include the following: not, no, never, cannot, should not, and would not. The dependency-based aspects of word words were derived from Dependency trees provide a simple tree-based method to express the structural relationships between the sentence's words. These relationships are denoted by triplets that

contain the name of the relationship type, the relationship's governor, and the relationship's dependent [24]. While these systems do not require training datasets, they strain to adapt to new data trends caused by the changing nature of language, the growth in high-dimensional data, the structural and cultural intricacies of brief text such as tweets, and the usage of emoticons and abbreviations [22]. To address the challenge of rapidly assessing these new data types, machine learning algorithms have demonstrated their superiority over lexicon-based methods.

3.2. Machine learning approaches

Machine Learning methods are based on machine learning algorithms that classify annotated data into its intended category. They can be classified into three broad categories: supervised, semi-supervised, and unsupervised learning methods. When humans do not annotate data, individuals opt for lexicon-based approaches. The machine learning approach is fully automated, convenient, and capable of handling extensive collections of data. These methods require a training dataset to assist in automating the Classifier; it is used to develop a classification model that classifies feature vectors [25] and a test dataset to predict the class label of unseen feature vectors.

On numerous sentiment analysis jobs, machine learning is used; however, the two classical tactics to sentiment analysis can be merged to maximize each benefit. In [26], the authors demonstrated promising results when they combined machine learning and lexicon-based practices. Several machine learning approaches are used to categorize data, including Naive Bayes (NB), Maximum Entropy (ME), and Support Vector Machines (SVM). Sentiment classification studies on polarity determination, utilizing the Support Vector Machine SVM algorithm and a variety of characteristics selection methods, including word frequency, binary occurrence, and term occurrence [27].

In [28], the authors created a method for text categorization that combined a genetic algorithm and a Random forests classifier and achieved extremely high accuracy. Crawford et al. investigated opinion spam detection using SVM and Naive Bayes [29]. They created a collection of unfavorable deceptive opinions that included both genuine and misleading reviews. They illustrate that linear SVM produced results more than 80% of the time. In [30], authors published a new technique for sentiment analysis that is based on Bayesian network classifiers. The authors employed the Bayes factor approach to reduce the number of edges during the training procedure automatically. The proposed technique was evaluated using two Spanish datasets. The evaluation's outcome demonstrated the superiority of the submitted work above previously published efforts.

A hybrid technique combining machine learning (SVM classifier) and semantic orientation was proposed [31]. We applied content-neutral, content-specific, and sentiment-based features. Senti WordNet was used to derive the weights for the sentiment characteristics. These methodologies strive to keep up with the new data trend of a dynamic language, the growth of high-dimensional data, and the structural and cultural intricacies of short text such as tweets. To address the challenge of efficiently processing this type of data, deep learning algorithms have demonstrated their superiority over more traditional methods.

3.3. Deep learning approaches

Deep learning is a rapidly growing subfield of machine learning techniques based on artificial neural networks (ANNs) built based on the biological brain and capable of self-learning. The web comprises dozens to hundreds of neurons referred to as layers; each layer receives and interprets data from the preceding layer, allowing for multiple processing. For instance, the program will learn to detect letters before targeting words in a text or identify whether a photo has a face before determining who it belongs to. Deep learning algorithms have been primarily used in sentiment analysis due to their ability to automatically learn and generate input models from datasets [32].

Recent studies present numerous ways to sentiment analysis based on deep learning, including supervised and unsupervised algorithms. We will start by reviewing word embedding approaches as the basic data processing layer of deep learning and then examine many different deep learning approaches; We recommend to read [31] for mathematical details.

3.3.1. Word embedding

Word embedding is a collection of machine learning approaches aiming to express words in a text using vectors of real numbers, with similar vectors representing words with similar interpretations. These new forms of textual data have increased the efficiency of systems for automatic language processing. The first approach to word embedding dates back to the 1960s and is based on dimensionality reduction techniques. Recent advances in performance have been facilitated by innovative methods based on probabilistic models and neural networks, such as Word2Vec [33]. The following sections will discuss various word embeddings that are frequently utilized in sentiment analysis. Words having comparable semantic and contextual properties form a class. Authors in [34] describe a language model that simultaneously learns both distributed representations and the probability function for word sequences. Numerous studies have been conducted to strengthen this work. For instance, in [35] authors Suggested a classifier word embedding based on convolutional neural networks that can learn the characteristics associated with a particular task without basic info.

Authors in [33] proposed Word2vec which is a common pre-trained word embedding technique that uses two-layer neural networks to learn the vector representations of the words in a text so that similar numerical vectors express words with related comments. Word2Vec utilizes two neural architectures: CBOW and Skip-Gram. BOW takes the setting of a word as input and make efforts to identify the word in question. Skip-Gram, on the other hand, accepts a word as data and tries to guess its circumstances. In both situations, the network is trained by parsing the text and adjusting the neural weights to lower the algorithm's model complexity.

Glove2 is another popular term for embedding; it provides a global neural network and local context. In [36] authors presented a method for fast text word embedding in which n-grams encode words. Notably, the NLP Branch earned a millstone, Bidirectional encoder representation from transformers (Bert) language model [8]. This is an utterly fascinating work. The authors detail a masked LM (MLM) that enables bidirectional training in models

and subsequent acquisition of a strong sense of language context. In [3], authors obtained impressive findings by improving the Bert algorithm utilizing the lion method, in which the weights of the network of the BERT encoder were optimized using standard LA and also the sequence number length of the BERT encoder.

3.3.2. Convolution neural networks

A convolutional neural network (CNN) is a feedforward neural network established in 1989 and is the most widely used in computer vision. The visual cortex of animals inspires it. Each neuron in the visual brain has a small receptive area, which overlaps to perceive the complete object; these convolutional layers operate as local filters. In essence, CNN consists of three types of coatings: input, feature selection, and classification [37]. The input layer embeds raw data, the convolution layer learns and produces a feature map, and the neural network extracts practical attributes. Finally, these features are fed into a fully integrated categorization layer [2].

CNN has gotten perfect results in different domains, the same as in NLP and so sentiment analysis, and thus for the additional benefits. CNN utilizes few variables that don't need a lot of time to train the model. They have achieved good popularity in sentiment analysis since they can acquire contextual characteristics but are constrained in long-term connections. CNN's gets satisfactory accuracy when prepared in many samples, which is not always satisfied, where the development of Recurrent neural networks (RNNs) can manage the challenge of long-term dependencies.

3.3.3. Recurrent neural networks

RNNs are invented by ELMAN in 1990 [38], they have neural networks in which information can propagate in both directions, including from deep layers to early layers. They are closer to the true functioning of the nervous system, which is not one-way direction. These networks have recurrent connections in the sense that they keep information in memory: they can take into account a certain number of past states at a given moment t. For this reason, RNNs are particularly well suited to applications involving the context. Indeed, "classical" RNNs (simple recurrent neural networks or Vanilla RNNs) are only able to memorize the so called near past, and begin to "forget" after about fifty iterations. This two-way transfer of information makes their training much more complicated, and it is only recently that effective methods such as LSTM (Long Short-Term Memory) have been developed. Authors in [39] proposed long short-term memory network which is a special type of RNNs able of learning long term dependencies via the forget gate that allows the memory cell to keep information for a long time. GRU (Gated recurrent unit) is another type of RNNs to LSTM that have only two layers instead of three layers as in LSTM, it is a simplified version of LSTM that is more efficient computationally then Vanilla RNNs and LSTM.

A new technique called attention mechanism was introduced in 2015, which simulates the human reading capability of gaze-fixations by prioritizing relevant parts of data based on its weighted representation. Since then, more researches apply this technique [40]. RNNs approaches have been widely applied in sentiment analysis field, A neural network was proposed in [41] which is capable of learning document representation taking into consideration the sentence relationship. First it learns sentence representation with CNN and then a GRU is used to encode semantics.

3.3.4. Recursive neural networks

RecNNs are a generalization of RNNs, a form of neural network used to train a graph-based classification model from data and is familiar with structured inputs of any structure [42]. RecNN produces parent representations underside from the structural term of a sentence by merging tokens to obtain representations for phrases and soon the entire sentence. The representation at the sentence level can then be utilized to perform the final classification on the input text.

3.3.5. Deep Reinforcement Learning

Deep reinforcement learning (DRL) is a relatively new subfield of deep learning. It is a way of artificial intelligence that is not widely used but opens up entirely new possibilities for automation. The DRL enables the development of software capable of matching or exceeding human intellect in various domains. The most well-known system based on DRL is Deep Mind (Google's AI Platform) [43]. DRL employs validation learning algorithms, the most widely used of temporal difference learning (TD Learning) and Q-learning (QL). These learning models are described as the human (and animal) process of information acquiring through trial and error to learn from multiple exposures. To ensure that the action taken by the machine is in an intended way, these algorithms use a reward system to evaluate the machine's selections. The method is comparable to animal training. DRL, which is under the domain of automatic (or unsupervised) machine learning, integrates learning algorithms with neural networks to successfully estimate the validity of a "complex" strategy that incorporates a substantial percentage of decision criteria. The primary problem is to develop a system that promotes specific behaviors while avoiding adverse effects. The objective is to maximize rewards inside a simulation environment in which the computer acts and then receives feedback. The software agent is not informed in advance about the most probable activity and must select its strategy on its own through a process of interpretation.

DRL models are built on the Markov decision process (MDP), enabling an intelligent agent to be trained [2] by contact with the environment. They have demonstrated efficacy in improving structure, which is beneficial in NLP, particularly sentiment analysis. In [44], authors presented a reinforcement learning (RL) method for automatically learning phrase representation by discovering optimum structures. The reward in this research is the chance of correctly classifying the input sentence based on its structured presentation. The results demonstrate the algorithm's efficacy. However, developing a model that learns a function of rewards remains challenging.

3.3.6. Generative adversarial network

GANs are an unsupervised pre-trained network architecture first proposed by Good-fellow and colleagues in 2014 [45]. GANs enable pretraining the network layers with large datasets. We asynchronously train two models: a generative model G that discovers the distribution of the data and a discriminative model

D that indicates the likelihood that a compiled piece of data is from the classification model or the generative model. A GAN aims to analyze real data sets and develop its inventions that must resemble so natural that they cannot be identified as machine images without human interaction. The generative and discriminative models are perpetually in conflict with one another. When the discriminative network detects an error, it returns it to the other network. In this situation, the generative network has not yet reached its optimal state and must keep learning. Simultaneously, the discriminatory network has grown in sophistication. The generative network aims at generating data sets that appear to be so real that they are accepted as confirmed by the discriminative. On the other hand, the discriminative approach aims to thoroughly evaluate and comprehend real-world examples to eliminate the possibility of fraud being detected as accurate. In sentiment analysis, GAN models are the starting point.

3.3.7. *Hybrid deep neural networks*

Deep learning models have been extensively employed in the natural language processing sector and have shown impressive results, particularly sentiment analysis. Numerous researchers have advocated for merging models to maximize the performance of each kind. Multiple hybrid models have been supported for sentiment analysis; for example, [46] implemented a hybrid deep learning architecture for sentiment estimates based on a convolutional neural network for word embedding and a backing vector machine for sentiment classification; evaluation reveals that the proposed mechanism is exceptionally efficient. Authors in [47] created a model that needs to be extracted from several deep learning word embedding methods, mixes them, and classifies texts according to their sentiments.

4. Works on different sentiment analysis datasets

The table below summarizes and compares the performance of various Deep Learning techniques for sentiment analysis tasks.

Table 1: Performance and applications of DL approaches on various SA tasks.

References	Approach	Features	Dataset	Accuracy
[3]	Customized BERT based lion algorithm	-	Sentiment 120	92.2 %
[8]	Bidirectional Encoder Representations from transformers	-	CoNLL-2003	86.7%
[19]	SentiStrength + spelling correction, booster and negation word list + common emoticons	-	MySpace comments	66.7%
[25]	Dictionary based lexicon (SentiWordNet) + SVM	Lexical and syntactic features + unigrams and	Online product reviews	84.15% for kitchen appliances
		bigrams + POS		And 78.85% for books
[26]	SentiWordNet Lexicon + semantic rules + fuzzy set	Word features	Twitter Dataset, Movie review	88.02% for twitter and 75.85% for movie
[27]	SVM	Trigram and Binary Occurrence	Movie review	84.90%
[28]	Genetic algorithm + random forest		Business documents	98.7%
[30]	SVM	BOW	Spanish Twitter Dataset	85.8%
[33]	Recurrent neural network + skip-gram model	6-word window around the target word	Microsoft sentence completion challenge	58.9%
[34]	Neural network	Word features	AP News corpus	
[35]	Convolutional neural network		Prop Bank dataset	85.7%
[36]	Global logbilinear regression model (conditional random field)	Discrete feature from Stanford NER model + five word context	word analogy task	75%
[41]	Recurrent neural network (CNN/LSTM)	-	IMDB	45.3%
[44]	Reinforcement learning (HS-LSTM)	-	Subjectivity dataset	93.7%
[46]	Category sentence GAN with Reinforcement learning + generative adversarial networks + recurrent neural networks	-	Subset of Amazon review dataset and Emotion dataset	41.52% for Emotion and 86.4% for Amazon

5. Current issues and possible directions

Numerous approaches to sentiment analysis have been presented. The majority of the proposed technique for sentiment analysis is static. As a result, one of the significant obstacles in applying deep learning algorithms for sentiment analysis is dynamic sentiment analysis and tracking. For instance, on the Twitter social network, the dataset, vocabulary, and user count might change at any time, making dynamic analysis a problematic

process. On the other hand, dealing with language structure, which includes slang, is a significant difficulty. Additionally, heterogeneous information (i.e., short or long phrases) necessitates in-depth examination and specialized processing [48]. We recommend that researchers employ cutting-edge deep learning techniques to address these issues, such as the Bert algorithm for context representation learning [8], deep reinforcement learning (DRL), and generative adversarial networks (GANs), which are incredibly efficient at resolving complex problems.

6. Conclusion

Sentiment analysis is critical for decision-making in a variety of domains, including economics, business development, and social phenomenon research. Owing to its large use in potential implementation, there has been a meteoric rise in academic research and industrial applications. Recently, a significant number of academics have become interested in sentiment analysis utilizing deep learning algorithms; consequently, large varieties of efficient approaches for various tasks have been developed. We began this subject by discussing the history of sentiment analysis and its different degrees. There has been discussion of both traditional and machine learning methodologies, emphasizing deep learning approaches and their applications. We offered an application and performance analysis on a variety of real-world sentiment analysis datasets. Lastly, we outline existing issues that require attention and make some recommendations for change.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] B. Liu, "Sentiment analysis and opinion mining," Synthesis Lectures on Human Language Technologies, **5**(1), 1–167, 2012. Doi: 10.2200/S00416ED1V01Y201204HLT016

[2] O. Habimana, Y. Li, R. Li, X. Gu, G. Yu, "Sentiment analysis using deep learning approaches: an overview," Science China Information Sciences, **63**(1), 1–36, 2020.

[3] F. LAGRARI, Y. ELKETTANI, "Customized BERT with Convolution Model : A New Heuristic Enabled Encoder for Twitter Sentiment Analysis",2020

[4] S.C. Rachiraju, M. Revanth, "Feature Extraction and Classification of Movie Reviews using Advanced Machine Learning Models," in 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 814–817, 2020.

[5] A.M. Alayba, V. Palade, M. England, R. Iqbal, "Arabic language sentiment analysis on health services," in 2017 1st international workshop on arabic script analysis and recognition (asar), 114–118, 2017.

[6] S. Trinh, L. Nguyen, M. Vo, P. Do, Lexicon-based sentiment analysis of Facebook comments in Vietnamese language, Springer: 263–276, 2016.

[7] R. Alhajj, J. Rokne, Encyclopedia of social network analysis and mining, Springer, 2014.

[8] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," ArXiv Preprint ArXiv:1810.04805, 2018

[9] Y. Ma, H. Peng, E. Cambria, "Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM," in Thirty-second AAAI conference on artificial intelligence, 2018.

[10] M.H. Arif, J. Li, M. Iqbal, K. Liu, "Sentiment analysis and spam detection in short informal text using learning classifier systems," Soft Computing, **22**(21), 7281–7291, 2018. Doi: 10.1007/s00500-017-2729-x

[11] M. Al-Smadi, M. Al-Ayyoub, Y. Jararweh, O. Qawasmeh, "Enhancing aspect-based sentiment analysis of Arabic hotels' reviews using morphological, syntactic and semantic features," Information Processing \& Management, **56**(2), 308–319, 2019. Doi: 10.1016/j.ipm.2018.01.006

[12] I.P. Windasari, D. Eridani, "Sentiment analysis on travel destination in Indonesia," in 2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), 276–279, 2017.

[13] C. Fellbaum, "WordNet: An electronic lexical resource," The Oxford Handbook of Cognitive Science, 301–314, 2017.

[14] E.M. Alshari, A. Azman, S. Doraisamy, N. Mustapha, M. Alkeshr, "Effective method for sentiment lexical dictionary enrichment based on Word2Vec for sentiment analysis," in 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), 1–5, 2018.

[15] K. Ravi, V. Ravi, "A survey on opinion mining and sentiment analysis: tasks, approaches and applications," Knowledge-Based Systems, **89**, 14–46, 2015. Doi: 10.1016/j.knosys.2015.06.015

[16] M. del P. Salas-Zárate, J. Medina-Moreira, K. Lagos-Ortiz, H. Luna-Aveiga, M.A. Rodriguez-Garcia, R. Valencia-Garcia, "Sentiment analysis on tweets about diabetes: an aspect-level approach," Computational and Mathematical Methods in Medicine, 2017.

[17] A. Bittar, S. Velupillai, A. Roberts, R. Dutta, "Using General-purpose Sentiment Lexicons for Suicide Risk Assessment in Electronic Health Records: Corpus-Based Analysis," JMIR Medical Informatics, **9**(4), e22397, 2021. doi:10.2196/22397

[18] J. Zhang, C. Zhao, F. Xu, P. Zhang, "SVM-Based Sentiment Analysis Algorithm of Chinese Microblog Under Complex Sentence Pattern," in International Conference in Communications, Signal Processing, and Systems, 801–809, 2016.

[19] M.R. Islam, M.F. Zibran, "SentiStrength-SE: Exploiting domain specificity for improved sentiment analysis in software engineering text," Journal of Systems and Software, **145**, 125–146, 2018. Doi: 10.1007/s11042-019-7586

[20] M.K. Sohrabi, F. Hemmatian, "An efficient preprocessing method for supervised sentiment analysis by converting sentences to numerical vectors: a twitter case study," Multimedia Tools and Applications, **78**(17), 24863–24882, 2019. Doi: 10.1007/s11042-019-7586-4

[21] S. Jabri, A. Dahbi, T. Gadi, A. Bassir, "Ranking of text documents using TF-IDF weighting and association rules mining," in 2018 4th international conference on optimization and applications (ICOA), 1–6, 2018.

[22] A. Krouska, C. Troussas, M. Virvou, Deep learning for twitter sentiment analysis: the effect of pre-trained word embedding, Springer: 111–124, 2020. DOI: 10.1007/978-3-030-49724-8_5

[23] M. Ghio, K. Haegert, M.M. Vaghi, M. Tettamanti, "Sentential negation of abstract and concrete conceptual categories: a brain decoding multivariate pattern analysis study," Philosophical Transactions of the Royal Society B: Biological Sciences, **373**(1752), 20170124, 2018. Doi: 10.1098/rstb.2017.0124

[24] M. Xiao, Y. Guo, "Annotation projection-based representation learning for cross-lingual dependency parsing," in Proceedings of the Nineteenth Conference on Computational Natural Language Learning, 73–82, 2015.

[25] B.M. Hopkinson, A.C. King, D.P. Owen, M. Johnson-Roberson, M.H. Long, S.M. Bhandarkar, "Automated classification of three-dimensional reconstructions of coral reefs using convolutional neural networks," PloS One, **15**(3), e0230671, 2020.

[26] O. Appel, F. Chiclana, J. Carter, H. Fujita, "Cross-ratio uninorms as an effective aggregation mechanism in sentiment analysis," Knowledge-Based Systems, **124**, 16–22, 2017.

[27] A. Alsaeedi, M.Z. Khan, "A study on sentiment analysis techniques of Twitter data," International Journal of Advanced Computer Science and Applications, **10**(2), 361–374, 2019.

[28] F.-E. Lagrari, H. Ziyati, Y. El Kettani, "An efficient model of text categorization based on feature selection and random forests: case for business documents," in International Conference on Advanced Intelligent Systems for Sustainable Development, 465–476, 2018.

[29] M. Crawford, T.M. Khoshgoftaar, J.D. Prusa, A.N. Richter, H. Al Najada, Survey of review spam detection using machine learning techniques. J Big Data 2 (1): 23, 2015. Doi: 10.2200/S00416ED1V01504HLT016

[30] G.A. Ruz, P.A. Henr\'\iquez, A. Mascareño, "Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers," Future Generation Computer Systems, **106**, 92–104, 2020. Doi: 10.1016/j.future.2020.01.005

[31] Shoukry, A., & Rafea, A.," A hybrid approach for sentiment classification of Egyptian dialect tweets," 2015 First International Conference on Arabic Computational Linguistics (ACLing). https://doi.org/10.1109/acling.2015.18,(2015)

[32] L. Zhang, S. Wang, B. Liu, "Deep learning for sentiment analysis: A survey," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, **8**(4), e1253, 2018. Doi: 10.1098/rstb.2018.0124

[33] N. Alami, M. Meknassi, N. En-nahnahi, "Enhancing unsupervised neural networks based text summarization with word embedding and ensemble learning," Expert Systems with Applications, **123**, 195–211, 2019. 10.1016/j.systems.2019.01005

[34] F. Hill, K. Cho, A. Korhonen, "Learning distributed representations of sentences from unlabelled data," ArXiv Preprint ArXiv:1602.03483, 2016.

[35] E. Mansouri-Benssassi, J. Ye, "Synch-graph: Multisensory emotion recognition through neural synchrony via graph convolutional networks," in Proceedings of the AAAI Conference on Artificial Intelligence, 1351–1358, 2020.

[36] Z. Rahimi, M.M. Homayounpour, "TensSent: a tensor based sentimental word embedding method," Applied Intelligence, 1–16, 2021.

[37] M.F. Burg, S.A. Cadena, G.H. Denfield, E.Y. Walker, A.S. Tolias, M. Bethge, A.S. Ecker, "Learning divisive normalization in primary visual cortex," PLOS Computational Biology, **17**(6), e1009028, 2021. Doi: 10.1371/journal.pcbi.1009028

[38] S.Seker, E. Ayaz, E. Türkcan, "Elman's recurrent neural network applications to condition monitoring in nuclear power plant and rotating machinery," Engineering Applications of Artificial Intelligence, **16**(7–8), 647–656, 2003. Doi: 10.1016/j.engappai.2003.10.004

[39] L. Mou, Z. Jin, General Framework of Tree-Based Convolutional Neural Networks (TBCNNs), Springer: 37–40, 2018.

[40] D. Cazzato, M. Leo, C. Distante, H. Voos, "When i look into your eyes: A survey on computer vision contributions for human gaze estimation and tracking," Sensors, **20**(13), 3739, 2020. Doi: 10.3390/s20133739

[41] N. Capuano, L. Greco, P. Ritrovato, M. Vento, "Sentiment analysis for customer relationship management: An incremental learning approach," Applied Intelligence, **51**(6), 3339–3352, 2021.

[42] P. Tino, L. Benuskova, A. Sperduti, Artificial neural network models, Springer: 455–471, 2015.

[43] L. Lei, Y. Tan, K. Zheng, S. Liu, K. Zhang, X. Shen, "Deep reinforcement learning for autonomous internet of things: Model, applications and challenges," IEEE Communications Surveys \& Tutorials, **22**(3), 1722–1760, 2020.

[44] Z. Zhang, P. Cui, W. Zhu, "Deep learning on graphs: A survey," IEEE Transactions on Knowledge and Data Engineering, 2020.

[45] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative adversarial nets," Advances in Neural Information Processing Systems, **27**, 2014. Doi: 10.2200/S00204HLT016

[46] K. Pasupa, T.S.N. Ayutthaya, "Hybrid deep learning models for thai sentiment analysis," Cognitive Computation, 1–27, 2021.

[47] S.U. Hegde, A.S. Zaiba, Y. Nagaraju, others, "Hybrid CNN-LSTM Model with GloVe Word Vector for Sentiment Analysis on Football Specific Tweets," in 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), 1–8, 2021.

[48] J. Tang, M. Qu, Q. Mei, "Pte: Predictive text embedding through large-scale heterogeneous text networks," in Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, 1165–1174, 2015.

Predicting School Children Academic Performance Using Machine Learning Techniques

Radwan Qasrawi[1,*], Stephanny VicunaPolo[2] , Diala Abu Al-Halawa[3], Sameh Hallaq[4], Ziad Abdeen[3]

[1]*Department of Computer Science, Al Quds University, Jerusalem, 9103401, Palestine*

[2]*Center of Innovation Technology, Al Quds University, Jerusalem, 9103401, Palestine*

[3]*Faculty of Medicine, Al Quds University, Jerusalem, 9103401, Palestine*

[4]*Department of Economics, Al Quds University, Jerusalem, 9103401, Palestine*

A R T I C L E   I N F O

A B S T R A C T

The study aims to assess the machine learning techniques in predicting students' associated factors that affect their academic performance. The study sample consisted of 5084 middle and high school students between the ages of 10 and 17, attending public and UNRWA schools in the West Bank. The 'Health Behaviors School Children' questionnaire for the 2013-2014 academic year was used for data collection, and was then analyzed through machine learning techniques in order to evaluate their relationship with student academic outcomes. Six machine learning techniques (Random Forest, Neural Network, Support Vector Machine, Decision Tree, Naïve Bayes, and Logistic Regression) were used for prediction. The results indicated that the logistic regression and Naïve Bayes models had the highest accuracy levels (94.3%, 94%) respectively, followed by a decision tree, Neural Network, Random Forest, and Support Vector Machine (93.3%,91.9%,91.7%, and 80.2%) respectively. Thus, the Logistic Regression and Naïve Bayes had the best performance in classifying and predicting student academic performance with the associated factors. Furthermore, Decision Tree, Random Forest, and Neural Network had better predictive performance than Support Vector Machine. The results indicated that perception, Smoking, Depression, PTSD, Healthy Food Consumption, Age, gender, Grade Level, and Family income are the most important and significant factors that influence student academic performance. Overall, machine learning techniques prove efficient tools for identifying and predicting the features that influence student academic performance. The deployment of machine learning techniques within schools' information systems will facilitate the development of health prevention and intervention programs that will enhance students' academic performance.

1. Introduction

The prediction of students' academic performance is of great concern to educational institutions, governments, parents, and other stakeholders as it would enhance the development of curriculum, teaching methodologies, and planned interventions [1]. Several studies have evidenced that given the multi-causal nature of the concept of academic performance, a wide range of variables, from socio-economic to nutritional, health, and environmental, are directly correlated with it at all levels of education [2]. Good health, nutrition, and physical activity, for instance, are proven to enhance students' cognitive functions and therefore achieve better academic outcomes[3] [4].

Although the correlation between lifestyle habits and academic performance has been established [5], few studies have investigated the uses of Data Mining, namely Machine Learning Techniques among educational environments [6]-[10].

DM is used in various fields of research as it analyzes and extracts useful information and hidden patterns from large data

sets. The implications of DM, particularly ML techniques, in education are to provide new knowledge to help educational practitioners make predictions, informed decisions, as well as boost the shift of educational institutions from a traditional learning model towards automation[10]. Furthermore, ML techniques such as Clustering, Neural Networks (NNs), Decision Trees (DT), Random Forest (RTs), Support Vector Machine (SVMs), Logistic Regression (LRs), Naïve Bayes among others, have proven effective in predicting multicausal concepts such as academic performance [11]-[13].

2. Literature Review

2.1. Academic Performance-Associated Risk Factors

Abundant literature exists regarding Machine Learning and educational outcomes, although studies have found that certain variables do indeed significantly impact academic performance, there is no consistent agreement on a specific set of predictors or risk factors [14]. Nonetheless, some of the common variables yielding correlations include gender, parents occupation, family income, parents education, ethnicity, family size, mental health, post-traumatic stress disorder (PTSD), lifestyle habits, physical activity, and nutrition [3], [15].

In [16], the author demonstrates three key variable categories for improved academic performance: 1) Health behaviors, 2) Nutrition and Cognition, and 3) Obesity. The authors describe health behaviors as being influenced by risk associated factors such as physical activity, aerobic fitness, body mass, and energy consumption [13], [14]. Furthermore, relevant nutrition-associated risk factors, such as protein-energy consumption, breakfast consumption [19], short-term fasting, and sugar intake are among the key variables substantially influencing the development of brain functions.

In [10], the study provides a framework for researchers to use DM and ML models for the prediction of student academic performance. The guidelines consist of six stages: 1) data collection, 2) data initial preparation, 3) statistical analysis, 4) data preprocessing, 5) data mining implementation, and 6) result evaluation. The study has also found that early student performance prediction is not a widely studied area, yet it possesses the potential in aiding universities provide timely actions to improve students' academic outcomes. The study concluded that the most essential factors in predicting academic success are "prior-academic achievement, student demographics, e-learning activity, and psychosocial attributes" [10].

In [20], the authors aimed to predict middle and high-school students' academic achievement with a focus in Science and Arts classes by making use of the Chi-squared Automatic Interaction Detection (CHAID) model. The results yielded that students' performance in arts and science is strongly correlated to psychosocial features such as PTSD, self-reported health perception, age, and instances of depression [21]. However, variables such as good nutrition [15], family support, and gender weighed more on students' science performance. The study found that data mining, namely machine learning, possesses great potential for discovering hidden information that would otherwise

be difficult by using conventional analysis [20].

2.2. Machine Learning Techniques

Distinct studies have approached the analysis and prediction of academic achievement through a variety of methods. In [1], the authors review the ML techniques applied to the prediction of students' academic achievement through a study of published research, concluding the following:

- Most literature (around 70% of articles in their review) seeks to study student performance at the university level.
- The most widely used technique for predicting students' behavior is supervised learning.
- Support Vector Machine, DT, Naïves Bayes, and Random Forests are well-studied algorithmic proposals generating positive results.
- Neural Networks as less used, yet predict students' performance more accurately.
- Unsupervised learning has low accuracy in predicting students' behavior based on studies reviewed.

In [12], the authors seek to predict higher education student performance using the Random Forest Classification algorithm on a study of the UCI student performance dataset consisting of 8000 data. The study concludes that the RF algorithm predicts results that are 15% more accurate than the SVM algorithm.

The study in [22] analyzed the methodologies for the prediction of academic performance and the main variables impacting student outcomes. The study found that most experimental datasets focused on undergraduate Science, Technology, Engineering, and Mathematics (STEM) courses from a single educational entity. The majority of datasets were relatively small with a sample size of fewer than 1000 students. Furthermore, the prediction models were commonly developed as stand-alone modules instead of an ensemble technique despite it having better accuracy rates. Most ML techniques used are supervised through regression, neural networks, and tree-based models. Finally, the study concludes that the most effective predictive models are the Hybrid Random Forest, Feedforward 3-L Neural Network, and Naïve Bayes [22].

Seeking to introduce a prediction model for academic performance, the authors in [13] used supervised machine learning techniques like SVM and Logistic Regression. The results show that the Support Vector Machine algorithm yields more accurate results than logistic regression.

In [14], the authors employed a multi-dimensional methodological approach to predict first-year student academic performance at Babcock University with a sample size of 7,500 data. The study used classification algorithms such as Random Forest, Random Tree, REPtree, J48, OneR, Decision stump, JRip, ZeroR, Decision Table, PART, and Artificial Neural Networks. The study developed a Recommender System (RS) to enhance academic performance prediction. The results found that Random Forest, Reptree, J48, JRip, PART, Decision Table, and Artificial Neural Networks outperformed other classifiers with an accuracy rate of 96.78% (the lowest). Nonetheless, Random Tree, with an accuracy rate of 99.908%, proved ideal for this study's objectives [14].

In [9], the authors propose a prediction model for student performance based on a Decision Tree algorithm and through the application of an ensemble method on a UCI dataset for students in the Alentejo region of Portugal. The study used three supervised algorithms, J48, NN, and MLP. The results show that J48, with an accuracy rate of 96.78%, yields the most accurate results.

Thus, several studies have investigated academic performance from distinct perspectives, including teaching and educational factors, physical health, and educational environment [4], [5], [16]. The literature concludes that early prediction of risk factors affecting academic performance is essential in aiding struggling students in self-regulating the factors that can impact their cognitive development [3], [16], [19]. Nonetheless, in this study, we seek to assess machine learning models to predict students' academic status based on three levels of health factors (physical, mental, and social).

3. Research Methodology

In this research, we have used machine learning techniques for the classification and prediction of the influence of health factors on students' academic performance. Five machine learning models were utilized for data analysis. The methodology follows the guidelines outlined in [10] through data collection, data initial preparation, data analysis, data preprocessing, machine learning implementation of the models studied, and finally results evaluation through a confusion matrix and other visualization tools.

3.1. Data Collection

The data used for this research was collected from the National Health and School Behavior study conducted by Al-Quds University and the Ministry of Education (MoE) in the year 2013-2014. The data contains student's health, health behavior, and academic performance indicators. The survey was administered via the 'self-reporting' method of data collection while supervised by appointed school social workers. The research considers features that relate to the social, physical, and mental health of Palestinian children in school grades 5-9 (10-17 years). A representative sample of 5084 students was used in this study. The data includes associated risk factors variables related to mental health, physical health, social health, nutrition, and academic performance. The associated risk factors variables are listed in Table 1 below.

Table 1: Machine Learning Models' variables

Variable Name	Description	Value
Gender	Gender	Boys, Girls
Age	Age	Age (11-16 Years)
FAS	Economic Status	Low, Medium, High
ST	School Type	Public or Refugee
LP	Living Place	Urban or non-Urban)
FatherEdu	Father Education	≤ Secondary or > Secondary
MotherEdu	Mother Education	≤ Secondary or > Secondary
Healthy	Healthy Food Consumption	Low or High
Unhealthy	Unhealthy Food Consumption	Low or High
BMI	Body Mass Index	Normal, Overweight, or Obese
Anemia	Anemia hemoglobin blood test	Non-Anemic (+12g/dl), Anemic (<12 g/dl)
Smoking	Tobacco Risk	Yes, No
PA	Physical Activity	Low Active or Active
LA	Leisure Time Activity	Low Active or Active
FS	Family Support	Low, Medium, High
PS	Peer Support	Low, Medium, High
SS	School Support	Low, Medium, High
PTSD	Posttraumatic stress symptoms	Low, Moderate, Severe
Depression	Depression Symptoms	Low, Moderate, Severe
Psychosomatic_Symp	Psychosomatic symptoms	Low, Moderate, Severe
Positive_Health	Positive health perceptions	Low, Medium, High
Life_Satisfaction	Life Satisfaction	Low, Medium, High
Average_Score	Student Total Average Grades	Excellent/Very Good, Good, Weak/Fail

Socio-demographic variables: Include variables such as age, living place, gender, family income, parent's education, and school type.

Physical activity: Physical activity was measured in a two-scale system, the activity scale and the leisure scale. The activity scale was composed of three question items:
- In the last week, how many days were you physically active for more than 60 minutes
- Number of hours playing sports outside school
- Number of hours exercising per week

The leisure time scale similarly comprised three question items:
- Number of hours watching TV
- Number of hours playing video games
- Number of hours using the internet

All six items considered only weekdays (5 days a week). Respondents' answers were analyzed in quartiles for both scales. Physically active students were identified as those in the upper quartiles, and likewise, low and non-active students were identified as those in the upper quartile of the leisure scale.

Nutrition: Nutrition data were collected though and in-person self-reported 24h-food recall of one-day food intake, as well as through a food frequency questionnaire (FFQ). The food items were scaled by classifying them into 8 food categories (vegetables, fruits, dairy products, sweets, soft drinks, sweet beverages, and energy drinks) [23]. Response choices ranged from (1) never, (2) 1-2 times a week, (3) 3-4 times a week, to (4) 5-7 times a week (almost daily).

Furthermore, respondents were grouped into two classes based on consumption habits, healthy and unhealthy consumers. Respondents were classified into the *healthy* group if they formed part of the upper 2 quartiles and indicated they didn't consume unhealthy foods. Respondents were identified as *unhealthy* if they had indicated that they did not consume any of the healthy food groups, or if they were in the upper 2 quartiles on the unhealthy foods consumption frequency.

Anemia: Anemia was measured through the collection of blood samples for the assessment of hemoglobin levels. An instantly calibrated finger-prick was used for the collection of blood samples. Respondents with anemia were identified as those with a hemoglobin levels below 12 g dl^{-1} for children younger

than 14 years old, and those with levels below 13 g dl⁻¹ for respondents over 14 years of age.

Social Support: Social support was measured through three variables: 1) family support, 2) school support, and 3) peer support. Each variables asked to what extent respondents felt they were able to talk to their family, teachers, and peers.

Positive health perception: Self-reported health perception was measured through six question items:

- I like the way things are going for me
- my life is going well
- I have a good life
- I feel good about what is happening to me
- I would like to change many things in my life
- I wish I had a different kind of life—often or almost always.

Life satisfaction. The life satisfaction variable was measured through a 1-item scale based on the evaluative aspects of subjective well-being gathered from psychological research. The scale ranges from 0 to 10, where 0 represents the worst possible life, and 10 the best possible life satisfaction.

Posttraumatic stress disorder symptoms (PTSD)— PTSD symptoms were measured through 20- question items where respondents had to select their degree of association with the givens statement on a scale from 0 (not at all) to 4 (very much).

Academic performance: Respondents grades were obtained from prior school records, the average grades obtained in six courses (Arabic language, English language, Religion, Social studies, Science, and Mathematics) were collected and included in the total average estimation. Academic performance was considered by categorizing the average total score into Excellent/Very Good, Good, Weak/Fail.

3.2. Data Preparation

The raw data variables were selected, and cleaned the data prior to analysis. This step was taken to ensure the elimination of data redundancy, removed missing values, and derived new variables when applicable. The initial dataset consisted of 5084 student records. The key features in the dataset are balanced, whereas the gender variable is unbalanced.

3.3. Data Pre-Processing

The data has passed through a preprocessing phase before being ran in the machine learning algorithms. The data preprocessing includes techniques for organizing the data for machine learning models, such as data transformation, coding the categorical data, and feature selection, among others. The majority of the studied variables are composed of multiple questions, and thus need to be merged into a single variable with a preserved distance between the values. Preprocessing computes the variables from the scale of variables and converts them into meaningful numerical values. Common distance values were used depending on variable category levels: The three levels set as variable categories are 1) Low =0, 2) Moderate =1, and 3) High =2, and the two levels are set as 1) No =0, and 2) yes=1. The machine learning models are then used to analyze the students' academic achievement, as well as to predict the associated risk factors. The cross-validation technique that repetitively splits data into training and testing sets was used to overcome the problem of over fitting.

3.4. Statistical Analysis and Machine Learning Implementation

Six ML predictive models were built and compared to each other depending on their prediction accuracy from the given dataset. The ML algorithms used are described in the table below (Table 2).

Table 2: Description of Machine Learning Techniques

ML algorithm	Description
Artificial Neural Network (ANN or NN)	The Neural Networks algorithm is a computational model built up of many small aggregators that are clustered into groups, or layers that feed-forward from the input data to an output prediction [24]. Thus, the NN algorithm maps the predictor variables to an outcome variable through intermediate hidden layers.
Random Forest (RF)	Random Forest is an ensemble model that comprises decision trees, and then outputs the plurality of votes from the trees [25]. Each tree is exposed to a data subset and independently evaluates the features available to arrive in a conclusion [24]. RFs must be trained and are sensitive to the training data.
Support Vector Machine (SVM)	The SVM classifies linear and non-linear data; the model starts with a small number of the most similar cases across classes "support vectors", and the remaining cases are then classified based on their fall in the side of the support vectors [26].
Logistic Regression (LR)	Logistic regression is a statistical technique for classifying records based on values of input fields. LR is similar to linear regression, but it helps to find the probability that a certain item belongs in a specific class, thus the outcome lies between 0 and 1 [26].
Decision Tree (DT)	Decision trees are structures with multiple levels and a hierarchy of observations about an object in order to predict its class [24]. The tree "leaves" represent the classes the object could belong to and its "branches" reference the features that could lead the object to a certain class. DTs are used to predict and classify certain objects [27].
Naïve Bayes Classification	NB is analyzes training sets across the set of features and finds the probabilities of features' ability to predict the target variable. Naïve Bayes gives the probability of an event based on prior knowledge conditions without assuming that features are related even though they could be interdependent [26].

The machine learning algorithms were applied to predict the students' academic performance. The Orange data mining software [28] was used in testing and validating the machine learning models.

3.5. Results evaluation

The study found that Logistic Regression and Naïve Bayes had the highest accuracy rates (94.3% and 94% respectively), and are the most suitable machine-learning algorithms to predict student academic performance. The other algorithms similarly reported high accuracy rates and performed very well in predicting the associated factors related to academic performance. The results have been evaluated through a confusion matrix describing the performance of all machine learning models used (found under section 4.1- Table 4), as well as graphical representations of the data, and importance ranking of variables.

4. Results

A descriptive analysis was performed for describing the data distribution and for understanding the data set before the evaluation of machine learning techniques. The data set is composed of 5084 students with a mean age of 13±1.5, ranging from 10-17 years old. Among these students, approximately 32.8%

are boys and 67.2% are girls. Data in figure 1 shows the academic performance levels distributed by grade levels. The study found that the overall distribution of students' academic performance levels is almost similar in all grades. Results show that 46.7% of students obtained a low-average score, 20.6% obtained a medium-average score, and 32.7% obtained a high-average score.

Results in figure 2 show the average academic performance score by grade and gender. Results indicated that most of the students reported a good average score (70-75) in all grades for both girls and boys. We observed no significant difference found between the average performance score and grade or gender.

Figure 3 demonstrates the comparison between machine learning algorithms' accuracy rates for the algorithms used in predicting students' academic performance. Besides LR and NB models, which had the highest accuracy rates, the other ML algorithms had high-performance accuracy while SVM reported the lowest performance accuracy 80.2%. Logistic regression has 94.3% correct instances and 5.7% incorrect instances. For the further representation of prediction accuracy, the instances classification accuracy of the six models is shown in Table 2, while results in Table 4 show the algorithms' performance features through confusion matrices description.

Table 5 describes the details of the algorithms' performance. The classification results show that the Logistic regression and Naïve Bayes models are the best algorithms for predicting students' academic achievement. The logistic regression precision and area under curve (AUC) results (99.3% and 94.3% respectively) reflected in Table 4 indicate that most of the tested machine learning algorithms perform effectively in predicting positive values. Furthermore, the support vector machine algorithm reported the lowest acceptable performance, yet its significance is still valid for predicting students' academic performance.
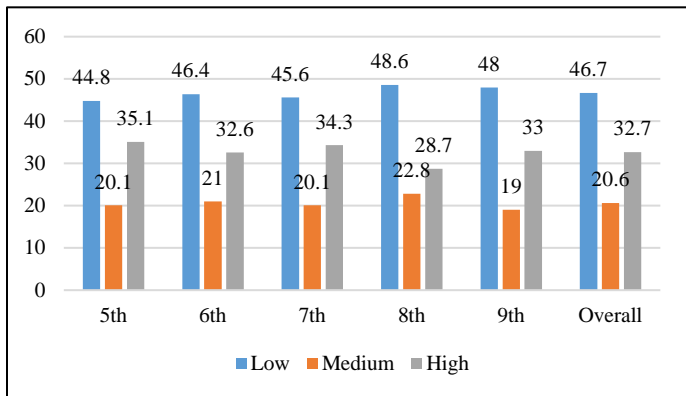
Figure 1: Students' academic performance by grade

The logistic regression ROC curve for the three academic performance classes is presented in figures 4.5, and 6. Three numerical categories for respondents' grades are included: 1) low, 2) medium, and 3) high. The ROC curve resides in the upper-left corner, thus the LR algorithm has a better prediction of positive value with an AUC of 99%. Furthermore, the Mathew Correlation Coefficient has been calculated for the 6 machine learning algorithms as shown in Table 3. Results indicated that the

different classifiers had constant and high-performance levels. Most of the models reported MCC values closed to one in the [-1,+1] interval, which means that the machine learning models performed very well in predicting the students' academic performance.
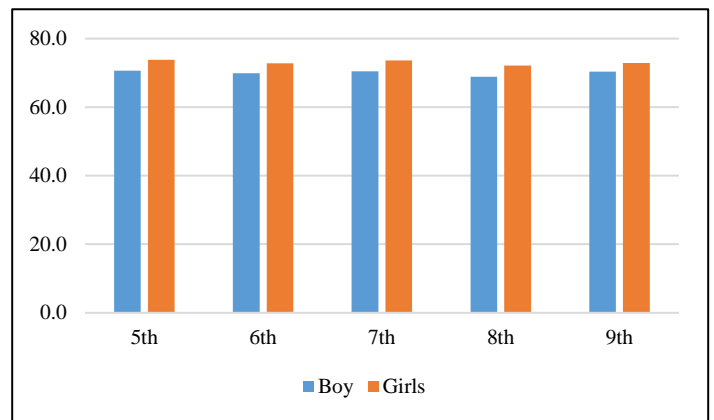
Figure 2: The students' average academic performance by grade and gender

Table 3: The estimated Mathew correlation Coefficient for the 6 ML models

Machine Learning Model	Mathew Correlation Coefficient
Logistic Regression	0.93
Naïve Bayes	0.927
Random Forest	0.92
Decision Tree	0.92
Neural Network	0.915
SVM	0.78
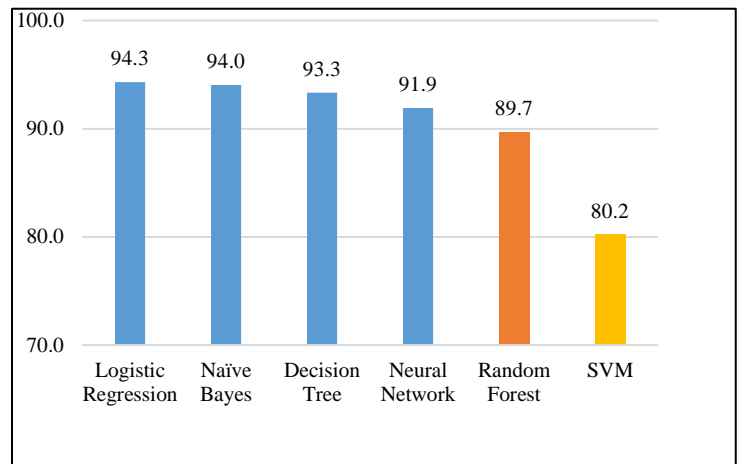
Figure 3: Comparison of Prediction Accuracy Percentage among the 6 ML Models

The machine learning models features importance ranking results are shown in Table 6. The ranking analysis includes Information Gain, Gain ratio, Gini and Chi-Square algorithms. Results in Table 6 indicated that positive health perception, Smoking, Depression, PTSD, Healthy Food Consumption, Age, gender, Grade Level, and Family income are the most important features for predicting students' academic performance.

Table 3: The confusion matrix that describes the performance of Machine Learning models.

Random Forest Predicted Value				
Actual Value	L	M	H	∑
L	94.2	5.5	0.5	**2523**
M	5.3	86.5	7.3	**1257**
H	0.5	8.0	92.3	**1304**
∑	**2604**	**1185**	**1295**	**5084**
Actual Value	Support Vector Machine Predicted Value			∑
L	88.3	21.5	0.0	**2523**
M	11.5	60.4	16.8	**1257**
H	0.2	18.1	83.2	**1304**
∑	**2555**	**1237**	**1292**	**5084**
Actual Value	Decision Tree Predicted Value			∑
L	94.3	4.7	0.0	**2523**
M	5.6	90.3	6.1	**1257**
H	0.0	5.0	93.9	**1304**
∑	**2618**	**1139**	**1327**	**5084**
Actual Value	Logistic Regression Predicted Value			∑
L	96.5	6.4	0.0	**2523**
M	3.5	88.8	4.6	**1257**
H	0.0	4.8	95.4	**1304**
∑	**2532**	**1249**	**1303**	**5084**
Actual Value	Neural Network Predicted Value			∑
L	95.5	8.1	0.0	**2523**
M	4.5	83.6	7.1	**1257**
H	0.0	8.3	92.9	**1304**
∑	**2535**	**1259**	**1290**	**5084**
Naïve Bayes Predicted Value				
L	95.8	6.1	0.0	**2523**
M	4.2	87.9	3.4	**1257**
H	0.1	6.0	96.6	**1304**
∑		**2554**	**1260**	**1270**	**5084**

Table 4: Performance of algorithmic models

Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.993	0.943	0.943	0.943	0.943
Naïve Bayes	0.975	0.940	0.940	0.941	0.940
Decision Tree	0.939	0.933	0.932	0.932	0.933
Neural Network	0.986	0.919	0.919	0.919	0.919
Random Forest	0.970	0.897	0.895	0.895	0.897
SVM	0.938	0.802	0.802	0.801	0.802

Figure 5: Logistic Regression ROC curve for moderate academic performance
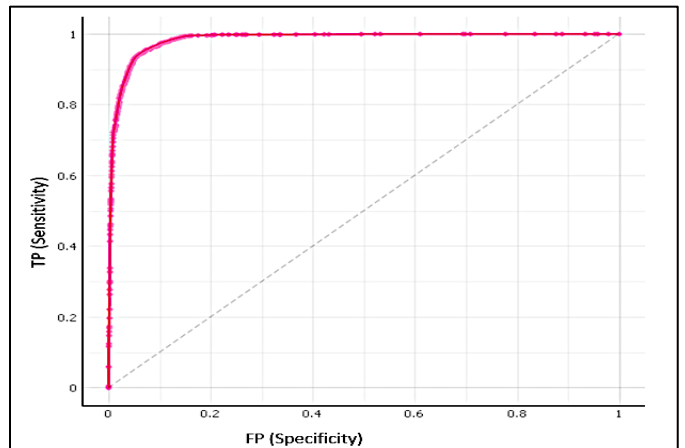
Figure 6. Logistic regression ROC curve for high academic performance

5. Discussion

The study aimed to predict the association between health factors (mental, physical, and social) with student academic performance. With this in mind, the results have identified that healthy students indeed perform better at school.

Twenty-two relevant features were utilized after performing feature selection using machine learning algorithms. These features were employed as input variables, while the average academic score was considered as the target variable. Among the tested models, logistic regression showed the best performance results with sensitivity =94.3%, specificity=96.3%, and AUC=99.3%. Out of the 22 features selected, 11 features are related to academic performance including physical, mental, and social health indicators.

The data has shown that positive health-associated factors are significantly interrelated with students' physical growth, cognitive development, and academic performance [7]. Furthermore, it has been found that unhealthy factors, such as obesity, depression, anxiety, and PTSD, have a direct negative impact on student's mental health and cognitive development,

Figure 4: Logistic Regression ROC curve for low academic

these findings were found to be consistent with other related studies [29], [30].

Table 5: Features' Importance Ranking

Features	Features Ranking Algorithms				
	Info. gain	Gain ratio	Gini	χ²	Relief
PTSD	0.040	0.025	0.020	43.8	0.006
Smoking	0.025	0.038	0.012	138.4	0.026
positive health perception	0.024	0.042	0.012	139.1	0.009
Depression	0.016	0.016	0.008	49.7	0.034
Healthy Food Consumption	0.010	0.010	0.005	30.6	0.029
Family Support	0.008	0.005	0.004	5.0	0.057
Age	0.006	0.003	0.003	27.9	0.011
Gender	0.006	0.006	0.003	12.9	0.010
Grade Level	0.004	0.002	0.002	16.8	0.013
Psychosomatic Symptom	0.004	0.002	0.002	5.7	0.015
Family Income	0.003	0.002	0.001	10.7	0.065
Mother Education	0.002	0.002	0.001	5.7	0.007
School Support	0.002	0.001	0.001	8.3	0.030
Residency	0.002	0.001	0.001	7.4	0.039
Unhealthy Consumption	0.001	0.001	0.001	2.7	0.012
Physical Activity	0.001	0.001	0.001	1.0	0.007
Anemia	0.001	0.001	0.000	5.5	0.003
Life Satisfaction	0.001	0.001	0.000	5.1	0.021
Peer Support	0.001	0.000	0.000	0.8	0.014
BMI	0.000	0.001	0.000	2.5	0.014
Leisure Time Activity	0.000	0.000	0.000	0.6	0.040
father Education	0.000	0.000	0.000	0.5	0.013

The prediction results indicated that the negative features had more effect on students with low academic performance. Students with low family income, smoking, anemia, depression, and unhealthy food consumption reported a significant association with low academic performance. Parents' education, life satisfaction, positive health perception, and physical inactivity had less significant effects on low academic performance than other features.

In the present paper, we evidence that several machine learning algorithms were able to predict academic performance associated with risk factors. All tested algorithms were successfully managed to predict the target variable. Nonetheless, we have found that the Logistic Regression and Naïve Bayes algorithms are the most effective in predicting academic performance when health (mental and physical) and social risk factors are considered. These findings strengthen previous findings that discuss the validity of data mining and ML techniques in predicting students' academic performance. Although the common ML model with highest accuracy in several literature is Naïve Bayes, we have found that when accounting for

only key subject areas and excluding commonly high-performance subjects (such as Arts and Physical Education), Logistic Regression is able to provide similar accuracy levels.

6. Conclusion

The study assessed the accuracy and performance of machine learning models in predicting the associated health factors on students' academic performance. Six ML models were used for this study (Random Forest, Logistic Regression, Neural Network, Support Vector Machine, Naïve Bayes, and Decision Tree). Although all models yielded valid and positive results, the logistic regression model proved to provide the best accuracy rate and performance, closely followed by the Naïve Bayes model. The study's findings demonstrate that other ML algorithms, such as Random Forest, Decision Tree, and Neural Networks are also able to satisfactorily predict students' academic performance, while support vector machine yielded the results with the least accuracy.

Furthermore, the obtained results show that the indicators that most affect academic performance include smoking, depression, PTSD, family support, anemia, leisure time, unhealthy food consumption, place of residence, and grade level. On the other hand, gender, physical activity, friends' support, school support, life satisfaction, mother's education, and positive health do not significantly influence students' academic performance.

Therefore, this research concludes that ML algorithms, particularly (yet not exclusively) Logistic Regression and Naïve Bayes, could be integrated into schools' information systems for the automatic prediction of students' academic performance based on key variables. In this manner, students, families, and school staff, and administration will be able to tackle issues that might affect students' performance through the use of the obtained prediction results. Likewise, through the use of accurate machine learning techniques, such as Random Forest, public health professionals, healthcare providers, and decision-makers will be able to predict rising issues, and implement relevant intervention programs to enhance students; health and education

7. Recommendations

Based on the findings presented in this paper, future research will benefit from expanding the study by adding additional associated factors, including cognitive development skills, in-school student behavior, social activities, and digital media activities. Further variables to consider that would enhance the aim of the research are factors related to the school environment and teaching methods, in particular those that are student-centered, motivational, and personalized as we believe they might positively influence academic performance.

Moreover, in a future study, we seek to deploy the above-mentioned machine learning algorithms by integrating them within schools' information management systems in Palestine to test the effectiveness of ML algorithms in a practical school environment. The results could aid researchers globally in developing frameworks for ML-led, automated school performance indicators.

Abbreviations

Data Mining (DM), Machine Learning (ML), Random Forest (RF), Artificial Neural Networks (ANN or NN), Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), Naives Bayes (NB), Post-Traumatic Stress Disorder, (PTSD), Chi-squared Automatic Interaction Detection (CHAID), Science, Technology, Engineering and Mathematics (STEM), United Nations Relief and Works Agency (UNRWA), Food Frequency Questionnaire (FFQ), Area Under Curve (AUC), Classification Accuracy (CA), Features (F), Intelligent Recommender System (IRS), Ministry of Education (MoE).

Conflict of Interest

The authors declare no conflict of interest

Acknowledgments

References

[1] J.L. Rastrollo-Guerrero, J.A. Gómez-Pulido, A. Durán-Domínguez, "Analyzing and predicting students' performance by means of machine learning: A review," Applied Sciences (Switzerland), **10**(3), 2020, doi:10.3390/app10031042.

[2] M.M. Dubuc, M. Aubertin-Leheudre, A.D. Karelis, "Relationship between interference control and working memory with academic performance in high school students: The Adolescent Student Academic Performance longitudinal study (ASAP)," Journal of Adolescence, **80**, 204–213, 2020, doi:10.1016/j.adolescence.2020.03.001.

[3] R. Qasrawi, "Links between nutrition, life style habits and academic acheivment in Palestinian schoolchildren: A cross-sectional study," Al-Quds Journal for Academic Research, **01**(1), 90–102, 2021, doi:10.47874/2021p6.

[4] E.L. Faught, J.P. Ekwaru, D. Gleddie, K.E. Storey, M. Asbridge, P.J. Veugelers, "The combined impact of diet, physical activity, sleep and screen time on academic achievement: A prospective study of elementary school students in Nova Scotia, Canada," International Journal of Behavioral Nutrition and Physical Activity, **14**(1), 1–13, 2017, doi:10.1186/s12966-017-0476-0.

[5] B. Asmare, M. Taddele, S. Berihun, F. Wagnew, "Nutritional status and correlation with academic performance among primary school children, northwest Ethiopia," BMC Research Notes, **11**(1), 1–6, 2018, doi:10.1186/s13104-018-3909-1.

[6] R.R. Halde, A. Deshpande, A. Mahajan, "Psychology assisted prediction of academic performance using machine learning," in 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 431–435, 2016, doi:10.1109/RTEICT.2016.7807857.

[7] H.F. Golino, C.M.A. Gomes, D. Andrade, "Predicting Academic Achievement of High-School Students Using Machine Learning," Psychology, **05**(18), 2046–2057, 2014, doi:10.4236/psych.2014.518207.

[8] V.L. Uskov, J.P. Bakken, A. Byerly, A. Shah, "Machine Learning-based Predictive Analytics of Student Academic Performance in STEM Education," in 2019 IEEE Global Engineering Education Conference (EDUCON), 1370–1376, 2019, doi:10.1109/EDUCON.2019.8725237.

[9] M. Imran, S. Latif, D. Mehmood, M.S. Shah, "Student academic performance prediction using supervised learning techniques," International Journal of Emerging Technologies in Learning, **14**(14), 92–104, 2019, doi:10.3991/ijet.v14i14.10310.

[10] E. Alyahyan, D. Düştegör, "Predicting academic success in higher education: literature review and best practices," International Journal of Educational

[11] A. Blanch, A. Aluja, "A regression tree of the aptitudes, personality, and academic performance relationship," Personality and Individual Differences, **54**(6), 703–708, 2013, doi:10.1016/j.paid.2012.11.032.

[12] A. Makwe, A.S. Rathore, An Empirical Study of Neural Network Hyperparameters, 2021, doi:10.1007/978-981-15-5788-0_36.

[13] D.I.F.S. Engr. Sana Bhutto, Dr. Qasim Ali Arain, Maleeha Anwar, "Through Supervised Machine Learning," Predicting Students' Academic Performance Through Supervised Machine Learning, 2020.

[14] M. Goga, S. Kuyoro, N. Goga, "A Recommender for Improving the Student Academic Performance," Procedia - Social and Behavioral Sciences, **180**(May), 1481–1488, 2015, doi:10.1016/j.sbspro.2015.02.296.

[15] T. Burrows, S. Goldman, K. Pursey, R. Lim, "Is there an association between dietary intake and academic achievement: a systematic review," Journal of Human Nutrition and Dietetics, **30**(2), 117–140, 2017, doi:10.1111/jhn.12407.

[16] T.M. Burkhalter, C.H. Hillman, "A narrative review of physical activity, nutrition, and obesity to cognition and scholastic performance across the human lifespan," Advances in Nutrition, **2**(2), 201–206, 2011, doi:10.3945/an.111.000331.

[17] I.K. Desai, A. V. Kurpad, V.R. Chomitz, T. Thomas, "Aerobic fitness, micronutrient status, and academic achievement in Indian school-aged children," PLoS ONE, **10**(3), 1–13, 2015, doi:10.1371/journal.pone.0122487.

[18] F.M. Asigbee, S.D. Whitney, C.E. Peterson, "The Link Between Nutrition and Physical," Journal of School Health, **88**(6), 407–415, 2018.

[19] K. Adolphus, C.L. Lawton, L. Dye, "The effects of breakfast on behaviour and academic performance in children and adolescents," Frontiers in Human Neuroscience, **7**(JUL), 1–28, 2013, doi:10.3389/fnhum.2013.00425.

[20] R. Qasrawi, Z. Abdeen, H. Taweel, M.A. Younis, D.A. Al-Halawa, "Data mining techniques in identifying factors associated with schoolchildren science and arts academic achievement," in Proceedings - 2020 International Conference on Promising Electronic Technologies, ICPET 2020, 2020, doi:10.1109/ICPET51420.2020.00023.

[21] P.A. Banerjee, "A systematic review of factors linked to poor academic performance of disadvantaged students in science and maths in schools," Cogent Education, **3**(1), 1–17, 2016, doi:10.1080/2331186X.2016.1178441.

[22] A. Namoun, A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," Applied Sciences (Switzerland), **11**(1), 1–28, 2021, doi:10.3390/app11010237.

[23] G.C. Frank, T.A. Nicklas, L.S. Webber, C. Major, J.F. Miller, G.S. Berenson, "A food frequency questionnaire for adolescents: defining eating patterns.," Journal of the American Dietetic Association, **92**(3), 313–318, 1992.

[24] H. Phwkrg, R.I. Frppxqlfdwlrq, W.S. Olyhv, "9 – 11 April , 2019 - American University in Dubai , Dubai , UAE 2019 IEEE Global Engineering Education Conference (EDUCON) Page 1370 9 – 11 April , 2019 - American University in Dubai , Dubai , UAE 2019 IEEE Global Engineering Education Conference (ED," 2019 IEEE Global Engineering Education Conference (EDUCON), 1370–1376, 2019.

[25] Y.L. Pavlov, "Random forests," Random Forests, 1–122, 2019, doi:10.1201/9780429469275-8.

[26] S. Uddin, A. Khan, M.E. Hossain, M.A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," BMC Medical Informatics and Decision Making, **19**(1), 1–16, 2019, doi:10.1186/s12911-019-1004-8.

[27] P. Geurts, A. Irrthum, L. Wehenkel, "Supervised learning with decision tree-based methods in computational and systems biology," Molecular BioSystems, **5**(12), 1593–1605, 2009, doi:10.1039/b907946g.

[28] J. Demšar, T. Curk, A. Erjavec, Č. Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, "Orange: data mining toolbox in Python," The Journal of Machine Learning Research, **14**(1), 2349–2353, 2013.

[29] C.A.C. Montanez, P. Fergus, A. Hussain, D. Al-Jumeily, B. Abdulaimma, J. Hind, N. Radi, "Machine learning approaches for the prediction of obesity using publicly available genetic profiles," Proceedings of the International Joint Conference on Neural Networks, **2017**-**May**, 2743–2750, 2017, doi:10.1109/IJCNN.2017.7966194.

[30] S. Roy, A. Garg, "Predicting academic performance of student using classification techniques," 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics, UPCON 2017, **2018**-**Janua**, 568–572, 2017, doi:10.1109/UPCON.2017.8251112.

SyncBIM: The Decision-Making BIM-Based Cloud Platform with Real-time Facial Recognition and Data Visualization

Chia-En Yang[*], Yang-Ting Shen, Shih-Hao Liao

National Cheng Kung University, architecture, Tainan, 701401, Taiwan

A R T I C L E I N F O	A B S T R A C T
	In this research we developed an BIM-based system to monitor and visualize the real-time building users information. Concentrating on building in-use stages, advantages in tracking facial recognition should be revealed through the availability of real-time information. In this way could explore the possibility of how BIM and IoT could improve data-oriented facility management. The integration system called SyncBIM. The five system elements of the construction platform are also proposed to allow more efficient management and data transmission of the building O&M system. This study integrates the three pieces of technology respectively known as BIM platform, Internet of Things, and computer vision to explore the architecture and technology required by the building O&M system, and establishes a "Decision-Making BIM-Based Cloud Platform" to allow the opportunities of information integration and collaborative operations for the application of BIM and the introduction of computer vision technology.

1. Introduction

The relevant definitions and scopes of Building Information Modelling (BIM) are constantly evolving over time. The American Institute of Architects (AIA) defines BIM as a multi-dimensional model presentation technology that adopts the database approach to integrate information of construction projects. In recent years, as the "Digital Twin" has been repeatedly regarded as one of the top ten global technology trends of the year by the information technology research organization [1], it could be understood that such trend is highly connected with BIM as it integrates multiple proposed applications in the technological and related fields.

In addition to implementing strategies in compliance with government policies, the BIM TASK GROUP, established under the promotion of the UK government, has its major goal of promoting the various duties of the building production and construction supply chain. However, after the government agencies having formulated various collaborative work specifications, starting from April 2016, it could be foreseen that building information models with highly collaborative functions would even become one of the basic elements of future architectures [2].

BIM plays an important role in all stages of the building life cycle, making it infrastructure of the future as one of the trends.

Because BIM not only serves as a beneficial tool or database in the construction phase, but also provides various flexibility and integration functions that can offer various requirements for building O&M, and exerts powerful application integration capabilities(see Figure 1).
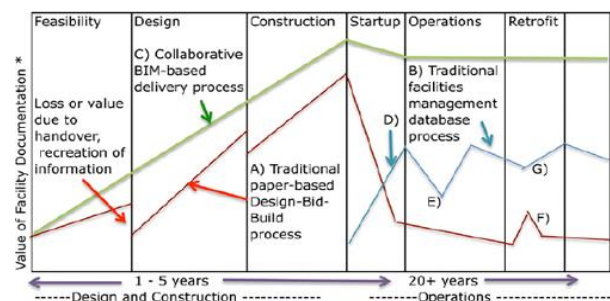
Figure 1: The value of documentation in the facility lifecycle
(Source: Eastman, 2008)

For the current application of building O&M, the importance of facility management (FM) and the improvement of building efficiency and functionality in facility management have gradually become prominent. Nowadays, increasingly complex building functions are making the building space and equipment more diversified and sophisticated. Building management is no longer simply checking whether it is operating normally, but finding out its unique usability based on the huge volume of building usage

information generated, real-time spatial analysis, and retrieval. The changeable[3] complexity and real-time nature of modern computer technology can assist decision-making and provide optimized applied artificial intelligence (AI). For a building space involving a wide range of levels, the integration of software and hardware and the deployment of various systems, such as environmental data, the number of users, and the analysis of crowds, etc., would have a certain degree of crucial impact on building O&M. In response to such impact, this study believes that the usage status of these buildings must be actively displayed through machine learning and the application of AI technology, and such relationship of changes requires to be reflected in the O&M process. In order to solve the problem of BIM in the O&M of building facilities, this study proposes the "BIM management platform for decision-making", which is hereby referred to as "SyncBIM"[4].

2. Related Work

In this Section, we present two popular aspects of adopting BIM in the building life cycle, which are Digital Twin and AI applications. The first lays the foundation for the cloud platform architecture, and diversified AI applications drive BIM to become the "sense" of architectures.

2.1. Application of Digital Twin to BIM

AEC includes buildings&architectures, engineering, and construction industries. At present, BIM technology has reached certain achievements under the promotion of government agencies of various countries. Meanwhile, with the evolution of information and communication technology, more application technologies such as computer vision, AI, and various sensors and environmental sensing devices such as LiDAR could be able to complement the possibility of collecting real-time and on-site data in a real environment reality, which would allow the BIM model to predict and plan the allocation of resources and transportation, and supervise the safety of the site environment. The introduction of Digital Twin at an earlier stage can further enhance the application of BIM advantages of the whole building life cycle.

An indicative example of a Digital Twin application is the "Virtual Singapore" developed by the French software company Dassault Systèmes for five years. On this virtual three-dimensional platform, one can freely browse all corners of Singapore City, regardless of the data of buildings, roads, public infrastructure, or the natural environment[5]. More importantly, it provides four main functions, namely "Virtual Experimentation", "Virtual Test-Bedding", "Urban Planning and Decision-Making", and "Research and Development" that allow decision-makers of urban affairs to predict the impact of different decisions and policy directions on the city. Therefore, there would be wider application in the fields of urban disaster prevention such as logistics and environmental conservation, etc.

Figure 2 shows the UI of Virtual Singapore. With real-time information and the large volume of data provided by sectors from public departments and communities, Virtual Singapore is able to solve the problem of livability in areas with limited resources and unlimited potential. For example, after importing annual data of insolation areas, temperatures, and humidity, the system would be

able to provide space decision-makers with the information for making a more comfortable and humanized environment.

Figure 2: Interface of Virtual Singapore (Source: nrf.gov.sg)

From the perspective of the "smart city", infrastructure is the element that composes a city. The system developed by Dassault Systèmes allows people to see the digitization of urban infrastructure, which is constructed in the digital domain of computers and adds value to or optimizes the decision-making for the city to shape the vision of smart city life of the future[6]. The BIM model plays the role of an information carrier, and after importing the information, it serves as a bridge between the real and virtual worlds. In the next stage, the health and safety issues at the construction site, which has the widest impact and draws the most attention during the construction phase, will be further reviewed and discussed in the scope and application focused on by this study, and the development of the research proposed by this study will be applied to the connection of the construction site to make trigger more innovative and progressive linkage[7].

2.2. Application of Artificial Intelligence in the Building Life Cycle

The current major technology of artificial intelligence (AI) aims at analyzing and realizing professional human techniques of various fields through extensive computer science training [8]. With its quality management and supervision functions, AI could replace a large amount of work previously requiring human judgment, which would achieve accurate and effective predictive capability [9]. For example, for a construction site facing the pressure from both construction schedule and cost while requiring the high integration of the construction specifications and interface of various fields, the introduction of AI and BIM technology for the integration of various information would provide a certain degree of efficiency improvement and added value[10].

According to the research conducted by the Science & Technology Policy Research and Information Center of the National Applied Research Laboratories, The assistance provided by AI to the construction industry can be roughly divided into four categories as shown in Table 1, which are respectively known as (1) Planning and Design, (2) Safety, (3) Automation Equipment, and (4) Monitoring and Maintenance.

The following four AI application areas proposed by the Science & Technology Policy Research and Information Center of the National Applied Research Laboratories are all able to be highly connected with BIM:

- Planning and Design

Table 1: Application of AI in the Construction Industry (Source: Science & Technology Policy Research and Information Center of the National Applied Research Laboratories, 2018)

Application of AI in the Construction Industry	
Categories	Contents
Planning and Design	The BIM model uses AI to automatically complete pipeline configuration, analysis, and machine learning to manage construction risks
Safety	AI with sound and image recognition system to automatically mark the construction site
Automation Equipment	Smart building management to make users feel comfortable
Monitoring and Maintenance	AI with robots and drones to improve construction efficiency

The construction site requires to consider structural, electrical and mechanical, and pipeline engineering issues, etc., which involve different contractors using different appliances and interfaces that may often cause delays and repetitive work in the final integration stage. Therefore, Building System Planning developed GenMEP by using BIM modeling software Autodesk Revit. By adopting Generative Design, the configuration of the electrical system in the building model can be automatically built using the GenMEP plug-in while also taking into account the design specifications and the spatial complexity of different buildings (see Figure 3).

Figure 3: Interface of GenMEP (Source: BuildingSP, Inc.)

• Safety

AI technology such as data analysis and machine learning would be adopted to manage construction risks. BIM 360 Project IQ launched by Autodesk is an example in which the BIM 360 Field is shaped by various images, audio, and other construction management information collected on site, which is analyzed to provide suggestions and reduce risks for relevant QA staff on site (see Figure 4).

• Automation Equipment

Jetson AI platform constructed by Nvidia and the Japanese company Komatsu is designed to perform 3D visualization and analysis of real-time interactive modes such as tracking the on-site staff, machinery, and objects, which could allow AI to take over tasks of on-site monitoring and management to provide round-the-

clock protection for site safety, as well as improving the quality of construction management.
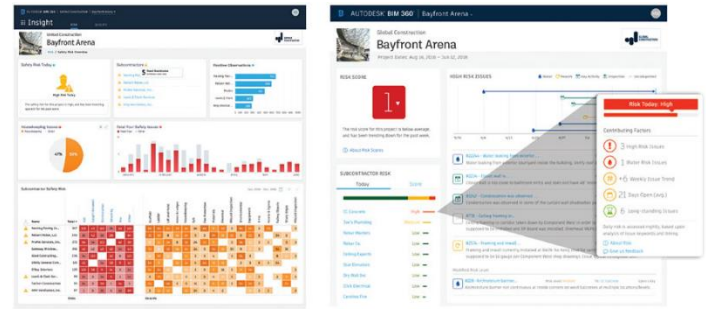
Figure 4: Interface of BIM 360 Project IQ (Source: BIM360)

• Monitoring and Maintenance

Suffolk Construction based in Boston adopts a virtual model constructed by VR to monitor the progress of on-site construction projects to cope with the rising construction costs and on-site safety management, as well as improving construction efficiency. In addition, Figure 5 shows that Doxel AI from Silicon Valley combines exploration robots and drones with photography equipment and LiDAR to scan and construct the real-time situation of the site, and adopts visualization data and deep learning algorithm processing to allow users and managers to automatically eliminate various errors and monitoring risks at the construction site by using the on-site construction progress analysis adopting AI technology.
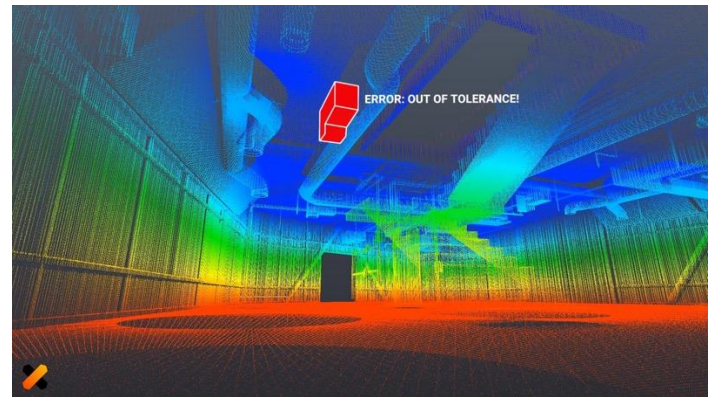
Figure 5: Visualization of On-site Construction Information by Doxel (Source: Doxel AI)

The key points and precautions of the occupational safety inspection of each working item are clearly-defined specifications, for example, the specifications for the categorization of hazardous areas and the key points of inspection. Hence, AI could serve as a 24/7 supervisor to check items such as whether safety equipment is worn and any staff is entering the hazardous areas by adopting the image recognition technology to conduct round-the-clock monitoring at the construction site. For example, the setting of geofencing allows personnel entering the hazardous areas to be adequately alerted and protected while notifications are also provided to have managers aware of the current presence of high-level risks of the staff (see Figure 6), and the real-time information could also be fed back to the Digital Twin's BIM environment integration platform and presented in a visualized manner. With

the integration of BIM and various AI techniques, the required time and cost for the construction site could be improved in terms of management efficiency, and the occupational safety and working environment for labors could also be assured, which would achieve a win-win situation[11], [12].
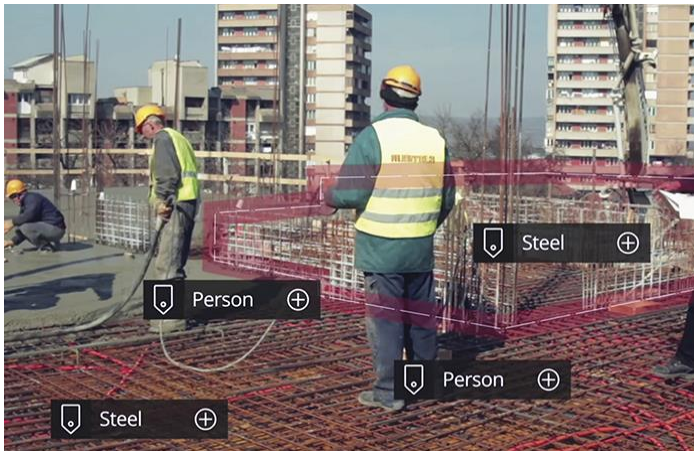
Figure 6: Construction Site Management Adopting AI Image Recognition Technology (Source: Smartvid.io)

The definition and application of BIM often evolve along with various technological advancements. And the pursuit of environmental friendliness and green living has always been its core value. Building O&M and management is a complex interface integration and multiple interdisciplinary coordination. However, when carrying out various intimate and highly customized O&M processes for future space users, a large number of people would be exposed to a high degree of on-site O&M risks for achieving the establishment of a smart city or its infrastructures. If the managers and participants could improve the building O&M mode in a safer and more efficient way, it will be more complete and environmentally friendly when establishing an overall smart city, which would also be a precious gift to the fields of building O&M.

3. The Planning and Design of Decision-Making BIM-Based Cloud Platform

This Section mainly explores the system adopting BIM in the auxiliary decision-making platform. It would discuss the information processing flow and architecture of the smart building O&M system in two parts respectively known as the information transmission architecture the system operates, and the information transmission loop of the system.

3.1. Information Transmission Architecture of the System

Figure 7 describe such information transmission architecture. The architecture itself adopts the IPO model as the application basis, and in the process of data transmission, a web portal is established between the endpoints in order to make the clearly defined division of labor of the system. The scalability and integration of the web portal also provide flexibility for the equipment and sensing devices installed in each period of the system's operation. And it is also helpful for the integration of information formats. For the transmission architecture specified in this Section, we will further describe it in detail in the following

three subsections respectively titled (1) Input, (2) Process, and (3) Output.
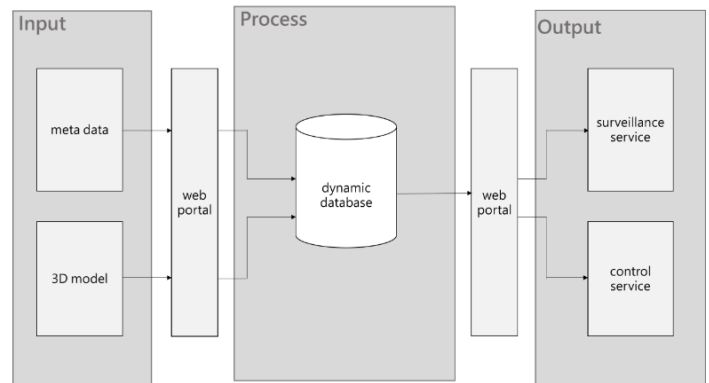
Figure 7: Diagram of Information Transmission Architecture for the Smart O&M System

- System Input: Information Collection

The information sources the smart building O&M requires could be summarized as three scales. One is the meta data that covers various real parameters and O&M mode scripts, which transmit the three-dimensional information model to the web portal and subsequently to the dynamic database for storage, integration, and further process. The meta data includes the real parameters, the scripts and modes related to various building O&M. If it is necessary to perform appropriate O&M for a certain building, the data monitoring operation would start the deployment independently without waiting for the completion of the system installation, and the scaling of input data monitoring could be performed even under the presence of an existing system which requires only the addition of new web portals. If the building usage data required by the system is transmitted to a specific web portal, such input would be designed to operate independently, and the system only retrieves the data from the web portal to achieve the division of labor and facilitate any subsequent scaling. Furthermore, upon the occurrence of equipment failure or update, the system would not be completely shut down but only some of the functions are temporarily suspended, which would also improve efficiency in troubleshooting for the system.

- System Process: Dynamic Database

Because the various O&M data and 3D models from the input would be updated continuously over time, a dynamic database that can be updated and accessed in real time shall be established for the process. Such a dynamic database for the process would retrieve the information of the web portal, then the system would analyze and compare the usage mode. Because the information carried by the web portal has been standardized, it is sufficient for the database to complete the data connection with the web portal while allowing the system to retain the scalability, which would be helpful for the maintenance and testing of some areas of the system or the addition of O&M areas.

- System Output: Provision of Services

The dynamic database compares and delivers the O&M data to the web portal of the output. Unlike the input web portal, the output web portal saves various service-related information which would be used for building O&M. Rather than being the big data, such

types of information required for O&M is integrated information, such as the parameter codes after script writing or the trigger values of execution, which are provided to the application services. The types of application services are divided into two types respectively known as "control type" and "service type". The output of the information transmission architecture would trigger related services according to different types of parameters.

3.2. Information Transmission Loop

In the adaptive O&M system, the transmission of information is not a straight-linear process but a continuous act of constant updating and modification. Therefore, a loop of information transmission shall be established to complete the architecture required for the O&M maintenance under the constant-changing relationship. When the data of the usage pattern is established in the dynamic database and compared with the analysis characteristics, these data can be used as a basis for system adaptation, learning, and reflective practice. Also, because both people and the built environment would continue to change and influence each other over time, the usage data will continue to be imported to the database, and then the updated results would be transmitted to the end sensors, which forms an information transmission loop accordingly. The three processes of this loop formation are respectively known as (1) Data Collection, (2) Database Comparison, and (3) Database Update. Figure 8 presents the data transmission loop formed by the system.
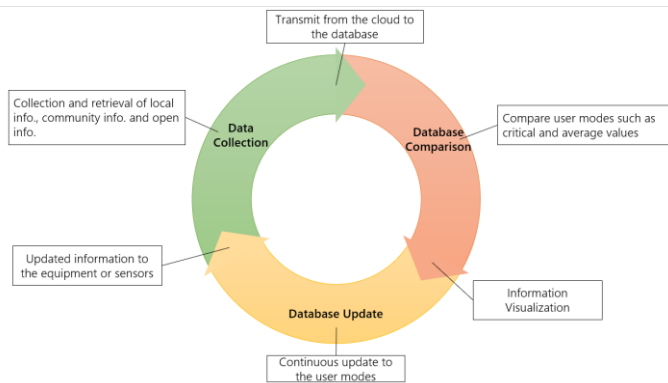
Figure 8: System Data Transmission Loop of SyncBIM

- Data Collection

After completing the layout of the sensor network for collecting building spatial data, the data via the cloud would start to be transmitted to the designated dynamic database. The types of data refer to the open information, local information, and community information mentioned in the previous paragraph. Such three types of information would be compared and analyzed within the database.

- Database Comparison

When the volume of data within the database is sufficient, it would be available for further analysis to find the average value, critical value, or usage habits of the particular space for the establishment of a spatial usage pattern. The usage pattern upon being determined established could be subject to continuous updates besides being compared with the data. By analyzing the dynamic pattern of the data, one could infer whether there is any device functioning abnormally or abnormal spatial activities, and

if so, the information visualization would display the location and time where such abnormality takes place to allow the recording by the system and update to the user mode of the database.

- Database Update

With the increase of usage data and the modification of user mode, the operation of the system will keep all usage information updated along with the building's existence period and usage history, and update its information to various building equipment and sensors, which also means that the database would continue to be updated over time to continuously adjust the balance between people and the building environment.

4. SyncBIM that Integrates Computer Vision Applications

This Section presents the construction of a SyncBIM system based on the system architecture mentioned in the previous Section. It would discuss in two parts respectively titled "The System Interface Planning", and "The Process of Integrating Computer Vision Applications".

4.1. The System Interface planning of SyncBIM

The SyncBIM system is established based on the following five principles respectively known as (1) 3D Information Model Aiding Decision-making, (2) Object Orientation Design Delivering Service, (3) Dynamic Database based on Real-time Monitoring, (4) Mapping Data-driven Visualized Results, and (5) Hierarchical Integration of Contextualized Spatial Information. (see Figure 9).

- 3D Information Model Aiding Decision-making:

The spatial information presented in a three-dimensional manner allows viewers to view in an intuitive way, while also being able to efficiently be aware of the information and contextualized spatial data carried by the objects.

- Object Orientation Design Delivering Service:

The BIM model constructed with individual objects not only uses object-oriented design to guide information to the building objects, but also treats each object as a node that can provide the required services, and synchronously connect the functions and services required for the operation of the O&M system.

- Dynamic Database based on Real-time Monitoring:

When the spatial information from various usage data in real buildings is collected, the system could continuously transmit the parameterized data to the database. In addition to using the database as the source of review for system learning and reflection, such a database would also become a dynamic database for real-time display of building usage.

- Mapping Data-driven Visualized Results:

With the import of real-time data, users would be able to customize the colorized levels for visual alters and the threshold for event trigger with the system's logical decision-making function, which would allow the decision-making process more adaptable.

- Hierarchical Integration of Contextualized Spatial Information:

The system would be presented in an adaptive hierarchical manner for different scales and service targets. In addition to allowing precise deployment of various services, users could also perceive the contextualized spatial relationship.
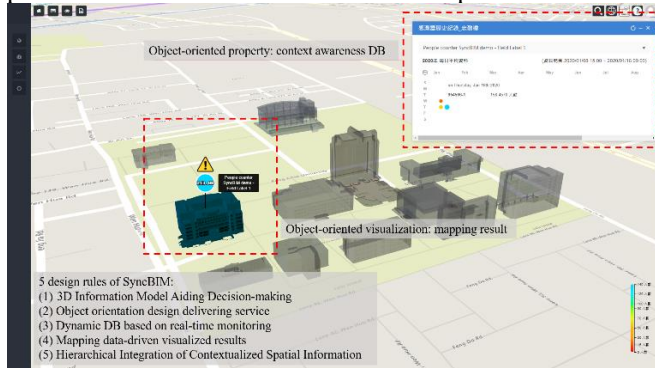
Figure 9: System Interface of SyncBIM

4.2. Importing SyncBIM into MATLAB for Computer Vision

MATLAB is software designed for currently commonly available algorithms, which is mainly used for data analysis and development of algorithms and also image processing and computer vision. The IoT platform "ThingSpeak" owned by MATLAB can transmit the analyzed results and parameters to the IoT platform, and further serve as a scalable modular web portal for the dynamic database of SyncBIM. The API could be used to access the data in the designated channel of the IoT platform (see Figure 10).

```
[initialNum,latestDate] = thingSpeakRead(954596,'ReadKey','QOX784947******');
if dateshift(latestDate,'start','day') < datetime('today','TimeZone','local')
    logThingSpeakData('reset')
end
```

Figure 10: ThingSpeak API

The designated IoT device would be set to ensure that the information is free of overlay (see Figure 11). This step would allow the continuous recording of the on-site data in the IoT platform and transmission to the SyncBIM database.

```
vidObj = webcam('STARTEC 1.3MP Webcam','Resolution','640x480');
faceDetector = vision.CascadeObjectDetector('MaxSize',[150 150]);
tracker = MultiObjectTrackerKLT;
if ~isempty(initialNum)
    tracker.NextId = initialNum(1) + 1;
end
```

Figure 11: IoT Device and Object Tracker

The timestamp would be generated for the collected (Figure 12), and such data would be continuously uploaded to the IoT channel shown in Figure 10. The timestamp of the data will be able to provide a future database as a basis for the decision-making of spatial events and situations.

```
tim = timer;
tim.ExecutionMode = 'FixedRate';
tim.Period = 20;
tim.TimerFcn = @(x,y) logThingSpeakData(tracker);
tim.StartDelay = 5;
end
```

Figure 12: Timer for Logging Data

After running such a program in the IoT device, it can continue to perform image recognition and facial recognition (Figure 13). It would not only calculate the number of people, but also record the time point, upload the data to the designated IoT platform, and continue to be available for retrieval by the SyncBIM System database. Such a device has scalability, which allows it to be installed in different spatial architecture and scaled as required. For the system itself, only the specified channel data retrieval settings are required to allow the division of labor for O&M for achieving a higher level of efficiency.
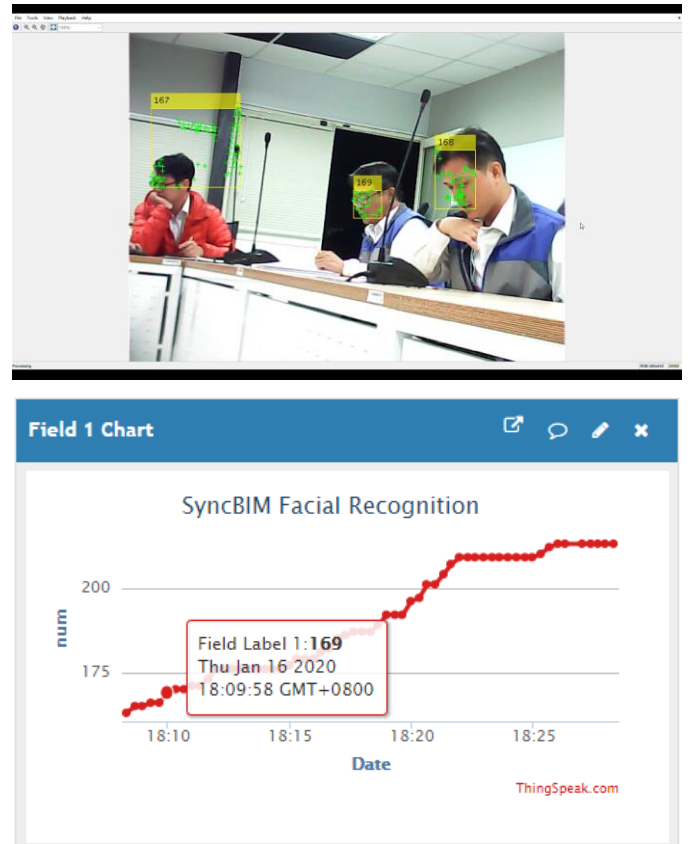
Figure 13: Top: Image Recognition Screen; Bottom: Data Collection Channel of IoT Platform

The ThingSpeak IoT platform would serve as a web portal to save the collected real-time data, while SyncBIM retrieves the data in the channel and continuously updates the dynamic database. The flexible architecture of the system allows the user preferences and spatial characteristics to be visualized for display (see Figure 14). The number and time data of user facial recognition collected in this study, combined with the building spatial information in BIM, could provide system operators with the required actions for interpreting the O&M situation, and even further adjusting the types of data collection such as various real-time information like air quality, temperature, and humidity. The adjustment for software and hardware would be feasible from simple video lenses to temperature detection cameras, and even computer vision algorithms could be modified according to events, such as detecting whether the person is wearing a mask (which is one of the major safety issues for spaces under the current global pandemic of COVID-19 and more application scenarios and services will become necessary). The spatial information and information integration in SyncBIM is the most appropriate

platform to provide these derivative services. In the future building environment, it can be foreseen that such type of cloud service platform will become a necessary asset, even as a part of infrastructures. For O&M, it can also effectively improve efficiency and further enhance the mutually friendly relationship between people and the environment.
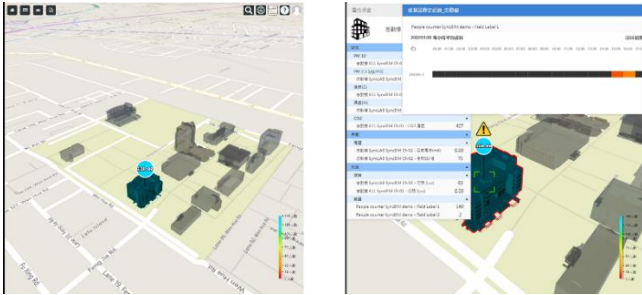
Figure 14: The system displays the facial recognition data screen. Left: The viewing panel of the 3D models. Right: The viewing panel of detailed object information, such as the time and current data of each sensor.

5. Conclusion

This study proposes a cloud platform that integrates computer vision and building information models, which is called SyncBIM system, it provides the monitoring and information visualization functions required for building O&M, and improves the passive and equipment troubleshooting-oriented building maintenance modes. In addition to being used for auxiliary O&M decision-making purposes, it can also immediately reflect the O&M state most suitable for the very moment and distinguish the relationship between the user and the spatial context, which is also a major feature of the SyncBIM system. By recording and reflecting the changes in the spatial context and usage situation to provide various instant feedback information required for O&M of buildings, it is expected that to use the SyncBIM system as the best driver for the common prosperity between people and the environment and the pursuit of well-being in the increasingly complex and multi-faceted building O&M context.

References

[1] Q.Qi, F.Tao, Y.Zuo, D.Z.-P.Cirp, undefined2018, "Digital twin service towards smart manufacturing," Elsevier, Sep. 2021.

[2] A.Matthews, B.Ta, "Applying a National BIM Model to Vietnam's National Implementation of BIM: Lessons Learned from the UK-Vietnam Collaboration for the Industry," Lecture Notes in Civil Engineering, **54**, 57–66, 2020, doi:10.1007/978-981-15-0802-8_6.

[3] Y.T.Shen, T.Y.Wu, "Sync-BIM: The interactive BIM-based platform for controlling data-driven kinetic façade," Communications in Computer and Information Science, **618**, 445–450, 2016, doi:10.1007/978-3-319-40542-1_72.

[4] C.E.Yang, Y.T.Shen, S.H.Liao, "Integration of Real-time Facial Recognition and Visualization in BIM," 2nd IEEE Eurasia Conference on IOT, Communication and Engineering 2020, ECICE 2020, 243–245, 2020, doi:10.1109/ECICE50847.2020.9301999.

[5] J.Kim, H.Kim, Y.Ham, "Mapping Local Vulnerabilities into a 3D City Model through Social Sensing and the CAVE System toward Digital Twin City," Computing in Civil Engineering 2019: Smart Cities, Sustainability, and Resilience - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2019, 451–458, 2019, doi:10.1061/9780784482445.058.

[6] F.Tao, M.Zhang, A.Nee, "Digital twin driven smart manufacturing," 2019.

[7] I.Din, M.Guizani, J.Rodrigues, …S.H.-F.G., undefined2019, "Machine learning in the Internet of Things: Designed techniques for smart cities," Elsevier, Sep. 2021.

[8] J.Chen, T.Tseng, …C.L.-2012 9th I., undefined2012, "An intelligent virtual fence security system for the detection of people invading," Ieeexplore.Ieee.Org, 2012, doi:10.1109/UIC-ATC.2012.64.

[9] M.Despotovic, D.Koch, S.Leiber, …M.D.-E. and, undefined2019, "Prediction and analysis of heating energy demand for detached houses by computer vision," Elsevier, Sep. 2021.

[10] E.I.Brunberg, K.E.Bøe, K.M.Sørheim, "Testing a new virtual fencing system on sheep," Acta Agriculturae Scandinavica A: Animal Sciences, **65**(3–4), 168–175, 2015, doi:10.1080/09064702.2015.1128478.

[11] B.Akinci, M.Fischer, J.Kunz, R.Levitt, "Representing Work Spaces Generically in Construction Method Models," Journal of Construction Engineering and Management, **128**(4), 296–305, 2002, doi:10.1061/(ASCE)0733-9364(2002)128:4(296).

[12] B.Akinci, M.Fischer, J.Kunz, "Automated Generation of Work Spaces Required by Construction Activities," Journal of Construction Engineering and Management, **128**(4), 306–315, 2002, doi:10.1061/(ASCE)0733-9364(2002)128:4(306).

Cyber Incident Handling and the Perceptions of Learners on Cyber Incidents in South African Schools

Naume Sonhera[*,1], Elmarie Kritzinger[2], Marianne Loock[2]

[1]Information and Communication Technology, Vaal University of Technology, Vanderbijlpark, 1911, South Africa

[2]School of computing, College of Science, Engineering & Technology, University of South Africa, Florida, 1709, South Africa

A R T I C L E I N F O	A B S T R A C T
	With increases in technological usage, cyber incidents are also on the rise and have become a major concern in schools across the globe. What is of significant concern is that cyber incidents in South African schools are also on the rise. Existing evidence suggests that, in South Africa there are no clear procedures that are consistently followed by schools on how to report cyber incidents. The aim of this research is therefore to propose cyber incident handling procedures to enhance the effectiveness in handling cyber incidents as well as ensuring that each role player has an important contribution in the intervention process that is designed to reduce cyber incidents in South African schools. The study also assessed the perception of learners on cyber incidents in South Africa. Using the literature review approach and thematic analysis of the data collected from learners the study highlighted the procedures and roles of role players that can assist in cyber incident handling in South African schools. The study also came up with a detailed analysis of the views of learners on cyber incidents in South Africa. The results presented can help to provide a framework that will act as a guide on reporting cyber incidents and directing school management, and all within the school, towards appropriate reporting procedures and intervention processes. The study also found out that the rise in cyber incidents in South African schools, if left unaddressed, can have a devastating effect on learners. Therefore, the government of South Africa, through the Department of Basic Education, must prioritize the handling of cyber incidents in schools as cyber incidents are now a threat to the efficient and effective execution of the mandate of the department.

1. Introduction

The 21st century came with changes in almost all the sectors of the world economies. In the education sector, both the academics and practitioners of the 21st century have applauded the paradigm shift in schools of encouraging learners to be computer literate [1-4]. The technological paradigm shift has been complemented by the efforts of schools through promoting access to technology in education [5-6]. The socialization patterns among learners have also changed drastically because of the growth and proliferation of electronic communication devices. Learners access social network sites to communicate with their friends at any time of the day or night [7]. However, access to new technology has led to an increase in misuse and abuse of technology by learners and this has brought about incidents of threatening, harassing, embarrassing, and humiliating behaviour and actions online [8]. Cyber incidents

have consequently become prevalent in schools worldwide, and it is evident that they present certain unique problems for schools in the regulation thereof [9-12]. What is of significant concern is the fact that cyber incidents in South African schools are also on the rise [13].

South African schools are adopting Information and Communication Technologies (ICTs) as part of an educational and social approach to prepare learners for the future. However, there is a lack of supporting cyber incident handling procedures to empower schools to deal with cyber incidents [14]. Cyber incidents are increasing among learners and yet there is very little guidance for schools on how to deal with cyber incidents [15]. With the potential for cyber incidents and their negative effects, schools have the responsibility to intervene in the occurrences [16]. Cyber incidents do not have an impact only on the physical and mental health of learners but also on the academic performance which in turn influences learners' learning outcomes negatively.

[*]Corresponding Author: Naume Sonhera, Email: nqume@vut.ac.za

Therefore, there is a need to ensure that schools are provided with cyber incident handling procedures that can contribute to a safe and caring environment for learners and all involved role players. Far from being limited to cyber victims, the effects of cyber incidents extend to the learners collectively, the school environment, and the entire school system [17]. To address the elements of cyber incidents in South African schools and to narrow the focus of this research to an educational environment, a cyber incident has been defined as: An act that violates explicit or implied cyber safety, computer safety or acceptable use of ICTs policies of South African schools or the Department of Basic Education (DBE) [18]. It is an event that involves the intentional or negligent use of digital platforms or electronic media to cause harm to other learners/self, or to impact on their confidentiality and integrity.

Some studies were done to try and understand the critical issue of cyber safety. In [14], the author investigated the maturity levels of cyber safety in South African schools. The study was intended to assess whether schools are prepared to help learners and educators to come up with cyber safety culture within the school environment. The study involved 24 schools in South Africa, and the investigation was through a critical evaluation of the elements that are important in improving cyber safety in schools. The elements in [14] which were evaluated include leadership and policies, infrastructure, education and standards, and inspection. After measuring cyber safety successfully using one of the UK approved measurement tool, the 360safe tool, the study discovered that all the schools had low levels of cyber safety maturity and compliance. In [14], the author further discovered that there is a clear lack of cyber safety awareness policies, practices, and procedures in the schools in South Africa. The study went on to propose a step-by-step cyber safety plan to empower schools to create their own cyber safety culture.

In [19], the authors stated that cyber safety is one of the critical aspects in the world today. Their study highlighted that learners require education on how to operate safely in cyber space and to protect themselves. In [19], the authors developed a cyber safety curriculum through open educational resources. The authors believed that this curriculum can empower educators in schools when educating learners about cyber safety. In [20], the author argued that business schools can meet the multidisciplinary needs of training cyber safety professionals. Motivated by these findings the study aims at proposing cyber incident handling procedures to enhance the effectiveness in handling cyber incidents as well as ensuring that each role player has an important contribution in the intervention process that is designed to reduce cyber incidents in South African schools. The study also assesses the views of learners on cyber incidents in schools.

2. The Education Sector in South Africa

The education sector in South Africa is divided into two departments, the Department of Basic Education (DBE) and the Department of Higher Education and Training (DHET) [21-22]. The DBE is responsible for primary and secondary education while the DHET is responsible directly for tertiary education and vocational training [23]. These two departments were once represented as a single department of education before 2009, then there was a split. The DBE looks after the public schools, the

private schools, Early Childhood Development centres (ECD) and the special needs schools. Private schools are also known as independent schools and these private schools and public schools are known collectively as the ordinary schools [22]. On the other hand, the DHET deals with tertiary education where we have further education and training, popularly known as the Technical and Vocational Education Training colleges (TVET) [21]. The DHET is also responsible for Adult Basic Education and Training (ABET) Centres and Higher Education Institutions (HEI) [24]. In South Africa, each province has its education department which is responsible for the implementation of policies of the national department. These departments are also there to deal with local issues that are related to the execution of the mandate of the department at national level.

As of 2016, the DHE reported that there were 29 749 public and registered independent schools in South Africa [23]. The number of ordinary schools was 25 574 while 4 175 were other educational institutions such as special schools and Early Childhood Development (ECD) centres. The DBE also reported that the total number of educators were 425 000 in 2012 and 440 151 in 2016 who were reporting to the national department and the provincial department in the 9 provinces and the 86 districts of South Africa [22]. District and provincial DBE offices in nine provinces and 86 districts administer all the schools and have considerable influence over the implementation of policy. The goals of the DBE and its district and provincial offices are to improve the quality of teaching, undertake regular assessments, and improve on the Early Childhood Development, to ensure a system of outcomes-focused accountability [23].

3. Literature Review

Schools worldwide are being charged to take action to reduce cyber incidents because of the potential of cyber incidents disrupting the educational processes, creating a hostile environment, and threatening learners' feelings of safety and mental well-being [8, 12, 25]. Due to the increase of cyber incidents that cross over into both home and school environments, coupled with confusion as to who should issue consequences and how, role players must work together to create effective cyber aggression strategies [16]. The intervention during cyber incidents takes group effort from all role players and learners themselves. While it may take a village to raise a child, the virtual village inclusive of social media platforms, Internet Service Providers, phone companies, non-profit youth organisations, non-governmental organizations, educational institutions, parents or guardians, attorneys, and media need to join forces to have dialogue nationally and globally on ways to protect learners from cyber incidents [26].

The 21st century has brought with it a different revolution - learners who are on the cutting edge of technological proficiency [7]. While awareness of the use of the Internet is growing among learners in South African schools, there is no increase in awareness of safe practices in the use of ICTs [27]. Learners receive mixed messages with regard to online behaviour as they strive for technology literacy, sometimes without appropriate support [28]. Learners seem unaware of the risks of inappropriate behaviour online, viewing them as trivial. Cyber incidents are causing major challenges for school officials who are called upon to respond to

cyber incidents involving learners [29]. Online threats take place off the radar screen of educators and parents; this makes it difficult to detect cyber incidents in schools and more impossible to monitor off school premises [10]. The language that the learners use on the Internet has evolved and created a generation gap between learners and adults around them. This gap enables learners to participate in cyber incidents without fear of being discovered by adults [15, 30].

There are no clear procedures that are followed consistently by South African schools, governing boards, and educators on how to report or handle cyber incidents [10, 31]. At present, the principals, educators, and parents are not sure what procedures to follow when learners are being harassed online [32-33]. Research in South Africa (SA) that examines the procedures for schools on how to report cyber incidents are limited [10]. The lack of clear procedures for reporting cyber incidents in schools, makes educators feel unsupported. Educators will rather ignore these unethical violations than follow ill-defined and unenforceable policies [34]. In [34], the authors investigated the reason learners should be educated about risks in cyber space and the ways in which the schools can intervene to assist. The study discovered that despite the usefulness of the Internet there are still negative issues that are related directly to the Internet use. Users of the Internet across the world are being subjected to various forms of abuse such as cyber incidents, cyber bullying, online fraud, racial abuse, pornography, and gambling. In [34], the authors believe that issues related to cyber incidents are on the rise, especially among learners, because of a lack of anti-cyber incident awareness. The other argument put forward by them was that the level of awareness among the Internet users is low or moderate. As a result, in [34], the authors stated that it is important that the anti-cyber incidents awareness be cultivated among learners through various initiatives like education so that learners are able to operate safely in cyber space.

In another study, in [35], the authors verified the effectiveness of gamification in reducing cyber incidents and offer a constructive application of a framework that can help to improve cyber safety skills and capabilities for learners in schools. The study also discovered that the rise in cyber incidents shows that the traditional methodologies used to train and bring awareness are insufficient to build the necessary cyber safety skills and capabilities. The study also discovered that gamification can be a promising methodology that can bring some changes to the behaviour of learners in cyber space at early stages. In [35], the authors came up with a proposal for a cyber hero framework for information security awareness and training programs that can fight human error in cyber space. In [35], the authors discovered massive progress in the skills and capabilities of cyber safety for learners using the cyber hero framework.

In [36], the authors discussed the applicability of the safety standards and frameworks used in various industries to avoid threats of cyber incidents due to the rise in the use of technology in schools. In [36], the authors discovered that it is very critical to have strong cyber safety controls in schools due to the growth in cyber incidents. The other notable finding in [36] was that it is very important to have a cyber safety framework that accommodates safety standards and frameworks in an education environment. In [15], the authors investigated the perceptions of cyberbullying

among student teachers in South African schools, in Eastern Cape. In [15], the authors found out that in South Africa cyber incidents are some of the issues affecting learners. The study went on to discover that the Department of Basic Education in South Africa is providing little guidance for schools on issues related to cyber incidents. Using a quantitative survey approach, the study discovered that cyber incidents are some of the critical issues in South African schools, but the topic has not been incorporated in the policy or even in the curriculum. In [15], the authors recommended that the Department of Basic Education should come up with a standardised policy that can be used to implement and enforce cyber safety behaviour in schools.

In [37], the authors did an investigation to assess the policies, procedures and measures which schools need to have in place to ensure that there is cyber safety awareness for learners in South African schools. They discovered that Information Communication Technology is now becoming part of the daily lives of learners, opening various opportunities and cyber risks as well. Therefore, just like in [15], [37], the authors believe that it is important for cyber users to become aware of the cyber safety issues so that they are able to protect themselves. It was also highlighted that in South Africa, cyber safety awareness and cyber education is still lacking, especially in schools. In [37], the authors also believe that it is of utmost importance for South Africa to begin the creation and implementation of cyber safety and cyber incident awareness policies, procedures, measures, and initiatives for the education sector, especially for schools and learners.

4. Research Methodology and Design

The research methodology is explaining how the literature review was done and how data was collected from learners to get their experiences and perceptions on cyber incidents. This research consists of two research methods. The first one is the literature review. The researcher conducted a systematic search on literature that reported the prevalence and impact of cyber incidents among learners in South African schools and the extent to which role players responded to cyber incidents in schools. The main idea behind the reviews was to identify the roles and responsibilities of role players in cyber incident handling for the Schools in South Africa. The systematic process included the electronic search in databases, title and abstract review and hand searches in Google scholar and article reference lists. Online databases contained bibliographic references to academic and peer-reviewed journal articles, as well as references to theses, books, and chapters. Secondly, the research used a qualitative approach and purposive sampling to collect data from the learners to get their experiences and perceptions on cyber incidents [38-40]. The rationale for selecting learners was based on reviews in [41-44], with the authors who indicated that adolescence is a peak period for involvement in cyber incidents. A total of 85 learners participated in three focus groups that ranged in size: group one had twenty participants, group two had thirty participants and group three had thirty-five participants as shown in the table1.

Forty-eight percent were male, and fifty-two percent were female. The interview questions enabled learners to explain their perspectives in depth. The questions were related to the themes that emerged from each group. All the conversations were recorded and transcribed. The data was then coded using ATLAS.ti. 7; themes

emerged, and the frequencies of these themes were noted by the researcher. Several ethical guidelines were followed to ensure that the research study was conducted ethically. In an endeavor to ensure the reliability and validity of data, the verbatim transcribed interviews were presented to the respondents to verify and sign off. The researcher reviewed all the transcripts from focus group interviews to check for accuracy, ensuring that no obvious mistakes appeared. The researcher also ensured that there was no drift in the definition of codes or shift in the meaning of codes during the coding process.

Table 1: Participants in the Three Focus Groups.

Focus Groups	Participates in the Groups
Group 1	20
Group 2	30
Group 3	35
Total	85

Source: Author's Analysis

Table 1 above shows the number of students who participated in each focus group.

5. Discussion of the Roles and Responsibilities of Role Players in Cyber Incident handling for the Schools in South Africa

The study came up with a theoretical framework that outlines the roles and responsibilities of role players in addressing the cyber incident handling needs. The responsibilities of role players within schools and relevant role players outside schools for specific and specialized interventions and support were identified during the literature review. To elaborate on the "Needs for Addressing Cyber Incidents", each need is linked to role players and recommended responsibilities. The aim of the theoretical framework and the roles of role players for schools in South Africa is to ensure that relevant role players are drawn in to create a supportive structure that can contribute to the development of the cyber incident handling procedures. To achieve this, the researcher highlighted the roles and responsibilities of role players within and outside the school when responding to cyber incidents. The framework is designed to complement normal schooling duties and activities and to integrate cyber school safety into the daily activities of the school. The framework also gives a guide on how relevant role players can assist in responding to cyber incidents. The framework may not make provision for all cyber incidents in schools or cyber safety concerns for schools but is a starting point to address cyber incidents in South African schools. The linking of "needs for addressing cyber incident" to "the role players" assisted the researcher to present the complex ideas and the theoretical concepts, from the literature review to a new structure which will assist in developing reporting procedures [41]. Table 2 is outlining the roles and responsibilities of role players in cyber handling in South Africa as discovered from the reviewed literature.

Table 2: Summary of the Roles and Responsibilities of Role Players in Cyber handling in South Africa as discovered from the Reviewed Literature.

Role Players	Responsibilities
School Principals, Learners, Parents, Guardians, school staff members, and School Governing Body	Role players should be competent in handling cyber incidents.
	All role players should be informed of the cyber incident reporting procedures.
	School procedures should be readily accessible to learners, parents, guardians, and school staff members.
	Need clear and easily understandable school procedures to address cyber incidents.
	Consistent, universal, and effective ICT policy aimed directly at regulating cyber incidents in schools, should be developed.
	Codes of conduct or code of behaviour should be formulated and adopted.
	Need a referral system to appropriate services and community partnerships to support learners and build a cyber school safe environment.

Source: Author's Analysis

Full explanation of the roles and responsibilities of role players in cyber incident handling in South Africa as discovered from the reviewed literature is given in the following section.

5.1. Role Players should be Competent in Handling Cyber Incidents.

The principal should ensure that school personnel and parents are trained in handling cyber incidents. The advisory Team should implement an effective approach in intervening and responding to all cyber incidents. Educators should participate in school cyber safety trainings. Educators should be aware of the steps to take and advice to give if learners notify them of cyber incidents. Parents should report their children's cyber aggression. This was in line with the findings that were discovered in [15] and [37], the authors discovered that it is important for cyber users to become aware of the cyber safety issues so that they will be able to protect themselves from the various forms of cyber incidents. It was also highlighted that in South Africa, cyber safety awareness and cyber education is still lacking, especially in the education sector [15]. In [36], the authors argued that it is very important to have a cyber safety framework that accommodates safety standards and frameworks in an educational environment.

"All Role Players should be informed of the Cyber Incident Reporting Procedures"

The school community should know who to approach if they become aware of or suspect any cyber incident taking place. The advisory team should establish a process that communicates the cyber incident, reporting procedures to all parents, learners, and all staff members. The School Management Board (SMB), School Governing Body (SGB) and Counsellors should explain the cyber incident reporting procedures to learners and their parents or

guardians. Educators should establish consultation and training for parents and other staff members on cyber incident reporting procedures. Social Media Service Providers (SMSP) should make their educational and safety materials easy to find and promote the materials to users regularly. Social Media Service Providers should improve the visibility, consistency, and accessibility of reporting tools on their platforms. Several studies support this. In [15] and [31], the authors pointed out that in South Africa there are no clear procedures that are followed consistently by schools, governing boards, and educators, on how to report or handle cyber incidents. The authors believe that it is very important for role players to be aware of the reporting procedures because at the moment, the principals, educators, and parents are not sure of what procedures to follow when learners are being harassed online [10, 32]. In [35], the authors emphasized that role players should be informed of the cyber incident reporting procedures especially if these frameworks are to be usable.

5.2. School Procedures should be readily Accessible to Learners, Parents, Guardians, and School Staff Members.

Provincial education departments must ensure that all schools are trained about procedures to address cyber incidents. The principal, Information Technology (IT) Unit, SGB, SMB and the Advisory Team should develop, support, and evaluate procedures to address cyber incidents. Officials from the National Department of Education, in collaboration with provincial, regional and district officials responsible for ICT school safety, should monitor the implementation of procedures to ensure that learners are safe from cyber incidents. Educators should assist in the development and implementation of anti-cyber incident procedures according to delegated roles and responsibilities. This has been supported in [37] by the authors who argued that for South Africa to have meaningful progress towards addressing cyber incidents in schools, the departments of education must be equipped with the appropriate policies, procedures, and measures to ensure that there is cyber safety awareness for leaners in South African schools. In [37], the authors discovered that Information Communication Technology is now becoming part of the daily lives of learners, with various cyber opportunities and cyber risks as well, hence the need for proper policies, education, and training for those responsible for cybersecurity.

5.3. Need Clear and Easily Understandable school procedures to address cyber incidents.

The other important aspect is that procedures outlining the school response to cyber incidents and how to report cyber incidents, should be in place and these procedures should be easily understandable. Provincial Departments of Education should ensure that school cyber incident reporting procedures are in place and should be adhered to strictly in schools. School cyber incident reports must be submitted to the district after which reports must be consolidated and forwarded to the Provincial School Cyber Safety Coordinator. A consolidated report from provinces must be submitted every quarter to the national Department of Basic Education (DBE). This was also supported in [14], [15], and [37] by the authors who agreed with many other scholars. These authors believe that it is important for learners to be aware of the cyber safety issues and the cyber incident reporting procedures so that they can get help whenever they need it. It was also highlighted

that in South Africa, cyber safety awareness and cyber incident handling education is still lacking, especially in schools [14, 15, 37]. Scholars also believe that it is of utmost importance for South Africa to begin the creation and implementation of cyber safety and anti-cyber incident awareness policies, procedures, measures, and initiatives for the education sector, schools and learners in the country [14, 15, 37].

5.4. Consistent, Universal, and Effective ICT Policy aimed at Regulating Cyber Aggression directly in schools, should be developed.

The other critical aspect is that the Department of Basic Education should take a lead to ensure that ICT policies for provinces have been developed. Provincial officials must provide direction to their regional or district counterparts to ensure that the development, enforcing and reviewing of ICT policies is being implemented in schools. Public policy should also be informed by ongoing research on the prevalence of cyber incidents. A school ICT policy that includes an ICT Acceptable Use Policy (AUP), ethical use of ICTs or digital technologies and anti-cyber incidents should be developed, implemented, and enforced. All role players should be aware of the contents of these policies. The principal, IT Unit, SGB, SMB and the Advisory Team should ensure that they have an appropriate ICT school policy in place to safeguard the overall cyber safety and wellbeing of learners. The IT Unit should manage the ICT policy through recording and monitoring Internet use for compliance with ICT policies. The SMB and the SGB should convene regular meetings to review and update the ICT Policy. Parents should ensure that their children adhere to the ICT policy.

5.5. Codes of Conduct or Code of Behaviour should be Formulated and Adopted.

The information gathered also reveals that the provincial education departments should ensure that all schools have a Cyber Safety Code of Conduct for learners. The principal, SGB, SMB and the Advisory Team should develop, support, and evaluate the Cyber Safety Codes of Conduct for learners. Learners should adhere to the Code of Conduct and to ensure appropriate behaviour always. All schools should be obliged to comply with Child Protection legislation, which includes Internet safety for learners. DBE needs to develop and integrate cyber safety programs into existing curricula to support human rights and child protection. The Child Protection and Abuse Organization should be contacted for help if a cyber incident is a suspected child protection issue.

5.6. Need a Referral System for appropriate Services and Community Partnerships to support Learners and build a Cyber School Safe Environment.

The South African Police Service (SAPS) should be obliged to investigate any unlawful cyber incident in schools. SAPS should advise and educate learners, parents, and school personnel about legal responses to serious cyber incidents. Law enforcement authorities must ensure that the all-role players have a clear awareness and understanding of how existing criminal offences can be applied to cyber aggression behaviour. Learners should contact SAPS in case of any cyber incident. Counsellors should give immediate support to all affected learners. Parents should

partner with schools to ensure that their children are following appropriate guidelines for online behaviour. Parents of the victims should be able to work with the Internet Service Provider, Cell Phone Service Provider, or Content Provider to investigate the cyber incident or to remove the offending material. Internet Service Providers should track instant messaging which could have been used as evidence in a court of law. SA Depression and Anxiety Group (SADAG) should assist learners suffering emotionally because of the cyber incident ordeal. Educational Organizations should continue to hold workshops, do research, and give presentations on cyber incidents and cyber safety topics to help the school communities. The attorneys should be able to provide parents with sound legal advice on how to open a possible criminal case and how to get restraining orders against the accused.

5.7. Findings of the Focus Group Interview

The interviews revealed learners' perspectives about cyber incidents and were specific to cyber incidents as they manifested within the schools. The results indicated the negative cyber incident effects on victims, the no or ineffective consequences for aggressors, and the inability to stop or control cyber incidents once they have started. The themes which emerged are Learners' Perceptions about Victims, Aggressors and Bystanders; Learners' Perceptions about Technology; Learners' Perceptions about Cyber Incidents, and Learners' Perceptions about Adults. A collection of the conceptual valid and reliable evidence contributed to a better understanding of learners' interactions in cyber space and how much help they need when they are in trouble. The results of the focus group interview also showed that learners are acutely aware of cyber incidents occurring in their schools and how they are reluctant to report the experiences of cyber incidents when they occur. The researcher recognized that cyber aggression, especially if left unaddressed, can have a devastating effect on learners; it can create a barrier to learning and have serious consequences for mental wellbeing. The views of learners are summarized in the table below.

Table 3: The Findings from the Focus Groups on the Views of Learners about Cyber Incidents

Findings From Focus Groups	Description
	Learners' Perceptions about Victims, Aggressors and Bystander
	Learners' Perceptions about Technology
	Learners' Perceptions about Cyber incidents
	Learners' Perceptions about Adults

Source: Author's Analysis

The table above is summarizing the views of learners on issues related to cyber incidents and the full description of these views is given below.

5.8. Learners' Perceptions about Victims, Aggressors and Bystanders

The themes generated from the focus group interviews were linked to the suggestions from learners and from the literature reviewed. Pillars for reporting procedures were derived from the whole integration process. Themes from Learners' Focus Group Interviews indicated the Challenges for Victims, Aggressors' Behaviour, and Learners' Perceptions about Bystanders Supporting Aggressors. The results revealed that the challenges for cyber victims are that cyber victims are humiliated and harassed online. Also, learners were of the view that victims fear aggressors and are affected psychologically which will lead them to suffer in silence. On the behaviour of aggressors, the results revealed that aggressors gain power by controlling victims and most of them are mean; they seek attention, are insecure, and they take out their pain or anger on other learners. The results also revealed that technology allows the aggressors to hide behind screens and in many circumstances, they are perceived as mentally unstable. The study went on to check the learner's perception about bystanders supporting aggressors. The study found out that usually bystanders are afraid to become the next victims and to suffer rejection by peers and sometimes uncertainty about whom to tell.

In summary, the results pointed to the fact that cyber victims feel like outcasts, are hurt, offended, embarrassed and feel inferior. These victims fear aggressors because aggressors take out their pain or anger on other learners. The results also pointed out that bystanders support aggressors because of fear of becoming the next victims, being rejected by peers or uncertain of whom to report to. The identified pillars for reporting procedures derived from the discussions are Legislation, Referral System, Cyber Incident Reporting Platform and Cyber Awareness. Learners suggested that all schools should be compelled to comply with Child Protection Legislation. There is a need for a referral system to appropriate services, and need for community partnerships to support learners. Learners should report cyber incidents to the principal or ICT personnel, school staff members should provide support and counselling, school psychologists should help affected learners, school assemblies and public service announcements should be used for awareness. Learners need to be educated about cyber incident problems, there should be a hotline for reporting cyber incidents and finally there must be consequences for the aggressors.

5.9. Learners' Perceptions about Technology

Leaners are of the view that technology aggravates cyber incidents, and, in many ways, technology has decreased empathy among learners. The results indicated that aggressors use profiles of innocent people to threaten other learners. The anonymity online makes threatening behaviour more readily achievable and protects aggressors. The other aspect that was highlighted is that technology enables aggressors to have the biggest audience, making it difficult for those who want to champion anti-cyber incident awareness. Learners also indicated that due to technology some learners put themselves at risk by taking naked videos of themselves and these videos end up going viral on social media. In a nutshell, the results from the interviews have revealed that technology has decreased empathy for victims due to space differences, instant gratification and communication with the biggest audience. As a result, learners are of the view that schools need to have a clear ICT policy, procedures, and codes of conduct in place to regulate cyber incidents among learners. Parents and learners should be given the sets of ICTs rules so as to guide learners against cyber incidents. Learners are also of the view that

schools should introduce education against cyber incidents and independent service providers should assist to solve cyber incident problems, for example, the network companies should be able to identify phones that could have sent the harassing messages. It was also highlighted that learners should stop taking photos which could later be used against them, they should avoid instigating cyber incidents. Lastly, it was discovered that the ICT Policy, School Codes of Conduct, Identification of Aggressors and Anti-Cyber Aggression Awareness were the pillars contributing to the development of reporting procedures for cyber incidents in schools.

5.10.   Learners' Perceptions about Cyber Incidents

The results indicated that cyber incidents among peers could be perceived as harmless or humorous and with no ways of stopping cyber incidents once they are viral. Learners are of the view that cyber incidents are not reported as often as they should. At times learners do not have the courage to talk about the incidents; in most cases victims are only willing to report to someone they feel comfortable with or someone they can trust. The other important observation is that in most cases it is not known how these issues are solved or what the outcomes are. Aggressors perpetrate incidents to seek revenge and sometimes a victim may become an aggressor due to revenge. The other way learners manage cyber incidents is physical fighting. The results also pointed out that to facilitate learner confidentiality and trust towards adults, learners must be able to report cyber incidents anonymously and their confidentiality must be respected. Learners' voices are essential, and their positive suggestions should be included in the design and implementation of the cyber incident handling procedures. However, without a confidential, voluntary, and anonymous reporting system, it is unlikely that learners will feel free to report cyber incidents. Learners fear being singled out; they fear that someone might retaliate against them. The results indicated that the Reporting System and Management of Cyber Incidents were the pillars for reporting procedures.

5.11.   Learners' Perceptions about Adults

The study also assessed the leaner's perception about adults. The results indicated that in most cases learners are scared that adults will worsen the situation because of them being too judgmental about the behaviour of learners. Many parents are seen to always remind their children of the issues which they would have done wrong. Leaners also believe that adults do not have any role in the intervention of cyber incidents and are ineffective when they attempt to intervene. Leaners are also of the view that educators are not helping when it comes to cyber incidents to an extent that some learners are not comfortable telling them about their cyber incident problems. This is why some learners end up being absent from classes, avoiding being embarrassed. All these results point to the fact that there is no open and good relationship between learners and the adults around them. The results indicated that learners wish their parents could be more involved in their online lives from an early age. Most learners want parents to have a good relationship with them, to communicate with them, to spend more time with them, to support them and even to join them in cyber space. Learners believe that the most important issues that many parents are interested in are about schoolwork. As a result, many parents do not know what is happening in their children's

daily lives. The other critical factor is the huge generation gap between children and their parents; there is a great difference between how children are being raised and how their parents were raised. As a result, this gives learners the idea that adults are not confident to help and intervene during cyber incidents. Some learners pointed out that it is not known how principals are handling cyber incident cases and the outcomes are not known. This is due to adults being absent from cyber space, which compromises their ability to assist. Leaners believe that the fact that most parents are not using social media, makes it very difficult for them to identify cyber aggressors.

In summary, learners indicated that they do not have confidence in adults, especially when it comes to their online problems. Role players need to be competent in handling cyber incidents and learners need to be assured of their safety when reporting cyber incidents, and their reports should be taken seriously and acted upon. All role players should be informed of the cyber incident reporting procedures and should be aware of how important it is to report cases as early as possible. The school community should know whom to approach if they become aware of or suspect any cyber incident taking place. Parents should check their children's phones if they notice any behaviour change. Parents should block the aggressor, notify their service provider, and report cyber incidents to the school and law enforcement. Parents should tell their children not to go about taking naked pictures of themselves and posting them online. All school role players should be educated about cyber incidents. Parents need to learn the legal issues associated with cyber aggression. Severe cyber incidents should be considered as crime. Law enforcement should have a specialised task team to deal with cyber incidents. SAPS should have a cyber division. The pillars for reporting procedures derived from the themes are Competence in Handling Cyber Incidents, Awareness of The Reporting Procedures, Learner Assistance from Parents, Cyber Aggression Awareness, Legislation, And Cyber Division for Law Enforcement.

6.   Summary and Conclusion

Access to new technology in schools has led to an increase in misuse and abuse of technology by learners and this has led to a rise in cyber incidents related to threats, harassment, embarrassments and humiliating behaviour, and various bad activities online. The study, therefore, proposes that cyber incident handling procedures, and the roles and responsibilities of role players should be in place for cyber incidents to be addressed in South African schools. Using the literature review approach and thematic analysis of the data collected, the results indicated that it is important that role players should be competent in handling cyber incidents, and they should be informed of the cyber incident reporting procedures. The other important finding was that the cyber incident handling procedures should be readily accessible to learners, parents, guardians, and school staff members and they should be clear and easily understandable. The other issue that was highlighted was the need for a consistent, universal, and effective ICT policy directly aimed at regulating cyber incidents in schools. The codes of conduct or a code of behaviour should be formulated and adopted. The study also discovered that there should be a need for a referral system to appropriate services, and community partnerships to support learners and build a cyber school safe environment.

The study also come up with the views of learners on various aspects of cyber incidents in schools which include their perceptions on victims, aggressors and bystanders; learners' perceptions about technology; learners' perceptions about cyber incidents and learners' perceptions about adults. For instance, learners' perceptions on cyber victims, aggressors and bystanders were that cyber victims feel like outcasts, hurt, offended, embarrassed and inferior. These victims fear aggressors because aggressors take out their pain or anger on other learners. The results also pointed out that bystanders support aggressors because of fear of becoming the next victims, being rejected by peers or uncertain of whom to report to. The results presented helps to provide a framework that will act as a guide on reporting cyber incidents and directing school management, and all within the school, towards appropriate reporting procedures and intervention processes. The study also found out that the rise in cyber incidents in South Africa if they are left unattended, can have a devastating effect on learners. Therefore, the government of South Africa through the department of basic education must prioritize the handling of cyber incidents in schools as cyber incidents are now a threat to the efficient and effective execution of the mandate of the department.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] E. B. Meier, "Designing and using digital platforms for 21st century learning," Educational Technology Research and Development, **2021**(1), 1–4, 2021, doi.org/10.1007/s11423-020-09880-4.

[2] M. Finn-Stevenson, Schools of the 21st century: Linking child care and education, Routledge, 2018.

[3] S. McCoy and S. Lyons, Digital Technologies and Student Learning. Ireland's yearbook of education 2018-2019, Education Matters, 2018.

[4] J. Remón, V. Sebastián, E. Romero, and J. Arauzo, "Effect of using smartphones as clickers and tablets as digital whiteboards on students' engagement and learning.," Active Learning in Higher Education, **18**(2), 173–187, 2017, doi.org/10.1177/1469787417707618.

[5] L. K. Butola, "E-learning-A New Trend of Learning in 21st Century During Covid-19 Pandemic.," Indian Journal of Forensic Medicine & Toxicology, **15**(1), 422–426, 2021.

[6] J. Byrne, D. Kardefelt-Winther, S. Livingstone, and M. Stoilova, Global Kids Online Research Synthesis, 2015-2016, London, 2016.

[7] I. . Ogaji, P. Okoyeukwu, I. Wanjiku, E. A. Osiro, and D. Ogutu, "Pattern of use of social media networking by Pharmacy students of Kenyatta university, Nairobi, Kenya," Computers in Human Behavior, **100**(66), 211–216, 2017, doi.org/10.1016/j.chb.2016.09.035.

[8] Queensland Government, "Adjust our settings: A community approach to address cyberbullying among children and young people in Queensland," Queensland, 2018.

[9] J. D. Shapka, H. Z. Onditi, R. J. Collie, and N. Lapidot-Lefler, "Cyberbullying and cybervictimization within a cross-cultural context: A study of Canadian and Tanzanian adolescents.," Child Development, **89**(1), 89–99, 2018, doi.org/10.1111/cdev.12829.

[10] C. A. Hills, Developing a law and policy framework to regulate cyber bullying in South African schools, PhD Thesis, University of South Africa, 2017.

[11] J. M. Lee, J. S. Hong, J. Yoon, A. A. Perguero, and H. J. Seok, "Correlates of adolescent cyberbullying in South Korea in multiple contexts: A review of the literature and implications for research and school practice.," Deviant Behavior, **39**(3), 293–308, 2017, doi.org/10.1080/01639625.2016.1269568.

[12] D. Cross, T. Shaw, K. Hadwen, P. Cardoso, P. Slee, C. Roberts, L. Thomas, and A. Barnes, "Longitudinal impact of the cyber friendly schools program on adolescents' cyberbullying behavior," Aggressive Beviour, **42**(2), 166–180, 2016, doi.org/10.1002/ab.21609.

[13] P. Burton, L. Leoschut, and J. Phyfer, "South African Kids Online: A glimpse into children's internet use and online activities, Cape Town,"

[14] http://www.cjcp.org.za/uploads/2/7/8/4/27845461/south_africa_kids_online_full_report.pdf, Accessed: 17-Nov-2017.

[15] E. Kritzinger, "Improving cybersafety maturity of South African schools," Information, **11**(10), 471–488, 2020, doi.org/10.3390/info11100471.

[16] L. Cilliers and W. Chinyamurindi, "Perceptions of cyber bullying in primary and secondary schools among student teachers in the Eastern Cape Province of South Africa.," Electronic Journal of Information Systems in Developing Countries, **86**(4), 1–10, 2020, doi.org/10.1002/isd2.12131.

[16] S. G. Denby, Cyberbullying: School Administrators' Perceptions of Law and Prevalence, and Roles in Prevention, Intervention, and Discpline., PhD Thesis, Virginia Commonwealth University, 2020, doi.org/10.25772/NQZM-FM42.

[17] M. Astatke, C. Weng, and S. Chen, "A literature review of the effects of social networking sites on secondary school students' academic achievement.," Interactive learning Environments, **2021**, 1–17, 2021.

[18] L. M. Hellsten, An Introduction to Cyberbullying, Macerata, 2017.

[19] R. Von Solms and S. Von Solms, "Cyber safety education in developing countries," Journal of Systemics, Cybermetics Informatics, **13**(2), 14–19, 2015.

[20] S. C. Yang, "A meta-model of cybersecurity curriculums: Assessing cybersecurity curricular frameworks for business schools.," Journal of Education for Business, **96**(2), 99–110, 2021, doi.org/10.1080/08832323.2020.1757594.

[21] D. Mhlanga, "Industry 4.0: The Challenges Associated with The Digital Transformation of Education in South Africa.," https://www.researchgate.net/publication/344230555_Industry_40_The_Ch allenges_Associated_with_The_Digital_Transformation_of_Education_in_ South_Africa, Accessed: 02-Aug-2021.

[22] South African Market Access, "Education Statistics - South African Market Insights," https://www.southafricanmi.com/education-statistics.html, Accessed: 02-Aug-2021.

[23] Department of Basic Education, Education Statistics in South Africa 2016. Department of Basic Education, 2018.

[24] D. Mhlanga and T. Moloi, "COVID-19 and the Digital Transformation of Education: What are we learning on 4IR in South Africa?," Educatin Sciences, **10**(7), 180, 2020, doi.org/10.3390/educsci10070180.

[25] A. DeSmet, N. Aelterman, S. Bastiaensens, K. Van Cleemput, K. Poels, H. Vandebosch, G. Cardon, and I. De Bourdeaudhuij, "Secondary school educators' perceptions and practices in handling cyberbullying among adolescents : a cluster analysis," Computers and Education, **88**, 192–201, 2015, doi.org/10.1016/j.compedu.2015.05.006.

[26] J. L. Tinstman Jones, L. O. Campbell, J. Stickl Haugen, and C. C. Sutter, "Cyberbullying considerations for school counselors: A social media content analysis," Professional School Counseling, **23**(1), 2020, doi.org/10.1177/2156759X20919365.

[27] C. A. Myers and H. Cowie, "Cyberbullying across the lifespan of education: Issues and interventions from school to university," International Journal of Environmental Research and Public Health, **16**(7), 1–14, 2019, doi.org/10.3390/ijerph16071217.

[28] P. Redmond, J. V. Lock, and V. Smart, "Developing a cyberbullying conceptual framework for educators.," Technology in Society, **60**, 101223, 2020, doi.org/10.1016/j.techsoc.2019.101223.

[29] C. Angus, "Cyberbullying of Children: E-Brief," Australia, 2016.

[30] D. H. Tustin, G. N. Zulu, and A. Basson, "Bullying among secondary school learners in South Africa with specific emphasis on cyber bullying; South African Professional Society on the Abuse of Children: ISSN 1562-1383 Child Abuse Research," A South African Journal 2014, **15**(2), 13–25, 2014.

[31] M. Bulger, P. Burton, B. O'Neill, and E. Staksrud, "Where policy and practice collide: Comparing United States, South African and European Union approaches to protecting children online," new media & society, http://www.cjcp.org.za/uploads/2/7/8/4/27845461/protecting_children_onli ne.pdf, Accessed: 03-Mar-2018.

[32] P. Burton, L. Leoschut, and J. Phyfer, "South African kids online: barriers, opportunities, and risks. A glimpse into South African children's internet use and online activities. Techical Report.," Cape Town, 2016.

[33] J. Phyfer, P. Burton, and L. Leoschut, "South African Kids Online.," Cape Town, 2017.

[34] N. A. A. Rahman, I. H. Sairi, N. A. M. Zizi, and F. Khalid, "The importance of cybersecurity education in school," International Journal of Information and Education Technology, **10**(5), 378–382, 2020, doi:10.18178/ijiet.2020.10.5.1393.

[35] H. Qusa and J. Tarazi, "Cyber-Hero: A Gamification Framework for Cyber Security Awareness for High Schools Students," in 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), 2021, pp. 0677–0682, doi:10.1109/CCWC51732.2021.9375847.

[36] M. Torres and N. Thompson, "Toward a Cyber Security Adoption Framework for Primary and Secondary Education Providers," in 31st Australasian Conference on Information Systems, 2020.

[37] D. Scholtz, E. Kritzinger, and A. Botha, "Cyber Safety Awareness Framework for South African Schools to Enhance Cyber Safety Awareness," in Computer Science Online Conference, 216–223, 2020, doi:10.1007/978-3-030-51974-2_19.

[38] A. S. De Vos, H. Strydom, and C. B. Fouche, Research at Grassroots for the Social Sciences andHuman Service Professions, Van Schaiks, 2009.

[39] J. H. McMillan and S. Schumacher, Research in education: A conceptual introduction, Longman, 2001.

[40] S. Hymel and S. M. Swearer, "Four decades of research on school bullying: An introduction.," American Psychologist, **70**(4), 293–299, 2015, doi.org/10.1037/a0038928.

[41] D. Adom, E. K. Hussein, and J. A. Agyem, "Theoretical and conceptual framework: mandatory ingredients of a quality research," International Journal of Scientific Research, **7**(1), 158–172, 2018.

[42] P. K. Smith, Cyberbullying and cyber aggression, Routledge, 2012.

[43] E. Maramwidze-Merrison, "Innovative methodologies in qualitative research: Social media window for accessing organisational elites for interviews," The Electronic Journal of Business Research Methods, **12**(2), 157–167, 2016.

[44] R. S. Tokunaga, "Following you home from school: A critical review and synthesis of research on cyberbullying victimization," Computers in Human Behavior, **26**(3), 277–287, 2010, doi.org/10.1016/j.chb.2009.11.014.

The Design and Implementation of Intelligent English Learning Chabot based on Transfer Learning Technology

Nuobei Shi, Qin Zeng, Raymond Shu Tak Lee[*]

Division of Science and Technology, Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, 519000, China

A R T I C L E   I N F O

A B S T R A C T

Chatbot operates task-oriented customer services in special and open domains at different mobile devices. Its related products such as knowledge base Question-Answer System also benefit daily activities. Chatbot functions generally include automatic speech recognition (ASR), natural language understanding (NLU), dialogue management (DM), natural language generation (NLG) and speech synthesis (SS). In this paper, we proposed a Transfer-based English Language learning chatbot with three learning system levels for real-world application, which integrate recognition service from Google and GPT-2 Open AI with dialogue tasks in NLU and NLG at a WeChat mini-program. From operational perspective, three levels for learning languages systematically were devised: phonetics, semantic and "free-style conversation" simulation in English. First level is to correct pronunciation in voice recognition and learning sentence syntactic. Second is a converse special-domain and the highest third level is a language chatbot communication as free-style conversation agent. From implementation perspective, the Language Learning agent integrates into a WeChat mini-program to devise three user interface levels and to fine-tune transfer learning GPT-2 as back-end language model to generate responses for users. With the combination of the two parts about operation and implementation, based on the Neural Network model of transfer learning technology, different users test the system with open-domain topic acquiring good communication experience and proved it ready to be the industrial application to be used. All of our source codes had uploaded to GitHub: https://github.com/p930203110/EnglishLanguageRobot.

1. Introduction

Chatbot operates similar to *virtual personal assistant (VPA)* and *question-answer (QA) system* in its development. In general, a practical chatbot consists of two categories, one has no *artificial intelligence (AI)* technologies i.e. rule-based and pattern recognition that collects tremendous high quality artificial corpus as database for *question-answer (QA)* matching [1]. Another has AI that uses current models and algorithms enabling chatbot to learn necessary articles before use. A typical example is speech recognition with text-to-speech interconversion to listen and converse at real-world human-machine interaction with database as machine brain or corpus with various meanings used as *natural language understanding (NLU)*, *natural language generation (NLG)*, *neural network (NN) models* with sufficient target language corpus training to equip the brain with thoughts and

communication. This new human-machine interaction method provides convenience but it generates interior questions on how the chatbot can perform intelligently. It has been an ongoing research topic on user interface and expected functions by many internet organizations. However, back-end support require both AI and/or non-AI technologies to operate effectively. A literature review on human-like chatbot with more than one algorithm to select the best response that similar to Microsoft's Xiao Ice, a software assistant with emotions will be presented. A chatbot's performance is dependent on core components enriched with sufficient algorithms. Thus, a transfer-based technology with different language learning models would be proposed in this paper.

Turing test progress enticed researches on rule-based human-machine interaction system using pattern recognition in 1950s. Corpus as response matching outperformed Turing test showed

[*]Corresponding Author: Raymond Lee, Email: raymondshtlee@uic.edu.cn

that database can be enlarged by sufficient human dialogue directing that human syntactic simulation is not necessary as semantic matched natural human dialogue for response. Since then, when natural language is generated by trained machine with big data, the rule-based chatbot is updated to an *information-retrieval (IR)* or a *neural network (NN)* based system automatically and reduced manpower considerably. IR based system is mostly applied in a search engine optimizer with constructed knowledge base whereas NN system is densely data-driven to perform high-level prediction and classification. It can be used as a transfer learning method to pre-train and fine-tune the basic concepts of a chatbot. As shown on Figure 1, machine linguistic is depended on its data learned such as the basic level with words, phrases and sentences etc. For *natural language processing (NLP)*, an initial step is to transform natural language into a word vector followed by mathematical computation to illustrate the basic meanings equivalent to real-world. Thus, NN has the ability to simulate human brain and self- generate natural language as compared with other methods.
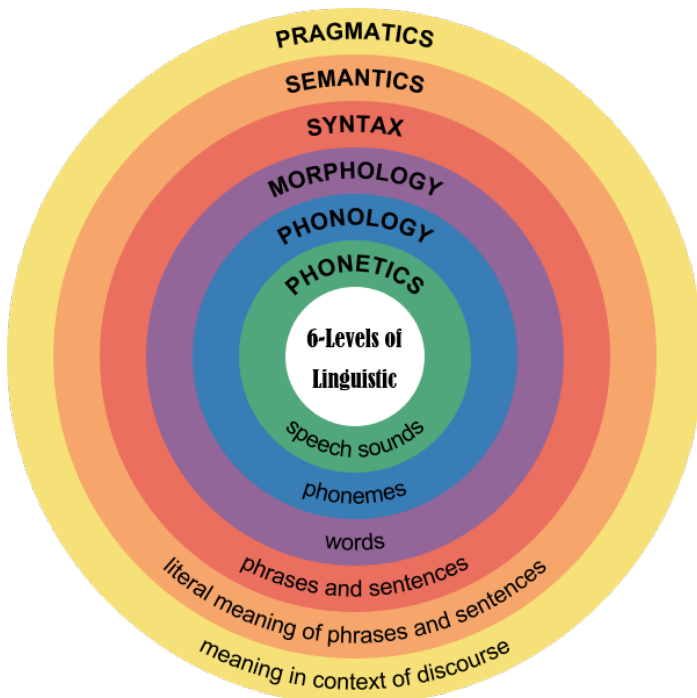
Figure 1: Levels of Linguistic in human languages [2]

There are numerous chatbot models equipped with different algorithms. Transfer learning is amongst performed better than various unsupervised learning sequence-to-sequence (seq2seq) models i.e. *recurrent neural network (RNN)* and its related bi-direction RNN and *long-short term memory (LSTM)* models. BERT from Google and GPT-2 from Open-AI are the most famous and pre-trained transformer models to generate natural language and classification. In this paper, GPT-2 is used as English language learning back-end model to converse daily dialogues because of GPT-2 inherent priority of natural language generation. For pre-trained transfer learning language model, a small dataset is used to fine-tune the transformer model to outperform traditional counterparts. Also, a self-attention mechanism is included. It is the main theme of transfer learning to find the relationship between sequences. Further, a language mini-program based on GPT-2 with

a fine-tuned model will be included to provide fluent and syntactic responses. Lastly, a mini-program equipped with a fine-tuned GPT-2 language model and speech recognition service from Google to deliver a three systematical English learning level functions at WeChat for implementation.

2. Literature Review

2.1. Overview of industrial English Learning apps

Duolingo is an English language learning application with growing popularity since coronavirus. It is accredited by universities in UK, US and Australia as English language proficiency qualifications similar to IELTS and TOEFL. It contains spoken tests' pronunciation and oral presentation on text questions or images understanding with scores given depend on voice recognition. Test contents are preinstalled without human participation and gradings are given by statistic algorithm that are different from IELTS by physical examiner. It covered more than 30 mainstream languages transmitted in English and more than 6 languages relayed to Chinese language. Prior course commencement, the application would match users' demands to set step by step learning schedules and provided tests to detect grammar and pronunciation proficiency. Level detection is an important component to improve language skills because it depended on applications or chatbots functions and users' willingness to select appropriate learning levels at different courses. [3] It shows that the online English learning or human-machine interaction English test will be the mainstream in these days. However, the technologies it used are not clear or transparent for users, the statistical algorithm or non-AI technologies also cannot substitute the official recognized IELTS which is communicate with real human or the machine records judge by human. On the contrary, our model are natural based on AI technologies and growing with the interaction between human and machine.

The model we proposed is based on IELTS Part ABC speaking module examination syllabus covering pronounciation, topic discussion and free chat. They are named as Level 1-3 where the highest level 3 is a human-machine interaction dialogue. As responses from reverse party generated by IELTS are in random so the same format is followed where responses are generated by a back-end transfer learning model similar to a proficient English Tutor. Corpus of big data used for transfer learning model training covers topic knowledge discussion.

Other reference is based on a different English language learning application called LiuliShuo. It contains 7 levels to improve verbal skills. Level 4 refers to proficiency equivalent to College English Test (CET)-4 or 6, where the highest level 7 is equivalent to IELTS 7.5 and TOEFL 105. LiuilShuo settings, compared with Duolingo in user interface and functions design are fitted to the model design. However, they are sufficient for grades evaluation but are lack of AI advanced intelligent technologies for response according to user's learning progress.

Thus, the design of the proposed model began with daliy dialogue for model training to improve its learning ability, then a database was added to track every users' learning history and levels improvement. For this segment, daily chitchat data in natural langauge are transferred into json format to improve machine learning. Those unstructed raw data or called heterogeneous data

required cleaning and sorting automatically in accordance with AI ecosystem.

GAMIFICATION EXAMPLES IN EDUCATION

Kickstarter
An American crowdfunding platform has launched a campaign on Canny bot - a robot toy that can be controlled remotely via a programme and is used to teach children practical coding skills.

Quizlet
A web and mobile application to study Math, Science, Languages, Arts quicker via games and flashcards for free. It's used by half of the universities of the USA and is said to help with progress a lot.

Khan Academy
Designs interactive video short lessons and supplies users with practice exercises and additional materials for creative learning of different subjects at their own pace.

Duolingo
Language-learning game with more than 8 millions downloads on Google Play. The social app encourages to level app earning points, get virtual currency and learn with friends.

ClassDojo
An interactive, funny and practical solution that connects children with teachers and helps create and manage better learning via virtual classes, sharing media and messaging.

Kahoot
A solution to use at school and at work for e-learning. While playing, anyone feels more comfortable and confident in vocabulary quizzes, multiplication, geography and more.

Figure 2: Gamification examples in Education [4]

Figure 2 shows a set of gamification examples in education. It is a game module to calculate grade levels to motivate users compete within app for improvement.

2.2. Academic research about Chatbot.

2.2.1. AliMeChat: A Sequence to Sequence and Rerank based Chatbot Engine (rule-based IR+seq2seq) [5]

Recurrent neural network (RNN) is capable to generate responses from end to end leading to sequence to sequence (seq2seq) model becomes chatbot mainstream generation prior 2018. Since then, NLP tasks were greatly improved using transfer learning model.

TaoBao App uses AliMe to substitute online human customer service for most of the predicted questions. Statistical data showed that 5% questions were chats from a commercial-domain Question-Answer system to an open-domain chatbot. This process resembled to English language learning agent similar to level 2 scenario dialogue to level 3 open-domain free style conversation of this paper. AliMe integrated a rule-based IR and pre-trained seq2seq application model to real-world industry and performed better than both IR and public chatbot. However, its pre-trained seq2seq model is used twice for response generation to re-rank a set of answer from matching input questions with knowledge base paired dialogues. It suggested that free style format can be based on usual conversation scenario exist in daily activities.

The proposed model has two similar AliMe settings in generation for information retrieval but in different formats. IR-based model uses natural language word to match knowledge base, the generation seq2seq model and model evaluation to re-rank

output response in generation are embedded words as vector. IR-based dataset is approximately 9,164,834 natural questions and answers paired dialogues from users and staff collected from commercial domain. Researchers use inverted index based on search engine concept to match these conversation with input sentence containing identical words and used BM25 algorithm to measure input sentence similarity to select questions and respond to input questions. Traditional chatbot back-end also avoided commonsense questions that cannot be responded by chatbot and generation model used seq2seq model.

AliMe also selected GRU, a type of RNN units to reduce computation time for response generation. Further, an optimizer SoftMax is used in calculate to time control sample words set covering 512 random words. Beam Search in decoder assisted to find the highest probability in Conditional Probability and obtain the optimized response sentence within parameters. Results performance showed that *IR+generation+rerank* shown on Figure 3, by seq2seq model and mean probability scoring function evaluation approach achieved the highest score compared with other separated methods.
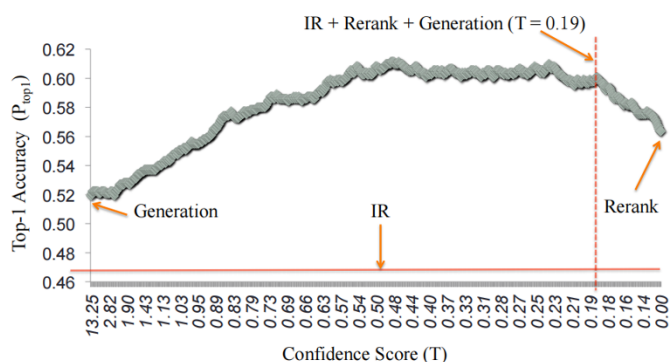
Figure 3: Top-1 accuracy of candidate approaches [5]

Due to the characteristic of NN and chatbot usage or a QA-based system, special-domain system in contents apart from the task-oriented or goal-oriented system for special usage than open-domain system applications are essential for model design. Microsoft chatbot is open-domain with emotions and AliMe specialized in business as Taobao customer service restricted in business corpus. Also, grounded knowledge contents becomes another important trend for chatbot developments[6]. In summary, rule-based dialogue with seq2seq RNN model are used frequently than others because most daily conversation cannot be generated by RNN model. Since attention mechanism proposed in 2017 in Attention is all you need [7], language model in natural language processing changed to *seq2seq+attention* and self-attention in transfer performed better than seq2seq.

2.2.2. The Design and Implementation of Xiao Ice, an Empathetic Social Chatbot（IR+seq2seq+KG）[8]

Microsoft released an empathetic social chatbot based on two segments at various platforms. One is intelligence quotient (IQ) that used for thinking and answer human questions. Another is emotional quotient (EQ) to understand and analysis the emotional meaning of natural language to provide 24 hours user services, research team use three back-end to equip the machine with human-like brain to realize these two segments. First is the traditional IR with corpus for matching. Second is a RNN of sequence to sequence model to generate response and the third is a knowledge graph for entities extraction to extract related entities

and sorting for responses. After that, an optimizer would rank potential responses and select the most suitable response to users.

Each Microsoft operation system has a unique Cortana for its computer owner [8]. Xiao Ice has different Cortana as a personal assistant resembled to Siri. It preferred to chat with IQ and EQ system design similar to human interaction. It is by far the most dialogue-oriented AI robot with emotions to provide 24 hours interaction. It has over 660 million users since its inception with *conversation-turns per session* (CPS) reached 23 which is higher than other chatbots. It is also an open-domain chatbot aimed to optimize content responses with emotions to understand users' thoughts. It has expanded to more than 5 countries with different names. Users willing to use not only depended on IQ but also dialogue contents provided a user-friendly first perception. Thus, IQ part would also be used in the proposed model.
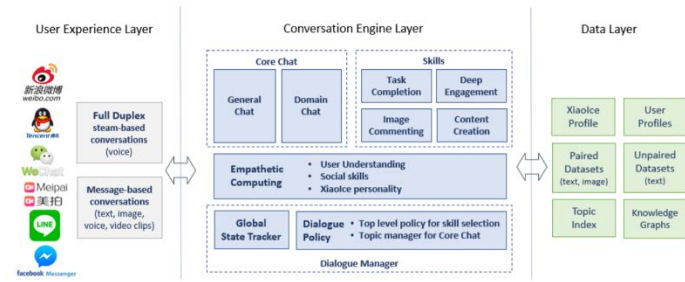
Figure 4: XiaoIce system architecture [8]

In Figure 4, for users implementation, Xiao Ice exists in 11 mainstream platforms including WeChat and Tencent QQ and at mini blog, Weibo and Facebook. It has the highest recognition function that include text, image voice and message-based conversation. Its conversation engine layer was developed after four years and its Core Chat can distinguish and change topics automatically. Thus, the proposed model is set with a second level at topic dialogues for special domain scenario and a third level for free style conversation to improve users' dialogue proficiency. A dialogue manager is to define dialogue management policy so that it can monitor the state of dialogue consisting Global State Tracker and Dialogue Policy to select a state action. Global State Tracker is a vector consisting of Xiao Ice responses and analyses text strings to extract entities and empathy. Dialogue policy is designed for long-term users to improve interactive engagement optimization. Initially, topic manager is the first to greet users and awake system with natural language understanding (NLU). It will then obtain users' responses' to adjust the "Core Chat" into open-domain or special scenario with users' interests. If system always repeats or response information is bland and within three words, the topic will be changed accordingly. Secondly, is the skill selection. Once user input is a particular format, the skills would activate to process different input. Its images can be classified into different task-oriented scenario e.g. if the image is food, a restaurant will be displayed. Lastly, a personal assistant would provide information such as weather, space availability and reservation for task completion.

In other words, it contained certain knowledge graphs in data layer where dataset is originated from popular forum such as Instagram in English or douban.com in Chinese. These datasets listed as several topics with relative small knowledge base as candidate answer and are sufficient for questions. When a new topic appears, system would obey the policy to refresh its knowledge base by machine-learning. However, not all new entity

or topic will be collected. It is similar to a ranking list unless the entity is contextual relevance or the topic has higher popularity or news latest update and users' personal interests will be considered.

The core of Xiao Ice is language model, using the RNN shown in Figure 5 to create the open and special domain knowledge.
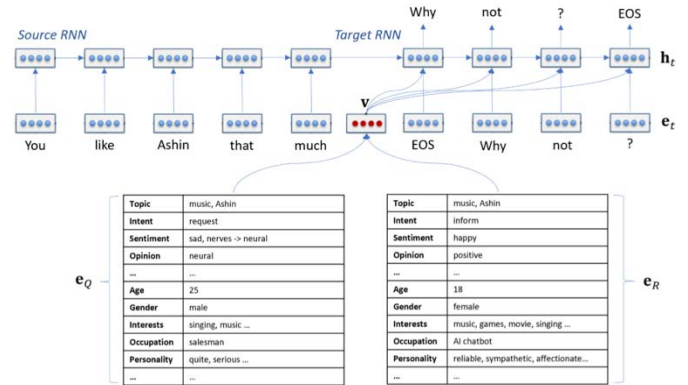
Figure 5: RNN-based neural response generator [8]

Figure 6: Examples of inconsistent responses generated using a seq2seq model [8]

Figure 6 shows a chatbot with emotional consciousness. It enhanced interaction to improve system EQ performance.

Xiao Ice, in general, has retrieval and generated-based back-end models. The former is a mature search engine used for decades. Its knowledge graph is a database for graph matching and entities extraction where the latter is trained by big data with related dialogue based on entities relations to supply contents and explain output sentences from NN. These two directions using neural model to generate responses and rank sentences to match relations and optimization. Once entity recognition is linked with neural generator conversation and dynamic knowledge graph and the knowledge-based will be refreshed. If user's feedbacks match with English syntax, the knowledge-based would always be collected together with users' interests. In addition, a rule-based chatbot with *artificial intelligence markup language (AIML)* as backup will be included to avoid senseless or repeat answers.

Thus, cognition, reinforcement learning of history dialogue[9] and empathetic computing are essential segments in the proposed model for competitiveness.

2.2.3. *Attention Is All You Need [7]*

Although chatbots mentioned above are acceptable for usage, its training back-end testing about sequence to sequence model showed good performance for text generation. Considering neural network relied on big data, data-driven model for better performance. Thus, transfer learning model shown on Figure 7 with magnitude training and fine-tune data has been the architecture mainstream. With the recognition of attention mechanism, the updated self-attention created the priority of text generation to be proved by BERT and GPT and GPT has natural advantage for text generation within sequence.
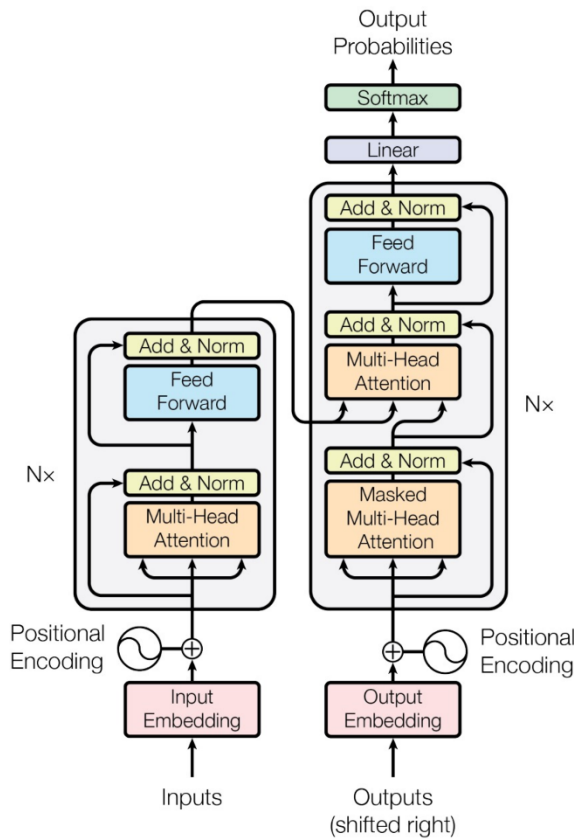
Figure 7: Transfer learning-model architecture [7].

The proposed model used as back-end is *Transfer Transfo*[1], as a daily dialogue system with common sense which integrate transfer learning training architecture and high-quality dataset. By fine-tuned persona-chat, the original utterance of GPT-2 with text generation based on context changed to format dialogue with fine-tune content dataset and English skills with original magnitude GPT-2 model. Also, the fine-tuned dataset also added the contents of English Tutor with users' personality and background resemble to native IELTS examiner.

Attention is human preference as a weight added to other words computation. It means that the encoding related to context, input and output. Transformer use self-attention that is search relations within sequence finding dependency among words in one sentence.

BERT is the first system proposed in transfer learning and fine-tuning. It has widely used in especially encoder part of pre-trained language model. However, it cannot predict decoder outputs. Open-AI GPT-2 has better application for generation with corpus and decoder limitation. The most famous performance of GPT-2 is text generation, since it releases larger dataset, more users are willing to make text prediction and dialogue. With the same transfer learning mechanism, GPT-2 not only can build dictionary use tokenizer like T, also play a role of language teacher with pre-trained grammar in natural language processing. Also, the GPT-2 original training source is 40G text from Internet, it has better adaption for text generation.

When the first edition of GPT-2 open source, researchers would like to test its performance with larger data. The 40G text in open-domain enabled GPT-2 equip with sufficient common sense similar to AIML(Artificial Intelligence Markup Language) which

consist of human daily dialogues, but with more contents like grounded knowledge of Microsoft chatbot. For researchers, fine-tuning is the process to make GPT-2 transfer as own language model with special tasks. Text generation, text classification, automatic summarization, music modeling even writing code, all are functions of fine-tuned GPT-2 by users.

Thomas Wolf and his partners applied GPT-2 into dialogue generations called *Transfer Transfo* [1] to develop conversational system. Besides rule-based and seq2seq dialogue system, Transfo is another data-driven dialogue agent based on GPT-2, it combines transfer learning method to training and the transfer learning model with abundant contents with high-capacity, which shows big progress compared with end to end like seq2seq and information retrieval models. For traditional language models, neural network cannot track dialogue history within the model, even it learns the conversation with users. It is only big data to learn which cannot be tracked like rule-based system with database to store history dialogues. For training and generation algorithm, Transfo has priority to improve in relevance, coherence also grammar and fluency for output. For the model, multi-layers with 12 decoders and masked self-attention heads that token only range from left context. The large transfer learning model used in several practical task pays more attention to decoder part for the down-stream NLP tasks [10]. Pre-trained is an important part for down-stream performance to avoid discrete sentence to construct systematical learning in grammar and continuous content, document level corpus is better than sentence level. So, an Open AI release different size corpus for researchers to select suitable pre-trained and fine-tuning model afterwards. Fine-tuning decides the down-stream task text format in this paper. Thomas Wolf use persona-chat dataset as fine-tuning data to train input and output utterance change from long-text to dialogue format. Persona-chat in real-world helped to shape the speakers' backgrounds to make system further define topics and better understand users' input sentences .

Figure 8: TranferTransfo's input representation [1]

In Figure 8, tokens embedding in one sequence is consist of 3 embedding of word, dialogue state and positional as shown in Fig 8. Rather than simple end to end dialogue with only one sentence, it collects usually 4-6 sentence and 3-5 of history dialogue belongs to same user in one sequence for training helping system to judge , state and position representation of every token to predict the next sentence in real application.

A dialogue system based on GPT-2 is included in the proposed model. In [11], the author made dialogue task more specific to goal-oriented personal assistant to help users complete special tasks. For task-oriented training, it has higher quality for corpus to fine-tune model and track conversation state of various situations. Not only it can interact with external system rather than pipelined system but also explain why the system generate such response.

End-to-end neural architecture for dialogue systems built on the top of GPT-2 with dialogue format. So, the dataset for fine-tuning should be dialogue as shown in Figure 9, which include external database of restaurants, hotels and others scenarios traveling in UK. The dialogues link to database contents to predict the system dialogue state and system action. Even more, following

the traditional end to end pipeline plus the database to be more explainable in output sentence that show why the system provides direction or operation advice as responds to users. Although it is a simple gradient descent, GPT-2 pre-trained model already equipped competitive dialogue as response.
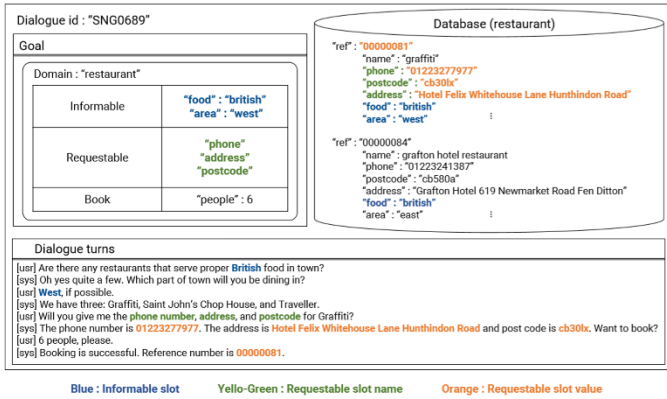
Figure 9: A single-domain example in MultiWOZ dataset [11]

Figure 10 shows the architecture of model similar to the model of grounded knowledge by Microsoft. Because the external system also a supplement for normal case which is a sequence to sequence model to generate substantial knowledge. When transfer learning-based system requires contents to engage specific task, knowledge base to fill the slot is another method to accomplish same object.
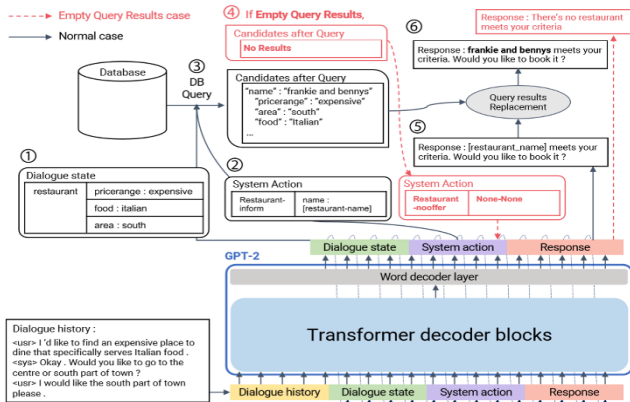
Figure 10: Architecture of end-to-end neural dialogue model[11]
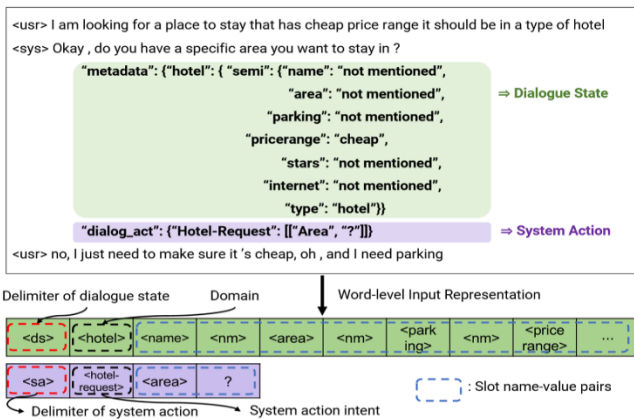
Figure 11: MultiWOZ dataset for test [11]

Dialogue state and system action, shown on Figure 11, is fundamental standard for the procedure of architecture. With personal assistant setting, this model also requires decoding strategy to optimize output sentence conditional on probability distribution. In this paper, top-k and top-p sampling is better than general beam search which are not suitable for high-entropy NLG and greedy decoding get highest probability over all probability such as for machine translation. To some extent, even though the database requires query to get real-world information tokens, there are not all sentence need external system to fill the slot, GPT-2 model will consider the situation to handle empty query for common sense responses and completion [12].

So far, progress in dialogue system in recent years due to the development in language model of natural language processing. In fact, most of techniques are for machine translation, sequence to sequence model and attention mechanism also the ones for machine translation. Until GPT-2, which natural has ability to generate text when the model is tested in below section. So, GPT-2 with transfer learning changeability for downstream natural language tasks with fine-tuning. Thus, we will use GPT-2 will be used as language model in the proposed model to obtain better response than seq2seq models.

After investigation and test of industrial products and some research projects about chatbot, with the development of AI technology, it was decided to use frameworks and transfer learning and fine-tune chatbot for English language learning. Chatbot in the proposed model will be a response generation chatbot with own common sense - the highest level is *free-style conversation agent* of a language system superior than other popular applications.

3. Methodology

In view of system framework, the design is firstly complied with the chatbot general flow at Figure 13, to decide whether it is functional. A top layer would be added followed by the application layer for systematically learning English which include three levels at User Interface designing scheme. As shown on Figure 12, the lower layer of architecture is the back-end model and the usage of language model using GPT-2 as language model responsible for the natural language understanding and natural language generation.
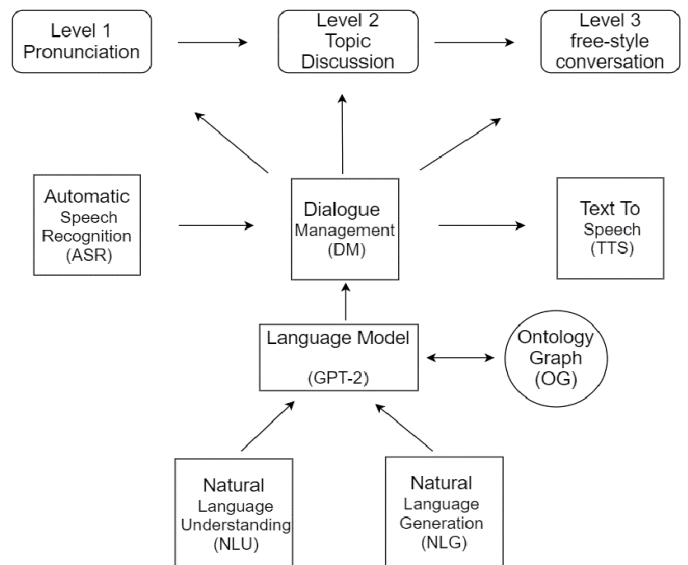
Figure 12: English learning chatbot architecture.

3.1. Chatbot system flow

In general, chatbot requires five parts to make communication with users and responses. The first step is voice recognition then change to text version natural language via ASR, the Automatic Speech Recognition is the same from text to voice in last step. For the central part, chatbot should handle natural language understanding and natural language generation using language model, it requires plenty of dataset as corpus to pre-train and fine-tune back-end model. About the optimization for daily dialogue and language model, with MySql database support the track of history dialogue distributed the user individually, the chatbot will become more intelligent and easier to modify with the long-term users.

Figure 13: System Flow of chatbot [13]

With the extension of system flow, in our architecture, the application layer are all English learning levels, which detail functions shown at implementation parts with mini-program User Interface and Develop Tool.

3.2. Open AI GPT-2 language model

Before we talk about Open-AI GPT-2, we should illustrate what is *language model (LM)*. For example, input editor is a language model which can predict the next word by language habits for behavior. From that, we can consider GPT-2 as a connection function of input editor. So far, GPT-2 already enlarge it language model with different size covering different needs of industrial NLP tasks.

Compared with the sequence to sequence we mentioned at section 2, the *transfer learning model* could provide better response that are more related to dialogue questions. The pre-trained dataset of Open AI GPT-2 are 40G text, it is no doubt that Open AI GPT-2 perform better than sequence to sequence model as it is data-driven neural network. The encoder part of GPT-2 will helps in the generation of natural language. Figure 14 shows the different size open source GPT-2 model for download and training.
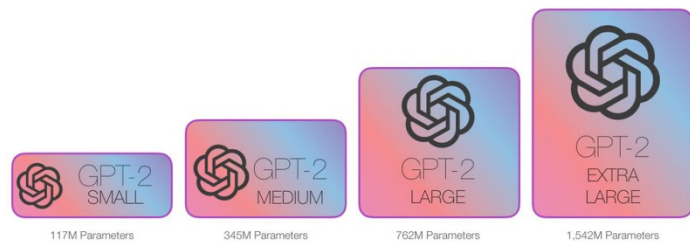
Figure 14: Size of Open AI GPT-2 pre-trained model [14]

Original GPT-2 always predicts the next token based on the information only at left hand of whole sentence. In summary, on probability condition, it uses history information with sufficient vocabulary to predict next token. After one prediction, it will provides a list for select, that is algorithm of top-k. In BERT, GPT-2 is also an architecture of transfer learning. Transfer learning consists of encoder and decoder, BERT focuses on the encoder and GPT-2 deletes encoder but keep decoder. Without encoder, decoder layers will be accumulated and magnified data. The idea

of self-regression used by GPT-2 resemble RNN where it puts the token generated at the back of sequence, which the sequence will be the new input for the condition of next token generated.

In Figure 15, the encoders and decoders have 6 encoder and decoder. It removed traditional RNN and the pre-trained model fine-tuned by researchers will become mainstream in the future.
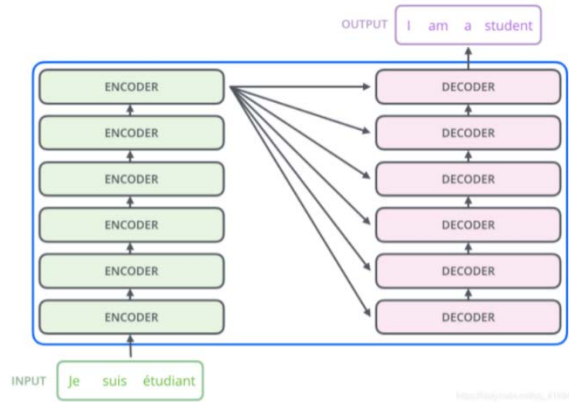
Figure 15: 6 layers encoder and decoder [14]
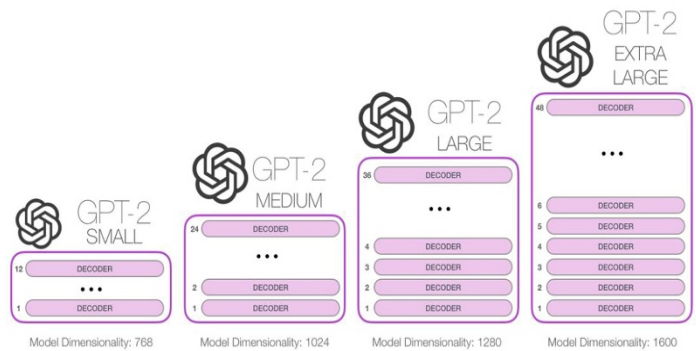
Figure 16: Decoder layers of Open AI GPT-2 [14]

The transfer learning change from BERT, shown on Figure 17, to GPT-2, the self-attention changes to masked self-attention shown on Figure 18. In Figure 16, due to GPT-2 have no encoder, input sentence as reminder is incomplete. The empty part of BERT covered with <eos> and other characters but GPT-2 with self-attention masked. So, it is masked self-attention. The impact for computation is only calculated the left words of token. The illustration of two different self-attention mechanism is shown in Figure 19.
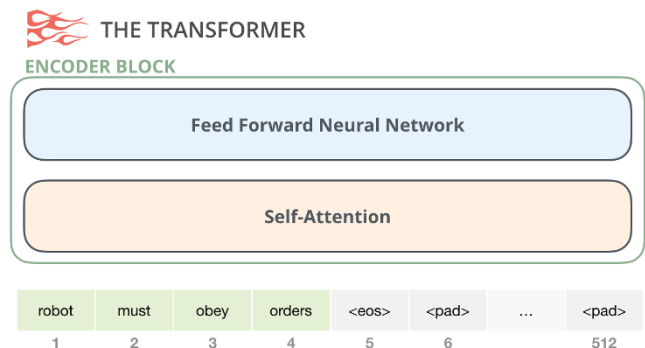
Figure 17: Encoder block of Google BERT [14]
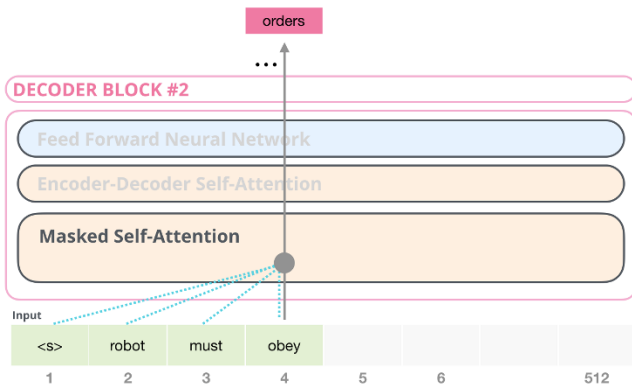
Figure 18: Decoder block of Open AI GPT-2 [14]

The masked self-attention GPT-2 used shows different with self-attention is that the masked one only know the left part of sentence, which include all information on left to predict next word to form the complete sentence even with articles. In addition, the token position encoding and embedding also considered into computation. From the start point of masked self-attention to the neural network to calculate the next word, the process will retain all the weight for the score of final values.
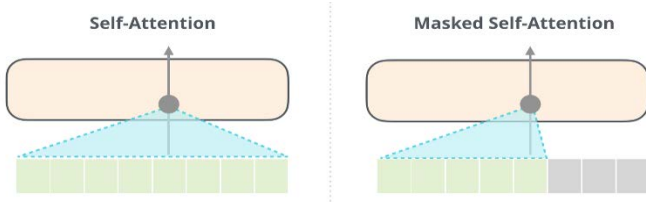
Figure 19: self-attention mechanism[14]
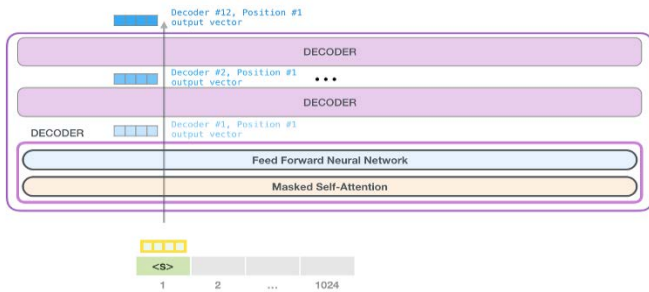
Figure 20: stack in Open AI GPT-2[14]

Figure 20 and 21 shows the process of prediction words. Sending a word to the first transfer learning block means looking up its embedding and adding up the positional encoding vector for position.
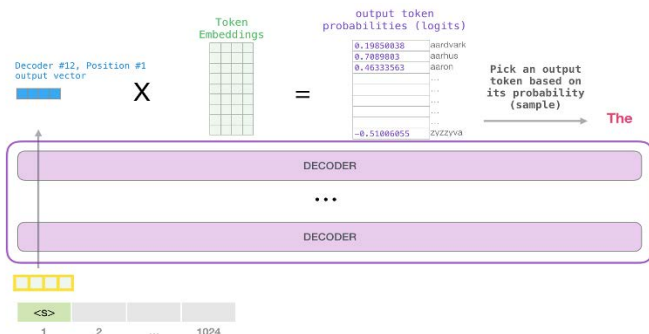
Figure 21: Open AI GPT-2 how to predict word in computation [14]

Initially, end-to-end and transfer learning are designed for machine translation with promising results. Since translation and dialogue are all sequence, more and more users used the same method to dialogue system. For research, only fine-tune with magnitude pre-trained model can obtain ideal language model for special tasks such as text classification, dialogue system etc.

4. System Implementation and Experimental Test

For the implementation of three levels English language learning, for users to access the system anytime, the platform of mini-program at WeChat for the front-end as User Interface to integrate our fine-tuned GPT-2 model based on the Google cloud server will be used in the proposed model. To realize the top layer applications of global architecture, Google speech recognition server is used as the recognizer of pronunciation due to Google has mature technology in speech recognition as compared with others. Level 1 includes the sentence distance comparison of recognized voice and sample sentence. Levels 2 and 3 use generate based language model to provide response to user. Level 2 is topic discussion from daily scenario with 8 topics is the representative of practical English. The highest level is called *free-style conversation* without any limitation for themes. In addition, the Open AI GPT-2 model also could process response to avoid meaningless response with common sense of English syntactic and semantic.

In Figure 22, our *user interface (UI)*, English learning agent has three levels for spoken English training. In this section, system usage and the different interaction performance levels with human users will be presented in this section.
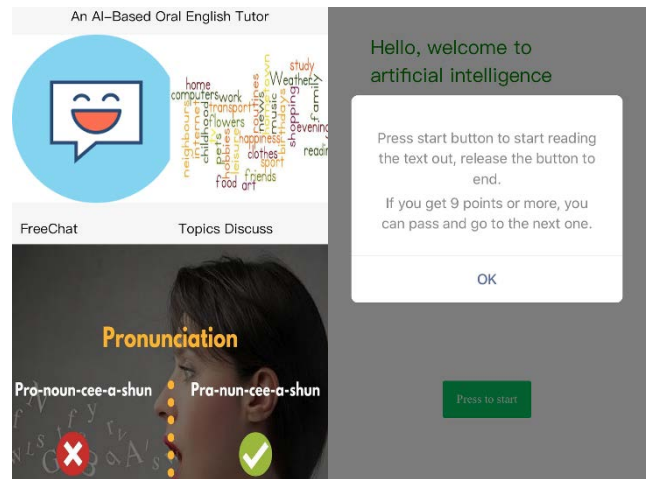
Figure 22: UI of English Learning Chatbot and Level 1 Pronunciation.

Pronunciation is the most important in English learning. This module is designed to practice. Once chatbot can recognize your voice to text and match the text read by user meaning that pronunciation is correct and can be recognized by a native speaker.

Firstly, the rule of pronunciation is the evaluation score shown on system more than 9 where the score comes from a native speaker's recognition set by Google speech recognition server. If user pass the pronunciation meaning that the words from users can be clearly recognized by voice recognition and match with the sample sentence. With the computation of voice recognized words and preinstalled sample sentence by *sequence matcher algorithm*, the essence of distance between two texts can be also calculated to distinguish the difference between user and the native speaker.
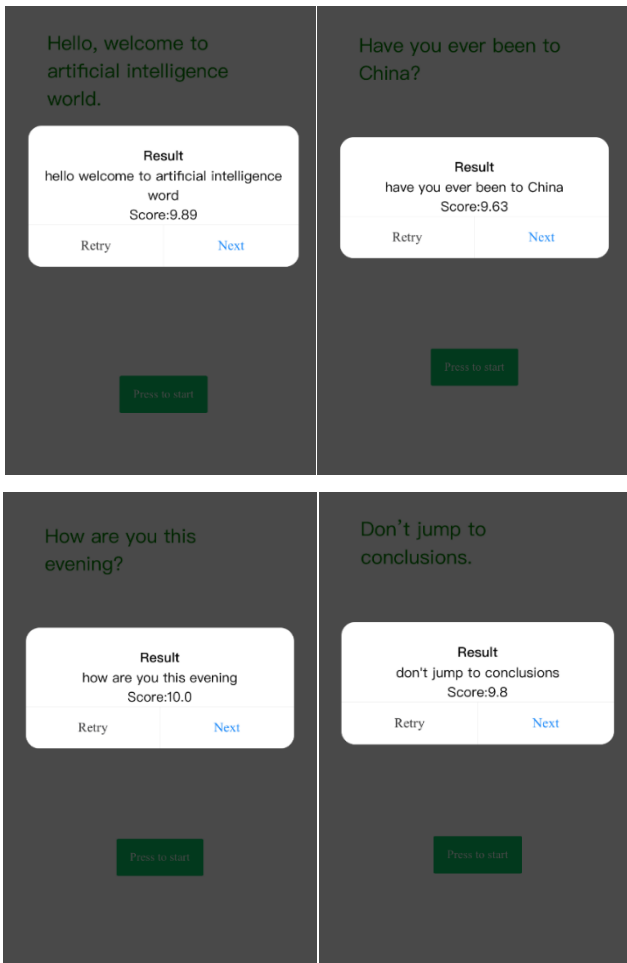
Figure 23: Test of Level 1 Pronunciation

meaningful sentences. So, the conversation contents cannot be limited within the topic as shown on Figure 24.

Figure 24: User Interface designed for our Chatbot Level 2 Topic Discussion.

During test in Figure 23, the first level only set 4 sample sentences as demonstration. After several rounds of test, the system gives higher score when user's voice is slow enough to distinguish every phonetic. Also, the testing or practice environment should be silent to avoid some acute noises. In addition, when different words with approximate phonetics such as 'word' and 'world', it will reduce the score pronunciation test in level 1, but the system may not transform voice to text accurately at level 2s and 3 which will influence the feedback from chatbot.

Second part is the process of a transfer learning model based on an Open-AI GPT-2 architecture so that the pre-trained model is able to generate high-quality text. The fine-tuned model used is *Transfer Transfo* where fine-tuned dataset is an open-domain daily dialogue. During the test, the UI settings are designed for dialogue scenario restricted in daily dialogue such as weather, travel, movies and other daily related activities.

For data training, the character always talks about movie, however, the range of movie contents are limited in US, with the influence of voice recognition. It is difficult to converse about the of movies contents. The model proposed has GPT-2t characteristic which according to the several reminders to write an article. Even the dialogue format limits the sentence length, it also follows the words from users to provide response and take no account of occasionally wrong input and fantastic output. Thus, the whole dialogue is worthy for further research. As mentioned, chatbot only focus on chatter. It has little knowledge base to generate

In Figure 25, the level 2 test shows topics without limitation, in general, when the chatbot greets users, the topic are pre-established and dialogue are also limited in this topic. However, since same language model for levels 2 and 3 are used, the discussion after several rounds would exceed the original one and can shows other key-words related to tutor personality in random. At shown in the movie module, the topic would be changed from movie to music at second one.
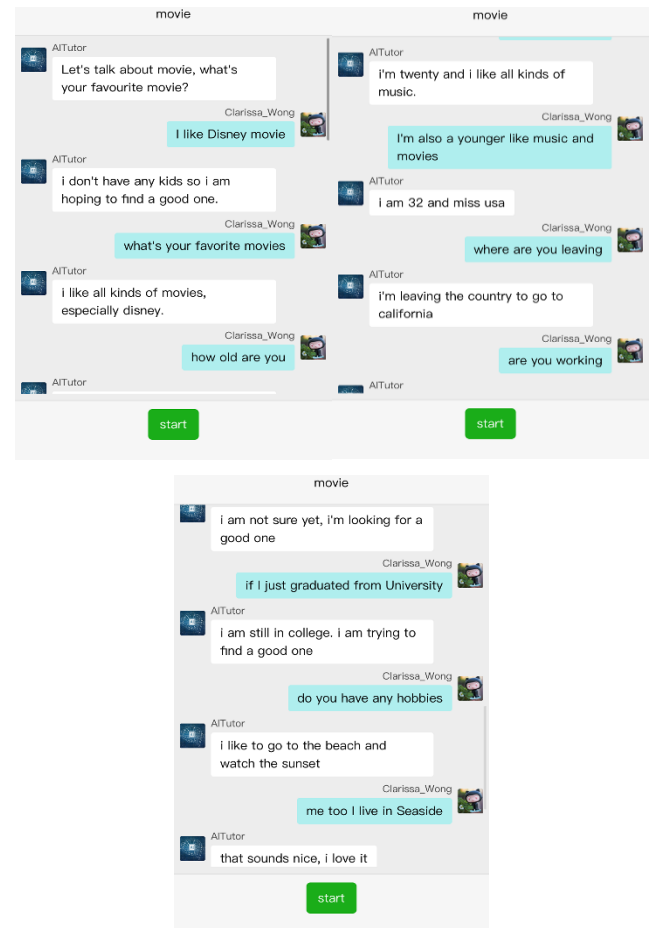
Figure 25: Test of Level 2 Topic Discussion.

At level 3 the same model is used for free chat. To be specific, the system cannot provide sustainable response about the query such as what, why or how. From the visualization of training dataset, the dialogue stayed at information level of the knowledge pyramid. Because the fine-tuned dataset is persona-chat dataset, it cannot provide response like specific *question-answer (QA) system* but prefer to chatbot direction with daily dialogue.

The highest level is free conversation. User and system can have several rounds of dialogue without any restriction . It was found that that the most response is related to fine-tuned dataset contents. Due to the transfer learning model and sequence to sequence model are all data-driven models, with more data used for training, the model has higher language level in syntax in English. Assuming that the text book of an *AI language tutor* is an English teacher, the fine-tuned data will be learned by the machine. After several rounds of test, the tutor uses text book contents to construct sentence content from 200M with high-quality grammar in English by data-driven of 40G text pre-trained.

According to the author of *Transfer Transfo*, the model is equipped with personality of responses generation. During test of Figure 26, every launch of system are different users with different personalities and personal information to chat with users.
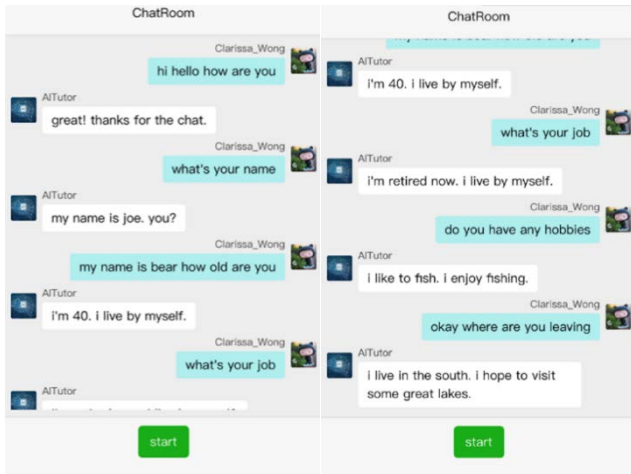
Figure 26: Test of Level 3 Free-Chat with different personality

The daily greetings initially similar to the chat with English teachers in some websites are face to face remote spoken English training. Users cannot guess the next sentence when it is unrelated to any practical topic. Every entry to chatroom has different personality, it is easier to follow the response to continue the subsequent rounds of dialogues.

Different with Level 2 *topic discussion*, Level 3 shows nothing is required to follow in UI. Users could converse anything with chatbot. However, in Figure 27, we can see the response is also limited in the corpus of fine-tune data even though the language model has 40G pre-trained corpus. Level 3 dialogue test results as above showed that English speaking practice with system, the contents cover normally personality with hobbies like music and reading at beginning . When the *AI tutor* ask the key-word 'work', the content will correspond to change and tie to some special words which is related to the key words. Also the contents will be changed from open domain chitchat to another topic proposed by the machine. In this test, "*animal*" and its related words from tutor can also be found at the fine-tune dataset. It was noted that when

output consists of semantic of relations amount among different words.
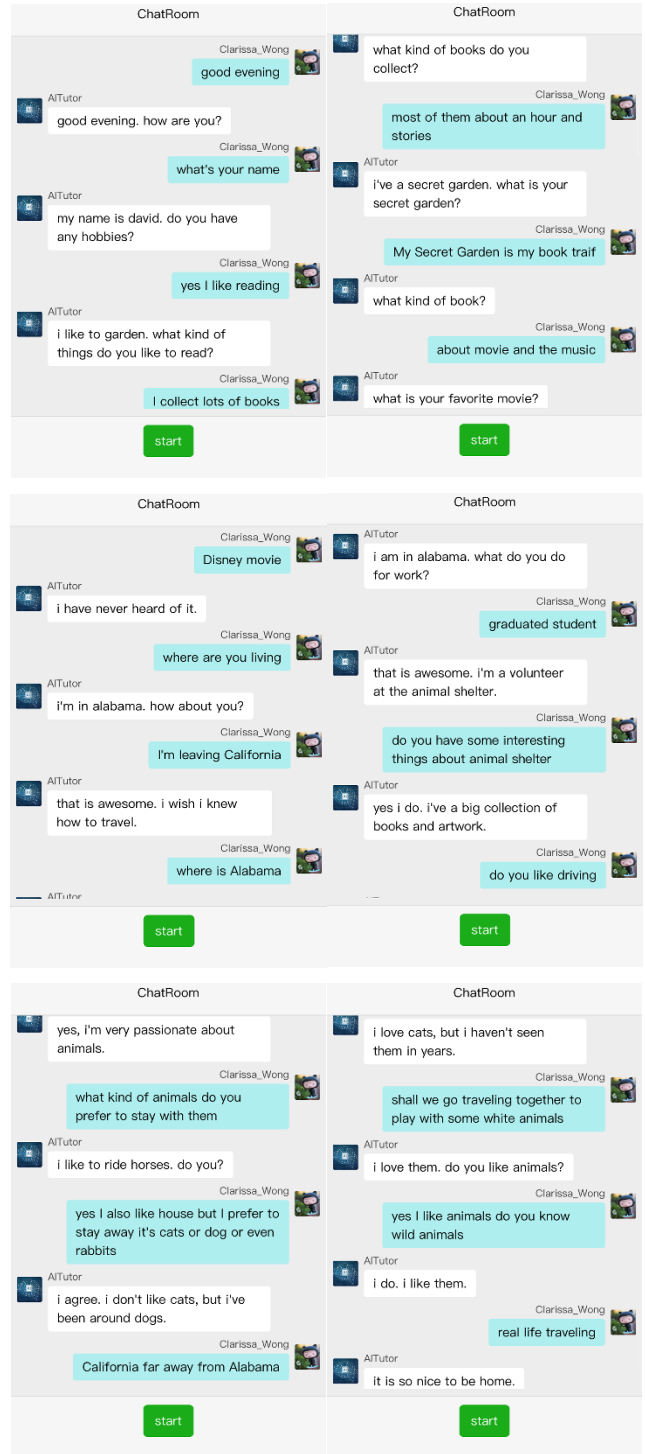
Figure 27: Test of Level 3 Free-Chat

5. Conclusion

In this paper, we used *transfer learning*, an advanced AI technology focusing on technology implementation and system construction to design and implement an *English language chatbot* which is cross-discipline project being proposed the first time and used by real-world university students for learning English round the clock. After around one year improvement, the stable back-end

supports the test with different users 24 hours. Our contribution and the system function shows:

- AI technologies in chatbot field can be well used and combine with the real-world learning needs, not limited in the usage only for personal assistant.

- With magnitude dataset and fine-tune corpus for machine against the AI model, it can generate the meaningful and good syntax response to imitate human response. It overcomes the problem with rule-based or knowledge search engine type chatbot relies on human to prepare the preliminary work, such as label or key-words matching construction. These two methods also range used by ongoing industrial chatbot.

For business aspects, especially in English learning market, there are many applications with splendid *user interface (UI)* and plenty of users declared that their back-end technology uses AI technology. However, during investigation and test commercial apps, the non-mature but with high-expectation neural network model required and relied on dataset quality. For users and system design, the advantages with learning language systematically combines three levels of phonetic, syntactic and semantic which are in both linguistic and the application of *natural language processing* of AI ecosystem.

Based on Turing test theory and *strong AI*, human-machine interaction agent is the main theme of the system. It is proficient in both grammar and knowledge grounds. An Open GPT-2 has already solved the problem to correct error character generated by RNN with pre-trained big data in grammar [3].

6. Future Work

Due to lack of hardware and server, our system performance and capability requires further improvement. At present, 8 GPU with GTX1080Ti on Tencent cloud is used to fine-tune GPT-2 with 345M objective data to make the language model with more meaningful entities to build a QA-based system. However, the computing capacity cannot reach the required magnitude. Also, the official GPT-3 is unavailable for download modification on our GPT-2 language model is continuous until the release of GPT-3. For further research, system optimization for system display is as follows:

- Find a higher capacity server for project implementation

- Optimize UI with better users' experiences and extend the three levels exercises with more options.

- Use 8 V100 GPU to train the largest GPT-2 with existing dataset to compare performance.

Acknowledgment

References

[1] T. Wolf, V. Sanh, J. Chaumond, C. Delangue, "TransferTransfo: A Transfer Learning Approach for Neural Network Based Conversational Agents," (ii), 2019.

[2] R.S.T. Lee, Artificial Intelligence in Daily Life, 2020, doi:10.1007/978-981-15-7695-9.

[3] N. Shi, Q. Zeng, R. Lee, "Language Chatbot-The Design and Implementation of English Language Transfer Learning Agent Apps," in 2020 IEEE 3rd International Conference on Automation, Electronics and Electrical Engineering, AUTEEE 2020, 403–407, 2020, doi:10.1109/AUTEEE50969.2020.9315567.

[4] Nazar Kvartalnyi, Gamification and Simulation in Education and Corporate Learning, Https://Inoxoft.Com/Gamification-and-Simulation-in-Education-and-Corporate-Learning/, 1, 2020.

[5] M. Qiu, F.L. Li, S. Wang, X. Gao, Y. Chen, W. Zhao, H. Chen, J. Huang, W. Chu, "AliMe chat: A sequence to sequence and rerank based chatbot engine," ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), **2**, 498–503, 2017, doi:10.18653/v1/P17-2079.

[6] M. Ghazvininejad, C. Brockett, M.W. Chang, B. Dolan, J. Gao, W.T. Yih, M. Galley, "A knowledge-grounded neural conversation model," 32nd AAAI Conference on Artificial Intelligence, AAAI 2018, 5110–5117, 2018.

[7] 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, **2017**-**Decem**(Nips), 5999–6009, 2017.

[8] L. Zhou, J. Gao, D. Li, H.Y. Shum, "The design and implementation of xiaoice, an empathetic social chatbot," Computational Linguistics, **46**(1), 53–93, 2020, doi:10.1162/COLI_a_00368.

[9] C. Agents, "applied sciences Human Annotated Dialogues Dataset for Natural Conversational Agents," 2020.

[10] S. Hoppe, M. Toussaint, "Qgraph-bounded Q-learning: Stabilizing Model-Free Off-Policy Deep Reinforcement Learning," 2020.

[11] D. Ham, J.-G. Lee, Y. Jang, K.-E. Kim, "End-to-End Neural Pipeline for Goal-Oriented Dialogue Systems using GPT-2," **2**(1), 583–592, 2020, doi:10.18653/v1/2020.acl-main.54.

[12] J.B. Tenenbaum, C. Kemp, T.L. Griffiths, N.D. Goodman, "How to grow a mind: Statistics, structure, and abstraction," Science, **331**(6022), 1279–1285, 2011, doi:10.1126/science.1192788.

[13] N. SHI, Q. Zeng, R. Lee, "The Design and Implementation of Language Learning Chatbot with XAI using Ontology and Transfer Learning," (Dm), 305–323, 2020, doi:10.5121/csit.2020.101124.

[14] Jay Alammar, The Illustrated GPT-2 (Visualizing Transformer Langua ge Models), Https://Jalammar.Github.Io/Illustrated-Gpt2/, 1, 2019.

Analysis of Grid Events Influenced by Different Levels of Renewable Integration on Extra-large Power Systems

Christoph Rüeger[*], Jean Dobrowolski, Petr Korba, Felix Rafael Segundo Sevilla

School of Engineering, Zurich University of Applied Sciences, IEFE, Winterthur, 8401, Switzerland

A R T I C L E   I N F O

A B S T R A C T

In this work, the impact of implementing a large amount of decentralized renewable energy sources (RES) of different scales on an extra-large power grid is investigated. Three scenarios are created, substituting 10%, 20%, and 30% of the conventional energy production by RES. For this purpose, the initial dynamic model of Continental Europe in combination with the industrial power system application DIgSILENT PowerFactory was used. In order to compare the behavior of different applied scenarios, a performance index was developed to evaluate and rank the effects of network disturbances by means of time-domain simulations. The performance index was designed based on three different criteria that analyze the oscillatory content and thus, the severity of a given event. The initial power flow of the dynamic model was identified as a limiting factor for the integration of RES, therefore two additional power flows were developed following an innovative procedure. Through the methodologies mentioned above, it was found that Turkey is the most sensitive to such changes, which are amplified by increasing implementation of RES and often lead to inter-area oscillation.

1   Introduction

The release of Al Gore's film "An Inconvenient Truth" in 2006 as a campaign to educate people about environmental awareness in the USA or the protest and notable criticism of the Swedish environmental activist Greta Thunberg to world leaders, that started in 2015, are good examples of events that have helped to draw people's attention on global warming problems, in particular among the young population. Pollution, climate change, and damage to the earth's ozone layer are some of the issues on the public agenda of governments around the world. Confinement, the considerable decrease in land and air mobility forced by the ongoing pandemic caused by the COVID-19 virus [1] has demonstrated how a sudden change in human behavior can have a positive impact on our environment. With lock-downs all over the world, major industrial processes at a standstill, lack of massive events, and a home-office policy, the daily routine of millions of people have been affected and might never be the same again. In world capitals such as New York, the measured carbon emissions dropped up to 50%, and satellite images have also shown a decrease in NO_2 (nitrogen dioxide) all over Europe, particularly in northern Italy, Spain, and England [2]. Even if it could be proved that the current pandemic situation was a short-term phenomenon, environmental effects have shown

that extreme measures provide positive results to reverse global warming. Moreover, it has been also shown that these extreme measures have a direct impact on the way of generating electricity in the upcoming years. In Switzerland, for instance, after a parliament vote and approval of a new referendum, a revision in the legislation entered into force in 2018 in the form of an amendment to the current Energy Strategy 2050, which is a document known as the New Energy Act [3]. The new reform considers measures to increase energy efficiency, measures to increase the use of renewable energy sources (RES) and withdrawal from nuclear energy. Furthermore, the implementation of further and more ambitious measures to the energy sector have not been limited only within the Swiss borders and similar measures have been implemented in parallel throughout the European Union. In December 2019, the European Green Deal [4] was released, containing 50 measures that, similar to those in Switzerland, intent to push the integration of RES in the electrical system and to promote an emissions trading system. An important point to highlight from this document, is the definition of nuclear power plants as an unsustainable source of energy, leading to nuclear phase-out across Europe in the future. As a result of the aforementioned measures, two main challenges for Transmission System Operators (TSOs) in Europe have envisaged namely 1) the decentralization of RES and 2) the exclusion of large

[*]Corresponding Author: Christoph Rüeger, Technikumstrasse 9, CH-8401 Switzerland, +41 (0) 58 934 69 06, christoph.rueeger@zhaw.ch

rotating masses from the system, presenting a loss of inertia into the system. Regarding the first challenge, a widespread decentralization represents a reshuffle of the energy flow in the transmission system as a result of the installation of RES in remote locations with weak interconnections. With respect to the second problem, existing rotating masses play an important role in the stability of the system as a source of inertia. Switzerland has been privileged with its geography, allowing it to produce more than half of his annual consumption from alpine hydro-power, which also contribute with a non-trivial amount of inertia [5]. However, not every country can benefit from this particular type of energy generation. Nevertheless, it is a fact that load and demand profiles on power systems are changing and thus, its impact must be assessed in order to manage the security of the supply [6, 7]. Contrarily, blackouts like in 2003 in Italy will happen more often [8]. Motivated by these questions, the current conditions of the European transmission power grid and the expected effects on the system as a result of the integration of the aforementioned measures are analyzed in this paper.

Table 1: Number of elements in the dynamic model

Code	Lines	Buses	Loads	Generators
Albania (AL)	193	339	110	77
Austria (AT)	123	104	40	31
Bosnia & Herz.(BA)	312	294	164	37
Belgium (BE)	178	140	36	52
Bulgaria (BG)	787	798	419	77
Switz. (CH)	244	193	82	88
C. Rep. (CZ)	106	288	76	113
Germany (DE)	3378	3939	859	898
Denmark (DK)	250	397	66	71
Spain (ES)	1338	1385	646	496
France (FR)	2599	2665	991	1564
Greece (GR)	1133	1312	367	128
Croatia (HR)	334	329	171	73
Hungary (HU)	94	120	37	28
Italy (IT)	817	1264	341	458
Luxembourg (LU)	41	38	11	12
Montenegro (ME)	80	94	35	18
Macedonia (MK)	146	163	85	25
Nether. (NL)	905	1031	244	178
Poland (PL)	988	648	199	140
Portugal (PT)	365	506	87	158
Romania (RO)	1194	1171	654	185
Serbia (RS)	619	553	303	62
Slovenia (SI)	111	230	65	73
SK Slovakia (SK)	52	48	18	82
Turkey (TR)	1943	4888	1245	1022
Europe (EU)*	9	316	26	1
Total	18339	23253	7377	6147

N_i = Elements not labeled for a specific country.i

The presented studies are based on the initial dynamic model of Continental Europe [9] developed by the European Network of

Transmission System Operators (ENTSO-E), which is the more comprehensive representation of the European transmission power grid available today and that has been extensively studied in previous projects [10].

This paper is organized as follows, as background, it is worth mentioning that this document describes an extension of [11], which is based on two additional results in order to give more insight to the main work. In the first section, the Lyapunov exponent is presented [12] and is used as preliminary stability evaluation of the most severe grid events already reported in [13]. Additionally, the characteristics of the dynamic model are provided in detail, as well as a ranking of events in terms of severity. In the subsequent section, three variations of the power system model are fully described. Considering different future scenarios, different levels of RES integration were evaluated [14]. Subsequently, the effects of faults (introduced on Section 2) for each model variation are investigated. Then, the results are analyzed for all simulated scenarios and ranked using the algorithm based on the Lyapunov exponent. From the time series resulting from Root Mean Square (RMS) simulations, special attention is paid to low-frequency fluctuations (in the range of 0.1-1Hz) also known as inter-area oscillations, which are the main problem to solve in extra-large power grids [15, 16]. One disadvantage of the dynamic model used for the analysis in this document is the fact that there is only a single initial power flow available. This constraint plays an important role in the investigation of the stability of the system. To overcome this limitation, new power flow scenarios were developed using an algorithm reported in Section 3.3. Finally, a possible solution will be discussed. The progressive replacement of the rotating masses and the consequent decrease of the grid inertia[17] makes the power grid more vulnerable to faults especially in the eastern part of the grid. Thus, three different approaches are presented using wide-area monitoring to control the inter-area oscillations. New in these methods is the architectural design of the controllers.

2 Background

2.1 The Dynamic Model of Continental Europe

The dimensions of the model, with more than 55,000 interconnected elements distributed over 27 grids (see table 1), is one of the main characteristics and made the selected system very challenging to work with. It is the most comprehensive representation of the continental European power grid, which was developed by ENTSO-E and that we have used to demonstrate the effectiveness of the proposed algorithms. In the following section, multiple variations of the topology of the system will be introduced but first, the base case is presented so it later can be used as reference for the comparison. To do so, different types of contingencies or faults were simulated and the response of the system was analyzed by means of frequency measurements on randomly selected buses at the high voltage level in all countries [9]. Given the size of the model, the number of contingencies to be investigated and thus the number of simulations to be performed, was narrowed down to 184 based on the seven criteria displayed in figure 1.
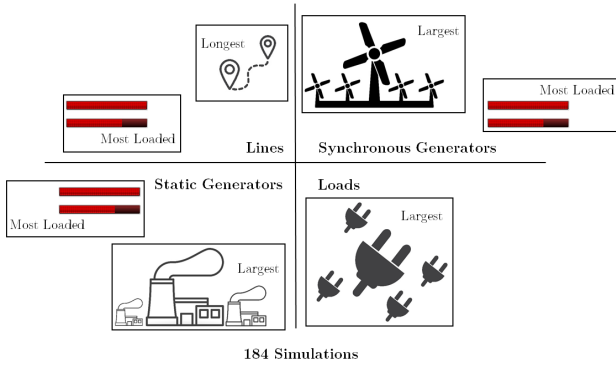
Figure 1: Event selection based on the seven criteria.

2.2 Severity evaluation and ranking of most dangerous events

For the base case, 184 simulations were performed representing one fault of an asset at the time, e.g. 3-phase short circuit on a line or the sudden disconnection of a generator or load. Following the event, the frequency was measured for a period of 20 seconds at 27 busbars randomly selected (one bus of each country). In order to evaluate the impact on the network and to rank the severity, the measurement results of each simulation were compared. To do this, a performance index that analyzed three aspects of the time series was first developed as follows:

1. Measuring the damping of the oscillations based on the Lyapunov Exponent [18] with f_j^i as the frequency in the country i at the time j (1),

$$\lambda_k^i = \frac{1}{N\Delta t} \times \sum_{m=1}^{N} \log \left(\frac{\left| f_{(k+m)\Delta t}^i - f_{(k+m-1)\Delta t}^i \right|}{\left| f_{(m)\Delta t}^i - f_{(m-1)\Delta t}^i \right|} \right) \qquad (1)$$

$$\alpha_{le} = \frac{\sum_{i=1}^{27} \alpha_{le}^i}{27} \qquad (2)$$

and where α_{le} in Equ. (2) is the average value of each simulation.

2. The next aspect was the analysis of the average amplitude of the oscillation in all countries [19] (3)

$$\alpha_{am} = \frac{\sum_{i=1}^{27} max \left(|f_0 - min(f_i)|, |f_0 - max(f_i)| \right)}{27} \qquad (3)$$

with the nominal frequency f_0 before the event and f_i being the frequency in the i-th country.

3. Finally, the rate of change of frequency or RoCoF for each country was used as last indicator [20]

$$RoCof = \frac{\partial f(t)}{\partial t} = \frac{f_m - f_{m-1}}{\Delta t} \qquad (4)$$

$$\alpha_{de} = \frac{\sum_{i=1}^{27} max \left(RoCoF_i \right)}{27} \qquad (5)$$

where in equation (4) f_m is the frequency at time t and similarly to the Lyapunov Exponent, α_{de} in Eq. (5) is the average value across the 27 countries.

In order to avoid that an individual criteria had more impact on the final index value than others, the individual indicators (α_{le}), (α_{am}) and (α_{de}) were multiplied by a weighting coefficient as shown in Equ. (6).

$$\alpha = \alpha_{le} * \omega_{le} + \alpha_{am} * \omega_{am} + \alpha_{de} * \omega_{de} \qquad (6)$$

The final outcome, resulting from equ. (6) is a scalar that can be used to evaluate the severity of a given contingency.

A partial result is depicted in Figure 3, which displays the severity index for the base case, where the normalized contribution of the criterions defined on Eq. (1-3) to the final index are depicted in blue, red, and green, respectively. The worst event has a magnitude of +1 and as the magnitude decreases its value, the lower the relevance of the subsequent event. Fig. 3 displays the worst events (ranked 175 - 184) and their geographical location is depicted on Figure 2. The inter-area oscillations caused by the sudden disconnection of a large synchronous machine in Turkey, are displayed in Figure 4 for a selected number of countries across Europe, where the severe effects of this fault can be clearly observed. Table 2 summarized the events displayed in Fig. 2 and it can be observed that 9 out of 10 of the most critical faults are related to the sudden loss of a large synchronous generation, in contrast to only one event that was caused by the trip of an important transmission line. It is worth noticing that all machines listed in Table 2 are located in Eastern Europe.
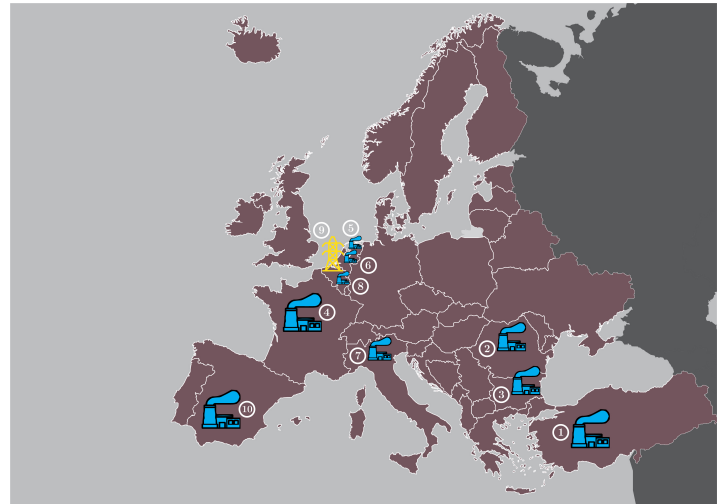
Figure 2: Location of the 10 worst events according to performance value.

Table 2: Performance Value of the ten worst events

Ranking	Country	Type	Perfomrance Value
184	TR	Sync. Machine	1
183	RO	Sync. Machine	0.903
182	BG	Sync. Machine	0.886
181	FR	Sync. Machine	0.699
180	NL	Sync. Machine	0.571
179	NL	Sync. Machine	0.569
178	IT	Sync. Machine	0.559
177	BE	Sync. Machine	0.518
176	BE	Line	0.517
175	ES	Sync. Machine	0.512
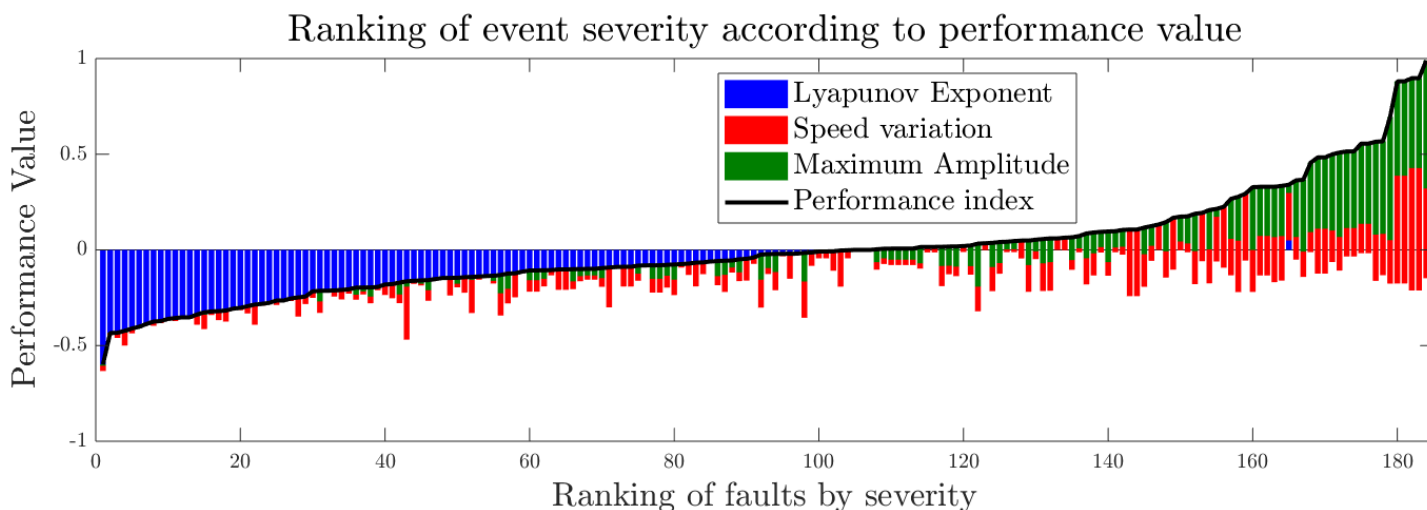
Figure 3: Ranking of events by severity.

3 Stability effects of RES integration

An insight of the results, obtained after the analysis of the base case, confirms the accuracy of the model in comparison with the real system. In this case, the base case corresponds to the conditions of the European power grid in 2016, approximately. In the next step, a modified grid is investigated, where conventional synchronous machines were replaced by decentralized renewable energy sources. In this form, the original dynamic model was topologically changed three times in order to represent different future scenarios.
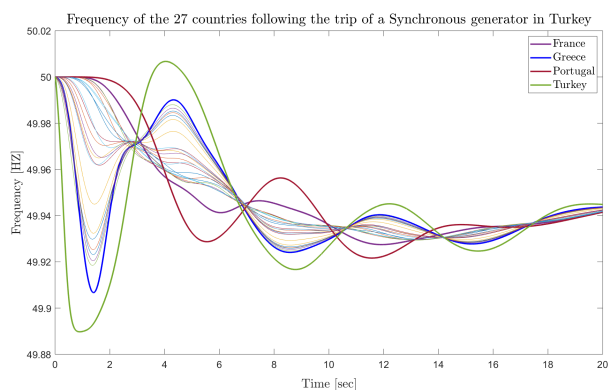
Figure 4: Inter-Area oscillations (0.125Hz) affirmed by the results of a tripped synchronous generator in Turkey.

3.1 Implementation of different levels of RES

Three future scenarios were developed with a systematic increase of RES. First, 10% of the country's conventional energy production was replaced with RES, this scenario is referred to as *S1* in the subsequent sections. Then, a 20% RES penetration was developed and labelled as scenario 2 or *S2* and finally a third scenario *S3*, with 30% of RES integration. It is important to underline that the procedure to achieve an increase in the different levels of RES is not trivial nor simple. In DIgSILENT PowerFactory, the power system software application in which the model was built, is possible to combine

renewable energy with existing models in different ways. However, since the focus of this work was to analyze stability phenomena, an accurate representation of renewable devices, namely one to one component representation, was omitted. In other words, there is no distinction between the replacement of conventional generation between Wind, PV or any other technology. The key idea was to substitute rotating masses with static elements and thus, indirectly affect the inertia of the system. Therefore, renewable assets were modelled as static generators connected via static converters [21].

Furthermore, a decentralized renewable generation is assumed [14], which has a direct influence on the topology of the different variants. In this form, synchronous generators representing nuclear power plants, coal-fired power plants and other conventional production, were replaced by several smaller static generators located on multiple locations. The voltage level at which these large quantities of RES were connected played also an important role. In [22] it is recommended to connect new elements at busbars where a load is also attached. In our case, the model has more than 23'000 buses with more than 7'000 loads. The new static generators integrated into the system were selected to operate in the 15kV voltage level.

From the countries considered in the model, not all of them are represented with the same level of detail or do not have a pronounced infrastructure as in Switzerland, Germany or Turkey. As a result, some countries were modelled with fewer amount of synchronous machines. This fact was taken into consideration in the development of scenarios 1, 2 and 3, respectively. Only countries modelled with less than 20 synchronous machines were excluded from the modification. Consequently, only 13 countries were modified and their names are listed in Table 3. The topology of these countries was changed individually to create the new system variations, leading to 39 model variations (13 countries times 3 different variations). In addition, Table 3 displays the amount of sync. machines (N_k) existing in each country and the total amount of active power in MW that these machines contribute. Similarly, this table indicates the static generators (N_i) required in order to replace the energy produced by the conventional machines N_k.

Technically, the 39 variants mentioned above were created and the share of active power to be replaced was calculated for the

respective country using an optimization procedure. Next, all synchronous machines available in a given country were analyzed in order to find the fewer amount of synchronous generators to be substituted by RES. Then, the selected machines were replaced by randomly distributed static generators to compensate the same amount of active power. Given that this task involved the modification of hundreds of elements, the procedure was automatized using python. In the same way, all simulations performed (7'000 resulting from analyzing 184 events at the time for 39 variations), were analyzed and compared, using the python interface of DIgSILENT PowerFactory and the results are presented in the following subsection.

Table 3: Transmutations in the selected countries

Country	RES	Scenario 1 10%		Scenario 2 20%		Scenario 3 30%	
	N_i	N_k	MW	N_k	MW	N_k	MW
Austria (AT)	40	4	409	4	1092	4	1640
Switz. (CH)	80	4	712	4	1458	4	2186
C. Rep. (CZ)	76	4	653	4	1304	4	1955
Germany (DE)	111	6	5301	13	8012	21	14046
Spain (ES)	646	4	2620	6	5241	9	7861
France (FR)	956	4	6600	11	13185	16	19784
Greece (GR)	363	4	726	4	1069	6	1604
Italy (IT)	319	4	2025	9	3211	16	4030
Nether. (NL)	166	4	1538	4	3076	5	4630
Poland (PL)	185	4	1125	4	2252	5	3376
Belgium (BE)	36	4	981	4	1959	4	2940
Slovenia (SI)	59	4	155	4	517	4	613
Turkey (TR)	316	5	3641	14	7278	23	10918

N_i = number of static generators, N_k = number of synchronous generators.

3.2 Simulation results using the initial Power Flow

In this section, the 7'000 simulations resulting from the 184 events on the different levels of RES integration are compared against the results of the 184 events on the original model [13] and presented in three parts as follow:

i. *Scenario 1 (S1) vs. Base case (B)*. The 13 variations implementing 10% of RES in different countries are compared against the original model.

ii. *Scenario 2 (S2) vs. Base case (B)*. The 13 variations implementing 20% of RES in different countries are compared against the original model.

iii. *Scenario 3 (S3) vs. Base case (B)*. The 13 variations implementing 30% of RES in different countries are compared against the original model.

The comparison was performed as follows:

a) The measure used to compare scenarios is the absolute value of the difference between the frequencies of each variation and the frequencies of the original model, more details are provided next.

b) The main results are presented in the form of 3D-plots where:

– The $x-$ axis depicts the country where the RES were deployed. For simplicity, only the country codes are presented and the country names are listed in Table 3.

– The $y-$ axis provides the simulation number from 1 to 184, which is refers to different types of events (discussed in Section 2) and is summarized below:

* 1-27   severe synchronous machine events
* 28-57   most loaded synchronous generators
* 58-84   most loaded lines
* 85-111   events related to loads
* 112-138   longest lines
* 139-161   largest static generators
* 162-184   most loaded static generators events

– the $z-$ axis corresponds to the sum of the frequency differences, which are given by:

$$\begin{pmatrix} \sum_{i=1}^{N_G} \left| f_{B_1}^i - f_{AT_1}^i \right| & \cdots & \sum_{i=1}^{N_G} \left| f_{B_{N_E}}^i - f_{AT_{N_E}}^i \right| \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{N_G} \left| f_{B_1}^i - f_{TR_1}^i \right| & \cdots & \sum_{i=1}^{N_G} \left| f_{B_{N_E}}^i - f_{TR_{N_E}}^i \right| \end{pmatrix} \quad (7)$$

where N_G is the number of measurements per simulation (27), N_E is the number of events (184), $f_{B_j}^i$ represents the frequency measured in the i-th country after the j-th event for the base case model [13] and $f_{AT_j}^i$ is the frequency measured in the i-th country (Austria in this case) after the j-th event for the n-th scenario (S_n) of renewable penetration.

c) After the frequency differences have been estimated, the stability indexes introduced on equation (6) are calculated for every model variation and then compared against the base case [13] using a heat map. In these graphical representation, the $y-$ axis represents the country where the RES were implemented and the $x-$ axis provides the simulation number related to the type of event as in the 3D plot. The colors of the heat map represent the relation among stability indexes, which are given as:

$$\begin{pmatrix} \alpha_{AT_1} - \alpha_{B_1} & \cdots & \alpha_{AT_{N_E}} - \alpha_{B_{N_E}} \\ \vdots & \ddots & \vdots \\ \alpha_{TR_1} - \alpha_{B_1} & \cdots & \alpha_{TR_{N_E}} - \alpha_{B_{N_E}} \end{pmatrix} \quad (8)$$

where α_{B_i} is the stability index of the base case for the i-th event and α_{AT_i} is the stability index in the j-th country (Austria in this case) after the i-th event for the n-th scenario (S_n) of renewable penetration. It is important to note that:

– An increase in stability is represented by negative values in matrix (8) as a consequence of RES penetration.

– Positive values imply a larger deviation from the base case, leading to a negative impact in terms of stability as a result of the RES implementation.

Figure 5: Comparison between scenario 1 (implementation of 10% RES) and the original model by the absolute difference of the frequency.

Figure 6: Heatmap visualizing the change in the stability index between scenario 1 (implementation of 10% RES) and the original model.

3.2.1 Scenario 1 (10%) vs. Base Case

Figure 5 shows how the compensation of 4-6 synchronous generators by RES affects the frequency. For this purpose, the results from the base case are compared with those from scenario 1. This corresponds to a 10% increase in energy generated by renewables, which means up to 163 Mvar of reactive power and up to 6.6 GW of active power, depending on the model variation. As mentioned in equation (7), the absolute difference of the frequencies serves as a metric for the comparison. By spanning a surface and plotting the frequency on the z-axis in Figure 5, unevenness indicates deviations from the base case. Apart from a few outliers and small peaks, no major differences can be detected. This observation is backed up by the heat map in Figure 6. Here, changes in the stability index are shown in color. Coloring in the dark blue area represents a negative value and thus indicates a more stable grid. The positive effect of an increment of 10% of RES in some countries, particularly in Greece (GR), is explained by the stability index on equation (6). As predicted on [15], reducing inertia produces an improvement of the rate of change of frequency (RoCoF) and at the same time produces an increase of the damping and the natural frequency. For most of the studied events in Greece, the increase in the RoCoF is lower than the increase on the damping, leading to a smaller value of α.

At this point, an event in the model variation Switzerland shall be briefly discussed. This event can be seen in Figures 5 triggering the largest outlier, as well as in Figure 6 as a dark red colouration. This is not a fault, but a disconnection of a synchronous generator, which is directly connected to the interconnection between Switzer-

land and Italy. In the ENTSO-E model, this element is subject to market restrictions, so the active power flowing through this line must remain constant. Additionally, one synchronous generator decommissioned in Switzerland was geographically close to this line. As result, the reallocation of the power was changed causing a large variation in the frequency.

3.2.2 Scenario 2 (20%) vs. Base Case

In this comparison, the same pattern is found in relation to the previous case. However, Figure 7 displays more frequency variations, particularly in Germany, Spain, France and Turkey. In the corresponding heat map depicted in Figure 8, it can be observed that Turkey happens to be the most sensible country to the RES variation and this is reflected by the number of times where the magnitude α exceed the value of 5, which is denoted with a turquoise/green color on the heat map. Similarly, it can be seen that 20% of renewable penetration has a positive impact in some events, especially in Greece and Italy because also, in this case, reducing the inertia has a lower negative influence on the RoCoF, than a positive one on the damping and the natural frequency in the system. Finally, it is important to note that similarly as in the previous case, some random negative events in Greece and Italy occur as a result of reallocation of the power after an event with respect to a trip of a static generator and a highly loaded line.
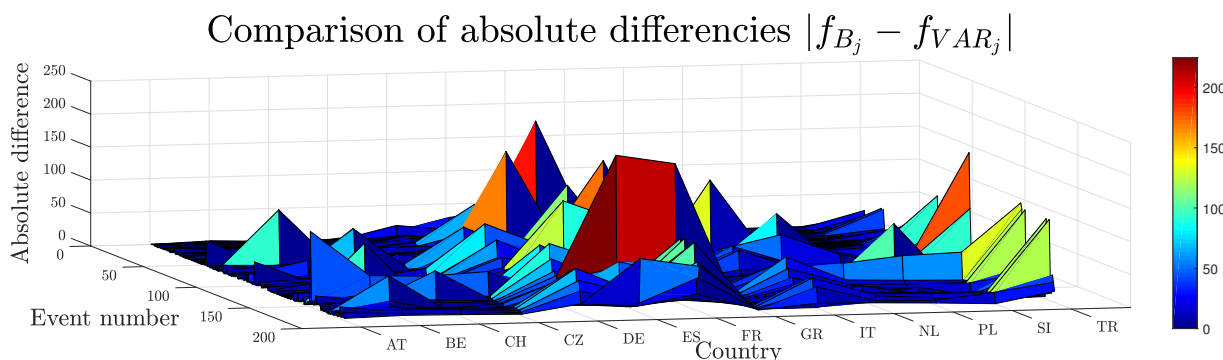
Figure 7: Comparison between scenario 2 (implementation of 20% RES) and the original model by the absolute difference of the frequency.
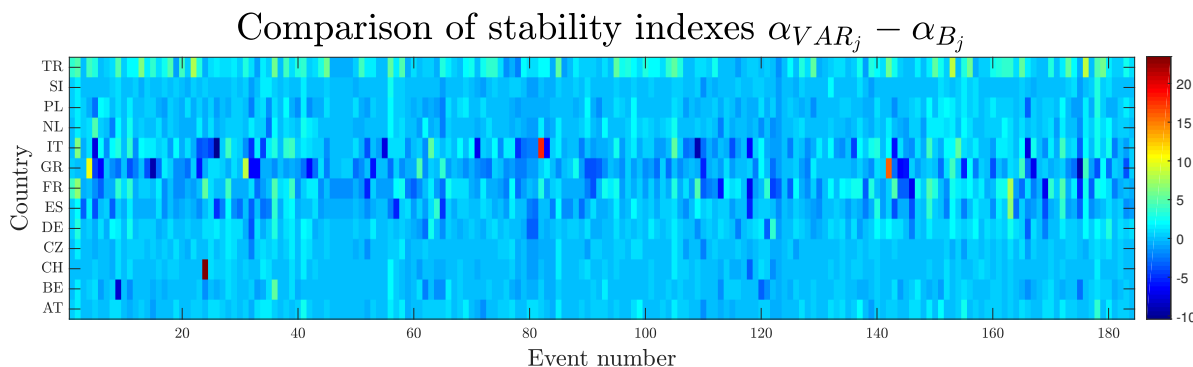
Figure 8: Heat map visualizing the change in the stability index between scenario 2 (implementation of 20% RES) and the original model.

3.2.3 Scenario 3 (30%) vs. Base Case

The largest topological changes were made in the last scenario and the effects on the models can be seen very well in Figure 9 and 10. When looking at the absolute frequency difference in Figure 9, there are clearly more irregularities than in subsection 3.2.1 and 3.2.2. What was only slightly apparent in scenario 2 can now be confirmed in this comparison. Events in countries resembling Germany, France, Spain, Italy and Turkey are most affected by the aggressive implementation of RES and show clear differences to the base case. It is noticeable that in all cases Turkey reacts most sensitively to the network changes. However, it is a fact that Turkey is the country with the highest change in share of renewable energy. The problems in Turkey can be explained by the results shown in Figure 12, where 27 measurements resulting after a severe event in one of the most loaded static generators in Turkey, are displayed (in blue) and the results are compared against the base case (in red). Through these results, it is evident that although the damping of the oscillations is improved for the 30% RES implementation (lines in red), there is a significant degradation of the amplitude and RoCoF compared to the base case (lines in blue). This results in electromechanical oscillations caused by groups of generators oscillating against each other (west versus east), which were already identified using the stability index developed in (6).

3.3 Development of additional power flows

It can be concluded from the results in the previous subsection that massive integration of RES can change the stability of the European power grid. In particular, Turkey is more prone to present stability constraints following RES integration and after analyzing the initial power flow, it can be seen that around 13% of the production in Central Europe comes from Turkey and that 70% of the global production is produced in the western part of Europe namely in France, Germany, Spain, Italy, and the Netherlands. The current generation distribution in the system anticipates the results obtained and opens the following question: does the redistribution of the current ratio play an important role in the stability of the system? The answer is not trivial and one potential solution is to redistribute the current generation and repeat a similar procedure presented in Section 3.2. This solution represented to create new power flow scenarios and to do so, an iterative algorithm to automatically randomly calculate a new set of initial conditions was developed and is explained in the next steps:

1. Calculation of the total active power in the model contributed from all participating countries. It is worth mentioning that in this initial approach the reactive power was not change.

2. Random reallocation of the active power among the countries of the model, respecting the maximum active power provide from each country.

3. Perform a load flow calculation in PowerFactory with the new conditions

4. Check if power flow converge (a solution to the new operating conditions exist). If solution exist: algorithm is finish, otherwise return to step 2.
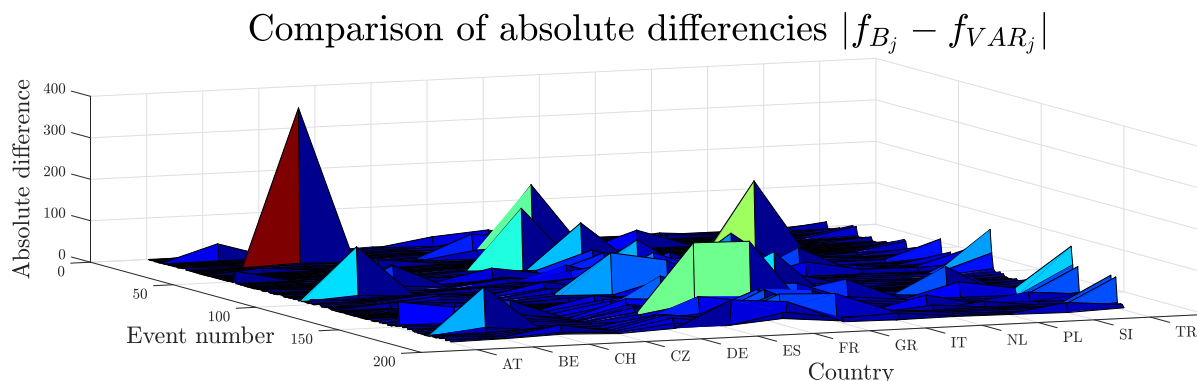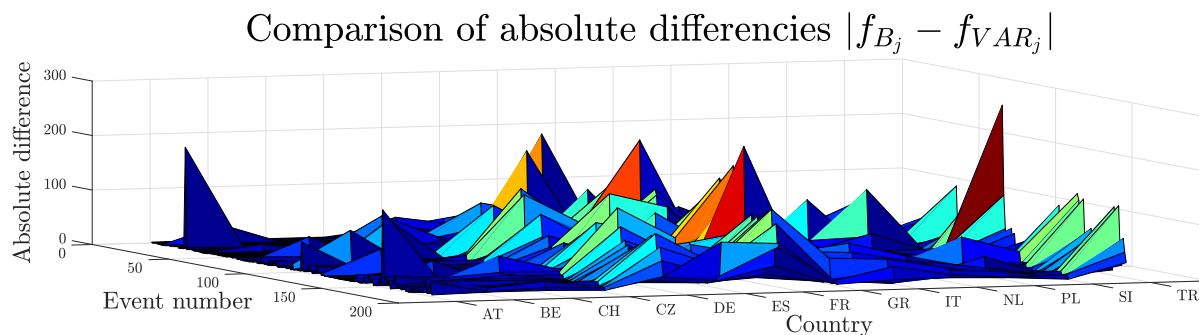
Figure 9: Comparison between scenario 3 (implementation of 30% RES) and the original model by the absolute difference of the frequency.
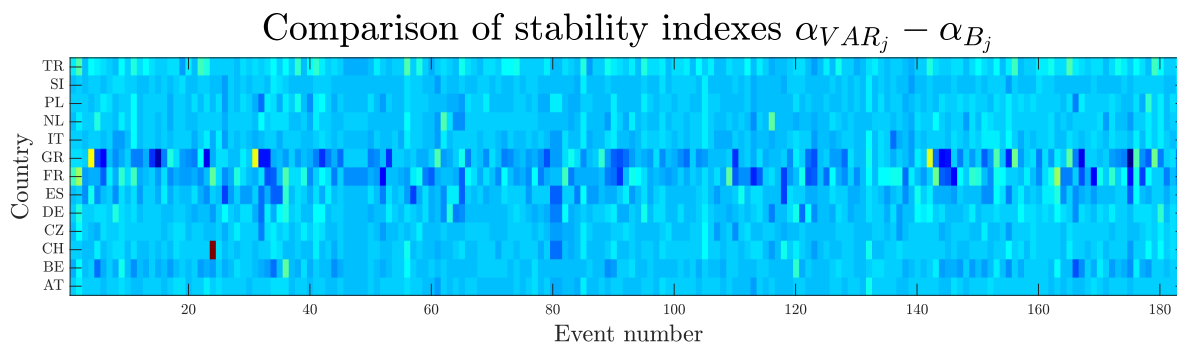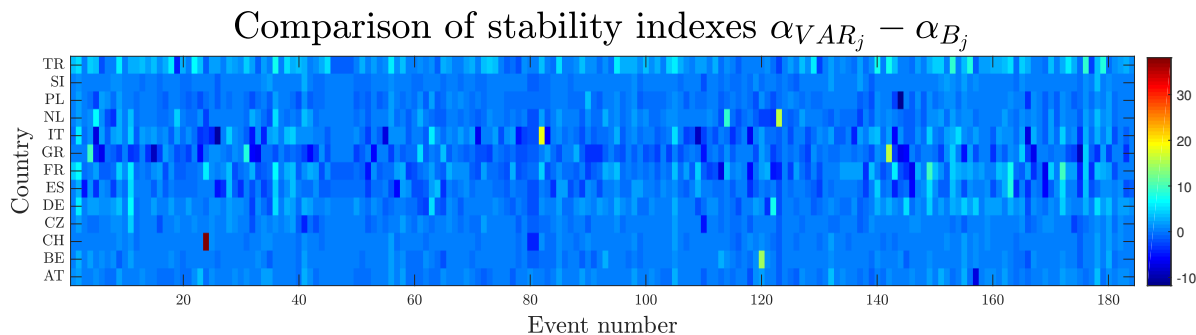

Figure 10: Heatmap visualizing the change in the stability index between scenario 3 (implementation of 30% RES) and the original model.

Given the complexity of the model, more than 500 iterations were required in order to obtain a new power flow. In this document, we present two additional initial power flows and the difference between the base case and the new operating points are depicted in Figure 11. In the first developed power flow (represented in orange bars), the shares of Turkey have been significantly reduced almost by half and the shares of Italy and Spain (Western-Europe) have been increased. In contrast, in the second developed power flow (green bars) the shares of Turkey have significantly increased and the share of Western-Europe has been reduced. After the two power flows were defined, the 184 simulations were performed again. The simulation results have been compared against the simulations performed using the original power flow and the results are summarized in Figure 13. In this figure, the evolution of the stability index (α) between the base case (presented in Section 2), and the RES integration for the new power flows is presented. For each simulation j, the performance index of the base case (α_{B_j}) and of the variation (α_{VAR_j}) are calculated and the results are compared in Figure 13, with the x-axis representing $\sum(\alpha_{VAR_j} - \alpha_{B_j})$. To get a better idea of these results, we can refer to Figure 3. In the same way, this ranking is created for each power flow and the performance value is added up.

The results displayed in Figure 13 show that the power flow plays an important role in the stability of the system and that incrementing the balance difference between the western and the eastern part of Europe is not realistic from a stability point of view. Particularly, it was noticed that the most critical event identified in Section 3 (a trip of a synchronous generator in Turkey) is even worst in the new power flow. On the contrary, the second power flow shows that the system becomes more robust or in other words,

produces a positive impact on the stability of the system. From the simulations results, it can be concluded that reducing the gap between the East and West seems to be a reasonable solution to improve the stability of the system.

4 Discussion & Future Work

The methodologies and results presented in this document represent a substantial extension to the works presented before [11]. The main objective was to investigate the dynamic behavior of the European power grid before and after the implementation of decentralized renewable energy sources. Thus, in a first part, a performance index was developed with which the measurement results from the numerous simulations could be compared. The final results are covered in the second part of this work. Three phases of RES penetration were investigated; 10%, 20%, and 30% increase respectively, resulting in 39 variations of the initial dynamic model. The 30% variation case caused the greatest amount of stability issues in the system in comparison with the 10% and 20% variation case. In general, the results of 10% and 20% more RES differ very marginally. However, what was already identified as a potential problem in the original model has been confirmed by the three tested and analyzed scenarios; Turkey is one of the most sensitive countries. As the percentage of implemented RES increases, the effects of the outage of a certain sync. machine (see Table 2) are getting worse. These negative effects are namely inter-area oscillations and are not rare, yet to be avoided in such super grids. With the increase of interconnections over longer distances, the risk of groups of generators in different geographical locations starting to oscillate against each other is
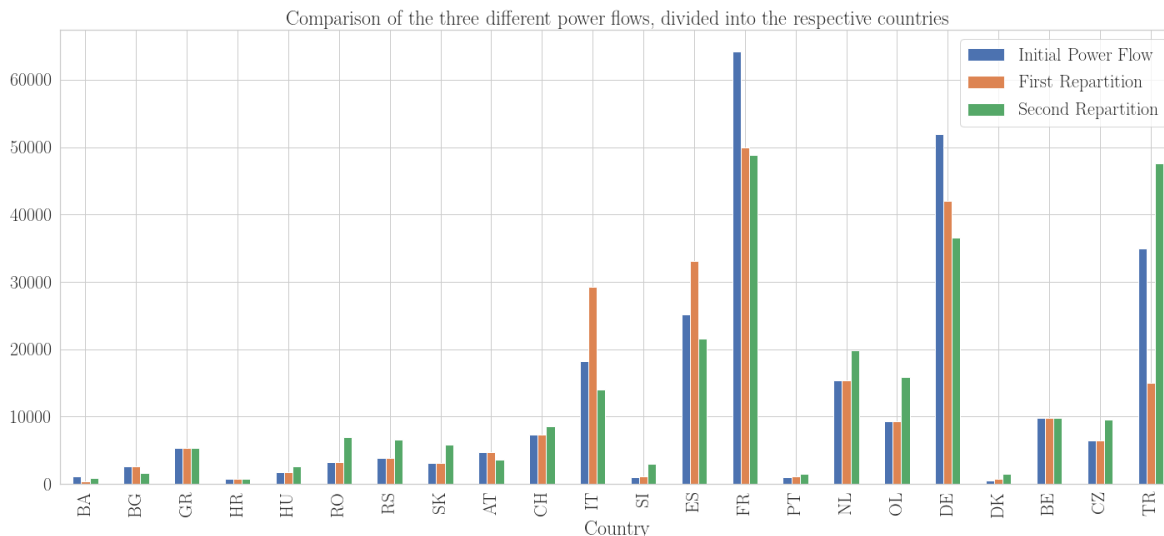
Figure 11: Representation of the three working power flows, divided into the different countries.

increasing respectively, creating electromechanical oscillations over the entire grid (in our case west to east). These events can limit the power transfer capacity and can even lead to a system collapse and blackouts if left undamped [23, 24].
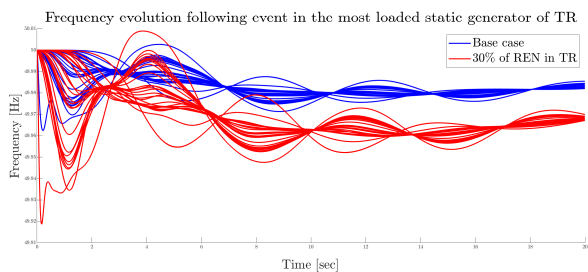
Figure 12: Frequency comparison of grid response to the trip of the most loaded static generator located in Turkey.
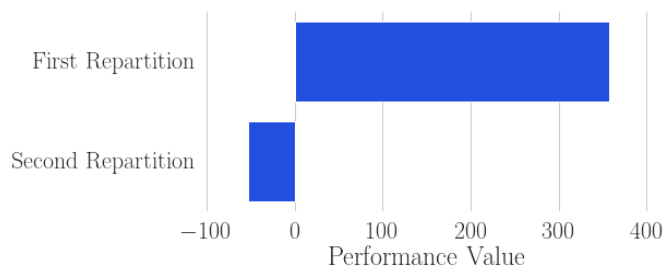
Figure 13: Stability index evolution depending on the initial power flows.

Power system stability problems are a topic that has been discussed for many years. Most approaches refer to the adjustment of Power System Stabilizers (PSS). However, considering the size of the European grid, dynamically adjusting all necessary PSS seems rather unrealistic. With the help of Wide Area Monitoring (WAM) and Wide Area Control (WAC), a different approach could be taken. By synchronously measuring the phase angle of the node voltages and currents via satellite-based Phasor Measurement Units (PMUs) distributed over a large area, a large system can be easily monitored and inter-area oscillations can be quickly detected. Wide Area Daming Controllers (WADC) play a key role in WAMs and help to damp such oscillations faster. The design of the controller is crucial to how efficiently such a WADC works.

In a next step, different methodologies for the design of such WADCs will be investigated and tested based on benchmark models. In a future step, these controllers will be implemented and tested in a real laboratory experiment. An adapted variant of Kundur's two area system has already been successfully developed in the smart grid laboratory in Winterthur, Switzerland [25]. Furthermore, electromechanical oscillations have been reproduced as well. This laboratory test will then be extended accordingly to enable the use of WACD. But even a test laboratory will eventually reach its limits in terms of size and complexity of a power & transmission grid. Therefore, these test trials are to be complemented by implementing the initial European model into Opal-RTs real-time simulator, coupled with control-in-the-loop to validate the effectiveness of a WADC and address main challenges like measurement noise & errors and control noise during the communication.

Conflict of Interest   The authors declare no conflict of interest.

References

[1] M. E. E. Zowalaty, S. G. Young, J. D. Järhult, "Environmental impact of the COVID-19 pandemic – a lesson for the future," Infection Ecology & Epidemiology, **10**(1), 1768023, 2020, doi:10.1080/20008686.2020.1768023.

[2] S. Saadat, D. Rawtani, C. M. Hussain, "Environmental perspective of COVID-19," Science of The Total Environment, **728**, 138870, 2020, doi: https://doi.org/10.1016/j.scitotenv.2020.138870.

[3] Swiss Federal Office of Energy SFOE, "Energy Strategy 2050 once the new Energy Act is in Force," PowerPoint Presentation, 2011. [Online].

[4] European Comission, "The European Green Deal," PDF Document, 2019. [Online].

[5] B. Schaefli, P. Manso, M. Fischer, M. Huss, D. Farinotti, "The role of glacier retreat for Swiss hydropower production," Renewable Energy, **132**, 615–627, 2019, doi:https://doi.org/10.1016/j.renene.2018.07.104.

[6] European Network of Transmission System Operator for Electricity (ENTSO-E), "EC Green paper: A 2030 Framework for climate and energy policies, an ENTSO-E consuiltation response," PDF Document, 2014. [Online].

[7] Swissgrid, "Overview of Ancillary Services," PDF Document, 2019. [Online].

[8] A. Berizzi, "The Italian 2003 blackout," in IEEE Power Engineering Society General Meeting, 2004., 1673–1679 Vol.2, 2004, doi:10.1109/PES.2004.1373159.

[9] European Network of Transmission System Operator for Electricity (ENTSO-E), "ENTSO-E Dynamic Model of Continental Europe," Operation Report, 2019. [Online].

[10] F. R. Segundo Sevilla, P. Korba, K. Uhlen, E. Hillberg, G. Lindahl, W. Sattinger, "Evaluation of the ENTSO-E initial dynamic model of continental Europe subject to parameter variations," in 2017 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT), 1–2, 2017, doi:10.1109/ISGT.2017.8085986.

[11] C. Rüeger, J. Dobrowolski, P. Korba, F. R. Segundo Sevilla, "Stability Effects After Massive Integration of Renewable Energy Sources on Extra-Large Power Systems," in 2020 IEEE PES Transmission Distribution Conference and Exhibition - Latin America (T D LA), 1–6, 2020, doi:10.1109/TDLA47668.2020.9326181.

[12] S. Dasgupta, M. Paramasivam, U. Vaidya, V. Ajjarapu, "Real-time monitoring of short-term voltage stability using PMU data," in 2014 IEEE PES General Meeting — Conference Exposition, 1–1, 2014, doi:10.1109/PESGM.2014.6939198.

[13] C. Rüeger, J. Dobrowolski, P. Korba, F. R. Segundo Sevilla, "Lyapunov Exponent for Evaluation and Ranking of the Severity of Grid Events on Extra-Large Power Systems," in 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe), 1–5, 2019, doi:10.1109/ISGTEurope.2019.8905564.

[14] European Network of Transmission System Operator for Electricity (ENTSO-E), "TYNDP 2016 Scenario Development Report," PDF Document, 2015. [Online].

[15] P. Kundur, Investigation of low frequency inter-area oscillation problems in large interconnected power systems, Canadian Electrical Association, Montreal, Canada, 1993.

[16] V. Vittal, N. Bhatia, A. Fouad, "Analysis of the inter-area mode phenomenon in power systems following large disturbances," in 1991., IEEE International Sympoisum on Circuits and Systems, 982–985 vol.2, 1991, doi: 10.1109/ISCAS.1991.176528.

[17] P. H. J. Nardelli, M. S. Baptista, C. A. Pomalaza-Raez, P. Cardieri, M. Latva-aho, "Network models for the structures of modern electric power grids," CoRR, **abs/1401.0260**, 2014.

[18] S. Dasgupta, M. Paramasivam, U. Vaidya, V. Ajjarapu, "Real-time monitoring of short-term voltage stability using PMU data," in 2014 IEEE PES General Meeting — Conference Exposition, 1–1, 2014, doi:10.1109/PESGM.2014.6939198.

[19] C. Sourkounis, P. Tourou, "Grid Code Requirements for Wind Power Integration in Europe," Conference Papers in Science, **1**, 9, 2013, doi:https://doi.org/10.1155/2013/437674.

[20] M. Bollen, M. Häger, "Power quality: interactions between distributed energy resources,the grid, and other customers," Leonardo Energy, **1**, 2005.

[21] DIgSILENT, Gomaringen, Germany, DIgSILENT PowerFactory 2018 User Manual, 2018.

[22] M. Reza, "Stability analysis of transmission system with high penetration of distributed generation," 2006.

[23] A. P. G. A. (APG), "Almost Blackout on January 8, APG Publishes Video-Simulation," Website, 2021. [Online].

[24] E. N. of Transmission System Operator for Electricity (ENTSO-E), "Analysis of ce inter-area oscillations of 1st december 2016," PDF Document, 2017. [Online].

[25] D. Baltensperger, J. Dobrowolski, A. Obushevs, F. R. Segundo Sevilla, P. Korba, "Scaling Version of Kundur's Two-Areas System for Electromechanical Oscillations Representation," in 2020 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM), 243–249, 2020, doi:10.1109/SPEEDAM48782.2020.9161960.

Model Reduction H_∞ Finite Frequency of Takagi-Sugeno Fuzzy Systems

Rim Mrani Alaoui[1], Abderrahim El-Amrani[2,*], Ismail Boumhidi[2]

[1]*Higher Institute of Engineering and Business, ISGA, Fez, 30000, Morroco*

[2]*LISAC Labs, Sidi Mohamed Ben Abdellah University, Fez, 30000, Morocco*

A R T I C L E   I N F O

A B S T R A C T

The daily treats model reduction finite frequency (FFMR) design for Takagi Sugeno (T S) systems. This work is to FFMR design in such a way whether augmented model is steady get a reduced H_∞ index in FF areas with noise is established as a prerequisite. To highlight the importance of suggested process, a practical application has been made.

1 Introduction

Fuzzy models T-S [1] call for wide observation of various practical industrial applications, mainly as the recognized T-S samples actually approximate nonlinear shapes. The essential characteristic on the sample TS is its general estimation of a nonlinear function. There are large number of results of literature that treat the globally difficulties utilizing the TS fuzzy samples, see [2]–[7].

The our existing sources on model reduction problem and disturbances are based on the whole full frequency (EF) area, which will give several types of model reduction design [8]–[13]. However, most practical industrial applications work in a FF domain. So far, a few applications have been made [14]–[19]. Thus, for this we will present new approaches to solve these problems.

The primary goal of our work is to define a fuzzy model reduction of discrete Model over FF ranges such a way that the augmented model is steady get a reduced H_∞ index in FF areas with disturbance is established as a prerequisite. we have also presented an example of simulation in order to exemplify the efficiency of the suggested method.

Notations :

- ”T” : Matrix transposition

- ”$*$” : Matrix symmetry

- $M > 0$: matrix M is positive

*A. EL-AMRANI & Email : abderrahim.elamrani@usmba.ac.ma

- **sym**$[\mathcal{D}]$: $\mathcal{D} + \mathcal{D}^*$

- **He**$[\mathcal{D}]$: $\frac{\mathcal{D}+\mathcal{D}^*}{2}$

2 Problem statement

2.1 System formulation

Envisage the nonlinear model presented by :
Rule x: IF $\zeta_1(\mu)$ is T_1^i,... $\zeta_n(\mu)$ is $\mathbf{T}_n^i$ THEN

$$
\begin{aligned}
\mathbf{z}_{\mu+1} &= \mathbf{M}_x\mathbf{z}(\mu) + \mathbf{N}_x\mathbf{d}(\mu) \\
\mathbf{w}(\mu) &= \mathbf{J}_x\mathbf{z}(\mu)
\end{aligned}
\tag{1}
$$

$$\tag{2}$$

with $\mathbf{z}(\mu)$ is the input; $\mathbf{w}(\mu)$ is deliberate output; $\mathbf{M}_x$, $\mathbf{N}_x$ and $\mathbf{J}_x$ are system parametres; $\zeta_1(\mu)$,..., $\zeta_n(\mu)$ the premise variables. $\mathbf{d}(\mu)$ is a known disturbance signal located in a following FF areas

$$
\varrho = \left\{ \tilde{\varrho} \in \mathbb{R}| \quad \tilde{\varrho}_1 \le \tilde{\varrho} \le \tilde{\varrho}_2; \quad \tilde{\varrho}_1, \tilde{\varrho}_2 \in [-\pi, +\pi], \right.
\tag{3}
$$

We describe the nonlinear system (1) employ singleton fuzzifer, center-average and inference product by the following relation :

$$
\begin{aligned}
\mathbf{z}_{\mu+1} &= \mathbf{M}(\mathbf{l})x(\mu) + \mathbf{N}(\mathbf{l})\mathbf{d}(\mu) \\
\mathbf{w}(\mu) &= \mathbf{J}(\mathbf{l})\mathbf{z}(\mu)
\end{aligned}
\tag{4}
$$

where

$$\mathbf{M(l)} = \sum_{x=1}^{q} \mathbf{l}_x \mathbf{M}_x; \quad \mathbf{N(l)} = \sum_{x=1}^{q} \mathbf{l}_x \mathbf{N}_x;$$

$$\mathbf{J(l)} = \sum_{x=1}^{q} \mathbf{l}_x \mathbf{J}_x \tag{5}$$

In this work, a fuzzy MR function is represented by :

Rule x: if $\zeta_1(\mu)$ is $\mathbf{T}_1^x$... $\zeta_n(\mu)$ is $\mathbf{T}_n^x$ then

$$\hat{\mathbf{z}}(\mu + 1) = \hat{\mathbf{M}}_x \hat{\mathbf{z}}(\mu) + \hat{\mathbf{N}}_x \mathbf{d}(\mu)$$
$$\hat{\mathbf{w}}(\mu) = \hat{\mathbf{J}}_x \hat{\mathbf{z}}(\mu) \tag{6}$$

with $\hat{\mathbf{z}}(\mu)$ is state MR vector; $\hat{\mathbf{w}}(\mu)$ is the output MR function, $\hat{\mathbf{M}}_x$, $\hat{\mathbf{N}}_x$ and $\hat{\mathbf{J}}_x$ are parameters should exist defined.

We get defuzzified for system (6) as following :

$$\hat{\mathbf{z}}(\mu + 1) = \hat{\mathbf{M}}(\hat{\mathbf{l}})\hat{\mathbf{z}}(\mu) + \hat{B}(\hat{\mathbf{l}})\mathbf{d}(\mu)$$
$$\hat{\mathbf{w}}(\mu) = \hat{\mathbf{J}}(\hat{\mathbf{l}})\hat{\mathbf{z}}(k) \tag{7}$$

with

$$\hat{\mathbf{M}}(\hat{\mathbf{l}}) = \sum_{x=1}^{q} \hat{\mathbf{l}}_x \hat{\mathbf{M}}_x, \quad \hat{B}(\hat{\mathbf{l}}) = \sum_{x=1}^{q} \hat{\mathbf{l}}_x \hat{\mathbf{N}}_x,$$

$$\hat{\mathbf{J}}(\hat{\mathbf{l}}) = \sum_{x=1}^{q} \hat{\mathbf{l}}_x \hat{\mathbf{J}}_x \tag{8}$$

Consider $\nabla(\mu) = \begin{bmatrix} \mathbf{z}(\mu) \\ \hat{\mathbf{z}}(\mu) \end{bmatrix}$, $\mathbf{g}(\mu) = \mathbf{w}(\mu) - \hat{\mathbf{w}}(\mu)$. Then, the error system is as follows:

$$\nabla(\mu + 1) = \bar{\mathbf{M}}(\mathbf{l}, \hat{\mathbf{l}})\nabla(\mu) + \bar{\mathbf{N}}(\mathbf{l}, \hat{\mathbf{l}})\mathbf{d}(\mu)$$
$$\mathbf{g}(\mu) = \bar{\mathbf{J}}(\mathbf{l}, \hat{\mathbf{l}})\nabla(\mu) \tag{9}$$

where

$$\bar{\mathbf{M}}(\mathbf{l}, \hat{\mathbf{l}}) = \begin{bmatrix} \mathbf{M(l)} & 0 \\ 0 & \hat{\mathbf{M}}(\hat{\mathbf{l}}) \end{bmatrix}; \quad \bar{\mathbf{N}}(\mathbf{l}, \hat{\mathbf{l}}) = \begin{bmatrix} \mathbf{N(l)} \\ \hat{\mathbf{N}}(\hat{\mathbf{l}}) \end{bmatrix};$$

$$\bar{\mathbf{J}}(\mathbf{l}, \hat{\mathbf{l}}) = \begin{bmatrix} \mathbf{J(l)} & -\hat{\mathbf{J}}(\hat{\mathbf{l}}) \end{bmatrix} \tag{10}$$

Definition 2.1 *[20] Consider $\beta > 0$, if the following inequality is verified:*

$$\sum_{\mu=0}^{\infty} \mathbf{g}^T(\mu)\mathbf{g}(\mu) \le \beta^2 \sum_{\mu=0}^{\infty} \mathbf{d}^T(\mu)\mathbf{d}(\mu) \tag{11}$$

$$\tag{12}$$

From Parseval function [21], we get :

$$\int_{-\pi}^{\pi} \mathbf{g}^T(\varrho)\mathbf{g}(\varrho)d\varrho \le \beta^2 \int_{-\pi}^{\pi} \mathbf{d}^T(\varrho)\mathbf{d}(\varrho)d\varrho \tag{13}$$

$$\tag{14}$$

We express the question of this work by : we design an appropriate fuzzy MR system (7) such that a error model is well-posed, stable satises the FF index:

$$\int_{\varrho \in \varrho} \mathbf{g}^T(\varrho)\mathbf{g}(\varrho)d\varrho \le \beta^2 \int_{\varrho \in \varrho} \mathbf{d}^T(\varrho)\mathbf{d}(\varrho)d\varrho \tag{15}$$

2.2 Preliminaries

Lemma 2.2 *[22] Let $\Phi \in \mathbb{R}^m$, $\mathcal{L} \in \mathbb{R}^{m \times m}$ and $\mathcal{S} \in \mathbb{R}^{p \times m}$. Then, The following equations are the same:*

- $\Phi^* \mathcal{L} \Phi < 0, \forall \Phi \ne 0 : \mathcal{S}\Phi = 0$
- $\mathcal{S}^{\perp T} \mathcal{L} \mathcal{S}^{\perp} < 0$
- $\exists \theta \in \mathbb{R} : \mathcal{L} - \theta \mathcal{S}^T \mathcal{S} < 0$
- $\exists \mathcal{D} \in \mathbb{R}^{m \times p} : \mathcal{L} + \mathcal{D}\mathcal{S} + \mathcal{S}^T \mathcal{D}^T < 0$

Lemma 2.3 *Error system (10) is stable and FF in (15) is fulfilled, on condition that there are $\mathcal{B}$, $0 < \mathcal{C}$, satisfying*

$$\begin{bmatrix} \bar{M}(l, \hat{l}) & \bar{N}(l, \hat{l}) \\ I & 0 \end{bmatrix}^T \begin{bmatrix} \mathcal{B} & e^{j\tilde{\varrho}_3}\mathcal{C} \\ e^{-j\tilde{\varrho}_3}\mathcal{C} & -\mathcal{B} - 2\cos(\tilde{\varrho}_4)\mathcal{C} \end{bmatrix}$$
$$\begin{bmatrix} \bar{M}(l, \hat{l}) & \bar{N}(l, \hat{l}) \\ I & 0 \end{bmatrix} + \begin{bmatrix} \bar{J}^T(l, \hat{l})\bar{J}(l, \hat{l}) & 0 \\ 0 & -\beta^2 I \end{bmatrix} < 0 \tag{16}$$

with $\quad \tilde{\varrho}_3 = \frac{\tilde{\varrho}_2 + \tilde{\varrho}_1}{2}$, $\tilde{\varrho}_4 = \frac{\tilde{\varrho}_2 - \tilde{\varrho}_1}{2}$.

Proof 2.4 *The substantiation of Lemma 2.3 is defined in appendix.*

3 FF performance analysis

Theorem 3.1 *Error model (10) is stable, FF index (15) is fulfilled, on condition that there are X, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{E}$, $\mathcal{Z}$, $\mathcal{T}$ satisfying $\mathcal{C} > 0$, $X > 0$ and*

$$\begin{bmatrix} \daleth_1 & \daleth_2 & \mathcal{Z}\bar{N}(l, \hat{l}) - \mathcal{T}^T & 0 \\ * & \daleth_3 & \daleth_4 & \bar{J}^T(l, \hat{l}) \\ * & * & \daleth_5 & 0 \\ * & * & * & -I \end{bmatrix} < 0 \tag{17}$$

$$\begin{bmatrix} X - \mathcal{Z} - \mathcal{Z}^T & \mathcal{Z}M(l, \hat{l}) - \mathcal{E}^T \\ * & \Psi_1 \end{bmatrix} < 0 \tag{18}$$

with

$$\daleth_1 = \mathcal{B} - \mathcal{Z} - \mathcal{Z}^T;$$
$$\daleth_2 = e^{j\tilde{\varrho}_3}\mathcal{C} - \mathcal{E}^T + \mathcal{Z}\bar{A}(l, \hat{l});$$
$$\daleth_3 = -\mathcal{B} - 2\cos(\tilde{\varrho}_4)\mathcal{C} + \mathcal{E}\bar{M}(l, \hat{l}) + \bar{M}^T(l, \hat{l})\mathcal{E}^T;$$
$$\daleth_4 = \mathcal{E}\bar{N}(l, \hat{l}) + \bar{M}^T(l, \hat{l})l^T(l)$$
$$\daleth_5 = -\beta^2 I + \mathcal{T}\bar{N}(l, \hat{l}) + \bar{N}^T(l, \hat{l})l^T;$$
$$\triangle_1 = -X + \mathcal{E}\bar{M}(l, \hat{l}) + \bar{M}^T(l, \hat{l})\mathcal{E}^T.$$

Proof 3.2 *First, Condition ([17](#)) maybe written as*

$$\begin{pmatrix} \bar{M}(l,\hat{l}) & \bar{N}(l,\hat{l}) \\ I & 0 \\ 0 & I \end{pmatrix}^T [\begin{pmatrix} 0 & \bar{J}(l,\hat{l}) & 0 \\ 0 & 0 & I \end{pmatrix}^T$$

$$\begin{pmatrix} I & 0 \\ 0 & -\beta^2 I \end{pmatrix} \begin{pmatrix} 0 & \bar{J}(l,\hat{l}) & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & I \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} \mathcal{B} & e^{j\varrho_c}C \\ e^{-j\tilde{\varrho}_3}C & -\mathcal{B} - 2cos(\tilde{\varrho}_4)C \end{pmatrix} \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \end{pmatrix}]$$

$$\begin{pmatrix} \bar{M}(l,\hat{l}) & \bar{N}(l,\hat{l}) \\ I & 0 \\ 0 & I \end{pmatrix} < 0 \qquad (19)$$

Denote

$$\mathcal{D} = \begin{bmatrix} \mathcal{E} \\ \mathcal{Z} \\ \mathcal{T} \end{bmatrix}; \quad \theta = \begin{bmatrix} \nabla(\mu+1) \\ \nabla(\mu) \\ d(\mu) \end{bmatrix};$$

$$\mathcal{S} = \begin{bmatrix} -I & \bar{M}(l,\hat{l}) & \bar{N}(l,\hat{l}) \end{bmatrix} \qquad (20)$$

With the help of Lemma [2.2](#), we have

$$\mathcal{D}\mathcal{S} + \mathcal{S}^T\mathcal{D}^T + \begin{bmatrix} 0 & \bar{J}(l,\hat{l}) & 0 \\ 0 & 0 & I \end{bmatrix}^T$$

$$\begin{bmatrix} I & 0 \\ 0 & -\beta^2 I \end{bmatrix} \begin{bmatrix} 0 & \bar{J}(l,\hat{l}) & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & I \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} \mathcal{B} & e^{j\tilde{\varrho}_3}C \\ e^{-j\tilde{\varrho}_3}C & -\mathcal{B} - 2cos(\tilde{\varrho}_4)C \end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \end{bmatrix} < 0 \quad (21)$$

With the help of Lemma [2.2](#), given ([16](#)).
   Consider the Lyapunov equation :

$$\begin{bmatrix} \bar{M}(l,\hat{l}) \\ I \end{bmatrix}^T \begin{bmatrix} X & 0 \\ 0 & -X \end{bmatrix} \begin{bmatrix} \bar{M}(l,\hat{l}) \\ I \end{bmatrix} < 0 \qquad (22)$$

Let

$$\theta = \begin{bmatrix} \nabla(\mu+1) \\ \nabla(\mu) \end{bmatrix}; \quad \mathcal{L} = \begin{bmatrix} X & 0 \\ 0 & -X \end{bmatrix};$$

$$\mathcal{S} = \begin{bmatrix} -I & \bar{M}(l,\hat{l}) \end{bmatrix}; \quad \mathcal{D} = \begin{bmatrix} \mathcal{E}^T & \mathcal{Z}^T \end{bmatrix}^T \qquad (23)$$

Using the conditions in [2.2](#), then

$$\begin{bmatrix} X & 0 \\ 0 & -X \end{bmatrix} + \begin{bmatrix} \mathcal{E} \\ \mathcal{Z} \end{bmatrix} \begin{bmatrix} -I & \bar{M}(l,\hat{l}) \end{bmatrix} +$$

$$\begin{bmatrix} -I & \bar{M}(l,\hat{l}) \end{bmatrix}^T \begin{bmatrix} \mathcal{E} \\ \mathcal{Z} \end{bmatrix}^T < 0 \qquad (24)$$

which is nothing but ([18](#)).

4 FF performance Design

Theorem 4.1 *Error model ([10](#)) is stable, FF index ([15](#)) is fulfilled, on condition that there are $\mathcal{B} = \begin{bmatrix} \mathcal{B}_1 & \mathcal{B}_2 \\ * & \mathcal{B}_3 \end{bmatrix}, C = \begin{bmatrix} C_1 & C_2 \\ * & C_3 \end{bmatrix} >$*

$0, X = \begin{bmatrix} X_1 & X_2 \\ * & X_3 \end{bmatrix} > 0, \check{M}(\hat{l}), \check{N}(\hat{l}), \check{J}(\hat{l}), \mathcal{E}_1, \mathcal{E}_2, \mathcal{Z}_1, \mathcal{Z}_2, \mathcal{T}_1, \mathcal{V},$
satisfying

$$\begin{bmatrix} \bar{\Delta}_{11} & \bar{\Delta}_{12} & \bar{\Delta}_{13} & -\mathcal{Z}_2^T + E\hat{M}(\hat{l}) \\ * & \bar{\Delta}_{22} & \bar{\Delta}_{23} & \hat{M}(\hat{l}) \\ * & * & \bar{\Delta}_{33} & -X_2 + M^T(l)\mathcal{Z}_2^T \\ * & * & * & -X_3 \end{bmatrix} < 0$$

$$(25)$$

$$\begin{bmatrix} \bar{\daleth}_{11} & \bar{\daleth}_{12} & \bar{\daleth}_{13} & \bar{\daleth}_{14} & \bar{\daleth}_{15} & 0 \\ * & \bar{\daleth}_{22} & \bar{\daleth}_{23} & \bar{\daleth}_{24} & \bar{\daleth}_{25} & 0 \\ * & * & \bar{\daleth}_{33} & \bar{\daleth}_{34} & \bar{\daleth}_{35} & J^T(l) \\ * & * & * & \bar{\daleth}_{44} & \bar{\daleth}_{45} & -\hat{J}^T(\hat{l}) \\ * & * & * & * & \bar{\daleth}_{55} & 0 \\ * & * & * & * & * & -I \end{bmatrix} < 0$$

$$(26)$$

with

$$\bar{\Delta}_{11} = \mathcal{B}_1 - \mathcal{E}_1 - \mathcal{E}_1^T; \ \bar{\Delta}_{12} = X_2 - E\mathcal{V} - \mathcal{E}_2^T;$$
$$\bar{\Delta}_{13} = \mathcal{E}_1 A(h) - \mathcal{Z}_1^T; \ \bar{\Delta}_{22} = X_3 - \mathcal{V} - \mathcal{V}^T;$$
$$\bar{\Delta}_{23} = \mathcal{E}_2 M(l); \ \bar{\Delta}_{33} = -X_1 + \mathcal{Z}_1 M(l) + M^T(l)\mathcal{Z}_1^T;$$

$$\bar{\daleth}_{11} = \mathcal{B}_1 - \mathcal{E}_1 - \mathcal{E}_1^T; \ \bar{\daleth}_{12} = \mathcal{B}_2 - E\mathcal{V} - \mathcal{E}_2^T;$$
$$\bar{\daleth}_{13} = e^{j\tilde{\varrho}_3}C_1 - \mathcal{Z}_1^T + \mathcal{E}_1 A(l); \ \bar{\daleth}_{22} = \mathcal{B}_3 - \mathcal{V} - \mathcal{V}^T;$$
$$\bar{\daleth}_{14} = e^{j\tilde{\varrho}_3}C_2 - \mathcal{Z}_2^T + E\tilde{M}(\hat{l});$$
$$\bar{\daleth}_{24} = e^{j\tilde{\varrho}_3}C_3 + \tilde{M}(\hat{l});$$
$$\bar{\daleth}_{15} = \mathcal{E}_1 N(l) + E\tilde{N}(\hat{l}) - \mathcal{T}_1^T;$$
$$\bar{\daleth}_{25} = \mathcal{Z}_2 N(l) + \tilde{M}(\hat{l});$$
$$\bar{\daleth}_{33} = -\mathcal{B}_1 - 2cos(\tilde{\varrho}_4)C_1 + \mathcal{Z}_1 M(l) + M^T(l)\mathcal{Z}_1^T;$$
$$\bar{\daleth}_{34} = -\mathcal{B}_2 - 2cos(\tilde{\varrho}_4)C_2 + M^T(l)\mathcal{Z}_2^T;$$
$$\bar{\daleth}_{35} = \mathcal{Z}_1 N(l) + M^T(l)\mathcal{T}_1^T;$$
$$\bar{\daleth}_{23} = e^{-j\tilde{\varrho}_3}C_2 + \mathcal{E}_2 M(l);$$
$$\bar{\daleth}_{44} = -\mathcal{B}_3 - 2cos(\tilde{\varrho}_4)C_3; \ \bar{\daleth}_{45} = \mathcal{Z}_2 N(l);$$
$$\bar{\daleth}_{55} = -\beta^2 I + sym[\mathcal{T}_1 N(l)];$$

The following parameters as

$$\hat{M}(\hat{l}) = \mathcal{V}^{-1}\check{M}(\hat{h});;$$
$$\hat{N}(\hat{l}) = \mathcal{V}^{-1}\check{N}(\hat{h}); \qquad (27)$$
$$\hat{J}(\hat{l}) = \check{J}(\hat{l})$$

Proof 4.2 *Parameterise slack matrices $\mathcal{E}, \mathcal{Z}, \mathcal{T}$ in Theorem [3.1](#) as*

$$\mathcal{E} = \begin{bmatrix} \mathcal{E}_1 & E\mathcal{V} \\ \mathcal{E}_2 & \mathcal{V} \end{bmatrix}; \quad \mathcal{Z} = \begin{bmatrix} \mathcal{Z}_1 & 0 \\ \mathcal{Z}_2 & 0 \end{bmatrix};$$

$$\mathcal{T} = \begin{bmatrix} \mathcal{T}_1 & 0 \end{bmatrix} \qquad (28)$$

where $E = \begin{pmatrix} I \\ 0 \end{pmatrix}.$

Moreover, the Theorem 4.3 is to solve the FFMR problem in the FF index (15).

Theorem 4.3 *Error model (10) is stable, FF index (15) is fulfilled, on condition that there are* $\mathcal{B} = \begin{bmatrix} \mathcal{B}_1 & \mathcal{B}_2 \\ * & \mathcal{B}_3 \end{bmatrix}$, $C = \begin{bmatrix} C_1 & C_2 \\ * & C_3 \end{bmatrix} > 0$, $\mathcal{X} = \begin{bmatrix} \mathcal{X}_1 & \mathcal{X}_2 \\ * & \mathcal{X}_3 \end{bmatrix} > 0$, $\check{M}_x$, $\check{N}_x$, $\check{J}_x$, $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{Z}_1$, $\mathcal{Z}_2$, $\mathcal{T}_1$, $\mathcal{V}$, *satisfying, with* $x, t \in \{1, 2, ..., m\}$

$$\begin{bmatrix} \tilde{\triangle}_{11} & \tilde{\triangle}_{12} & \tilde{\triangle}_{13} & -\mathcal{Z}_2^T + E\check{\mathcal{A}}_x \\ * & \tilde{\triangle}_{22} & \tilde{\triangle}_{23} & \check{\mathcal{A}}_x \\ * & * & \tilde{\triangle}_{33} & -\mathcal{X}_2 + A_t^T \mathcal{Z}_2^T \\ * & * & * & -\mathcal{X}_3 \end{bmatrix} < 0 \tag{29}$$

$$\begin{bmatrix} \tilde{\daleth}_{11} & \tilde{\daleth}_{12} & \tilde{\daleth}_{13} & \tilde{\daleth}_{14} & \tilde{\daleth}_{15} & 0 \\ * & \tilde{\daleth}_{22} & \tilde{\daleth}_{23} & \tilde{\daleth}_{24} & \tilde{\daleth}_{25} & 0 \\ * & * & \tilde{\daleth}_{33} & \tilde{\daleth}_{34} & \tilde{\daleth}_{35} & J_t^T \\ * & * & * & \tilde{\daleth}_{44} & \tilde{\daleth}_{45} & -\check{J}_x^T \\ * & * & * & * & \tilde{\daleth}_{55} & 0 \\ * & * & * & * & * & -I \end{bmatrix} < 0 \tag{30}$$

with

$$\begin{aligned}
\tilde{\triangle}_{11} &= \mathcal{B}_1 - \mathcal{E}_1 - \mathcal{E}_1^T; \ \tilde{\triangle}_{12} = \mathcal{X}_2 - E\mathcal{V} - \mathcal{E}_2^T; \\
\tilde{\triangle}_{13} &= \mathcal{E}_1 M_t - \mathcal{Z}_1^T; \\
\tilde{\triangle}_{22} &= \mathcal{X}_3 - \mathcal{V} - \mathcal{V}^T; \\
\tilde{\triangle}_{23} &= \mathcal{E}_2 M_t; \ \tilde{\triangle}_{33} = -\mathcal{X}_1 + \mathcal{Z}_1 M_t + M_t^T \mathcal{Z}_1^T; \\
\tilde{\daleth}_{11} &= \mathcal{B}_1 - \mathcal{E}_1 - \mathcal{E}_1^T; \ \tilde{\daleth}_{12} = \mathcal{B}_2 - E\mathcal{V} - \mathcal{E}_2^T; \\
\tilde{\daleth}_{13} &= e^{j\tilde{\varrho}_3} C_1 - \mathcal{Z}_1^T + \mathcal{E}_1 M_t; \ \tilde{\daleth}_{22} = \mathcal{B}_3 - \mathcal{V} - \mathcal{V}^T; \\
\tilde{\daleth}_{14} &= e^{j\tilde{\varrho}_3} C_2 - \mathcal{Z}_2^T + E\check{M}_x; \\
\tilde{\daleth}_{24} &= e^{j\tilde{\varrho}_3} C_3 + \check{M}_x; \\
\tilde{\daleth}_{15} &= \mathcal{E}_1 N_t + E\check{N}_x - \mathcal{T}_1^T; \ \tilde{\daleth}_{25} = \mathcal{Z}_2 N_t + \check{M}_x; \\
\tilde{\daleth}_{33} &= -\mathcal{B}_1 - 2cos(\tilde{\varrho}_4)C_1 + \mathcal{Z}_1 M_t + M_t^T \mathcal{Z}_1^T; \\
\tilde{\daleth}_{34} &= -\mathcal{B}_2 - 2cos(\tilde{\varrho}_4)C_2 + M_t^T \mathcal{Z}_2^T; \\
\tilde{\daleth}_{35} &= \mathcal{Z}_1 N_t + M_t^T \mathcal{T}_1^T; \ \tilde{\daleth}_{23} = e^{-j\tilde{\varrho}_3} C_2 + \mathcal{E}_2 M_t; \\
\tilde{\daleth}_{44} &= -\mathcal{B}_3 - 2cos(\tilde{\varrho}_4)C_3; \ \tilde{\daleth}_{45} = \mathcal{Z}_2 N_t; \\
\tilde{\daleth}_{55} &= -\beta^2 I + sym[\mathcal{T}_1 N_t].
\end{aligned}$$

The following reduced-order parameters as :

$$\begin{aligned}
\hat{M}_t &= \mathcal{V}^{-1}\check{M}_x \\
\hat{N}_t &= \mathcal{V}^{-1}\check{N}_x; \\
\hat{J}_t &= \check{J}_x
\end{aligned} \tag{31}$$

Proof 4.4 *We propose the following equations :*

$$\sum_{x=1}^{q} \sum_{t=1}^{m} \hat{l}_x l_t \tilde{\triangle}_{xt}$$

$$\sum_{x=1}^{q} \sum_{t=1}^{m} \hat{l}_x l_t \tilde{\daleth}_{xt}; \tag{32}$$

Therefore, us find Theorem 4.1.

Remark 4.5 *If us pick* $C = 0$, *we mastery employ theorem 4.3 to settle the* H_∞ *MR nonlinear systems over EF range.*

5 Numerical Example

Consider a fuzzy system in discrete time, represents in [12] :

$$\begin{aligned}
\mathbf{J}_1 &= \begin{bmatrix} 1.4419 & 0.6720 & 0.1387 & -0.8595 \end{bmatrix} \\
\mathbf{J}_2 &= \begin{bmatrix} 1.3329 & 0.6720 & 0.1387 & -0.8478 \end{bmatrix}; \\
\mathbf{M}_1 &= \begin{bmatrix} 0.1612 & 0.0574 & -0.0144 & 0.1846 \\ 0.0434 & -0.3638 & 0.5258 & -0.0357 \\ -0.0747 & -0.3146 & -0.0487 & -0.1043 \\ -0.1664 & 0.4031 & 0.0347 & 0.2864 \end{bmatrix};
\end{aligned}$$

$$\begin{aligned}
\mathbf{M}_2 &= \begin{bmatrix} 0.1312 & 0.0474 & -0.0044 & 0.1546 \\ 0.0234 & -0.3018 & 0.4258 & -0.0357 \\ -0.0554 & -0.2421 & -0.0367 & -0.0843 \\ -0.1551 & 0.3031 & 0.0247 & 0.1864 \end{bmatrix}; \\
\mathbf{N}_1 &= \begin{bmatrix} 0.2023 \\ -0.2313 \\ -0.1137 \\ 0.1279 \end{bmatrix}; \ \mathbf{N}_2 = \begin{bmatrix} 0.0123 \\ -0.1313 \\ -0.1138 \\ 0.1179 \end{bmatrix}; \\
E &= \begin{bmatrix} I & 0 \end{bmatrix}.
\end{aligned} \tag{33}$$

The normalized membership function :

$$\begin{aligned}
\mathbf{T}_1(\mathbf{z}_1(\mu)) &= \frac{1 - \sin(\mathbf{z}_1(\mu))}{2} \\
\mathbf{T}_2(\mathbf{z}_1(\mu)) &= 1 - \mathbf{T}_1(\mathbf{z}_1(\mu))
\end{aligned} \tag{34}$$

Via using Theorem 4, the obtained matrix parameters of FF H_∞ reduced order systems are the following :

- EF area : $(-\pi \leq \tilde{\varrho} \leq \pi)$

$$\left[\begin{array}{c|c} \hat{\mathbf{M}}_1 & \hat{\mathbf{N}}_1 \\ \hline \hat{\mathbf{J}}_1 & - \end{array} \right] = \left[\begin{array}{cc|c} -0.7415 & -0.7001 & -0.1925 \\ 0.7145 & 0.1925 & 0.4512 \\ \hline -0.8874 & -1.2145 & - \end{array} \right];$$

$$\left[\begin{array}{c|c} \hat{\mathbf{M}}_2 & \hat{\mathbf{N}}_2 \\ \hline \hat{\mathbf{J}}_2 & - \end{array} \right] = \left[\begin{array}{cc|c} -0.8840 & 0.7415 & 0.1325 \\ -0.6245 & 0.2415 & -0.4125 \\ \hline -1.2021 & -1.1325 & - \end{array} \right]$$
$$\tag{35}$$

- FF area : $(\frac{\pi}{7} \leq \tilde{\varrho} \leq \frac{\pi}{4})$

$$\left[\begin{array}{c|c} \hat{\mathbf{M}}_1 & \hat{\mathbf{N}}_1 \\ \hline \hat{\mathbf{J}}_1 & - \end{array}\right] = \left[\begin{array}{cc|c} -0.4415 & 0.2145 & -0.1125 \\ -0.1745 & 0.1325 & 0.1954 \\ \hline -0.2314 & 0.7458 & - \end{array}\right];$$

$$\left[\begin{array}{c|c} \hat{\mathbf{M}}_2 & \hat{\mathbf{N}}_2 \\ \hline \hat{\mathbf{J}}_2 & - \end{array}\right] = \left[\begin{array}{cc|c} -0.6614 & 0.2845 & -0.1354 \\ -0.3325 & 0.1325 & 0.2157 \\ \hline -0.3458 & 0.5842 & - \end{array}\right]$$

$$(36)$$

The comparison result with the technique proposed in Theorem 4.3 illustrate in Table 1, that indicate the little conservation of the method suggest in the paper.

The initial cases are proposed null. Figure 1 illustrate the errors $e(k)$ from various approaches. It illustrate the FF method have the best performance in comparison with the EF approach.

Table 1: H_∞ performance apply from various approaches

Frequency	Methods	β_{min}	Max error
$-\pi \le \varrho \le \pi$	[12]	0.9715	0.0915
$-\pi \le \varrho \le \pi$	**Th 4.3** (Q=0)	0.5514	0.0423
$\frac{\pi}{7} \le \varrho \le \frac{\pi}{3}$	**Th 4.3**	**0.1305**	**0.0088**
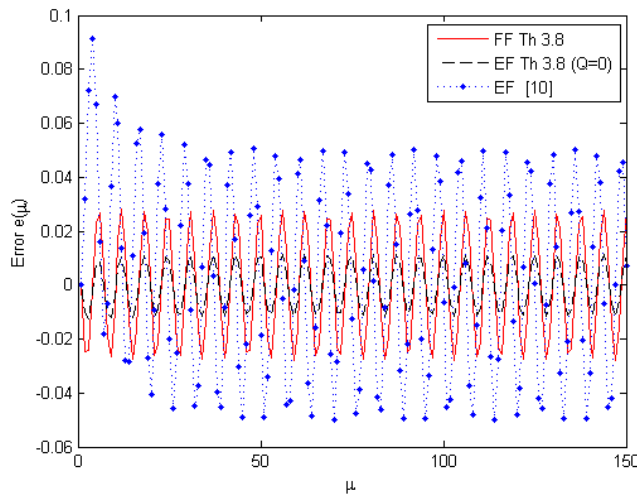
Figure 1: Error response of $\mathbf{g}_\mu$.

The ratio $\mathbf{W}(\mu)$ is presented as :

$$\mathbf{W}(\mu) = \sqrt{\sum_{\mu=0}^{\infty} \mathbf{g}^T(\mu)\mathbf{g}(\mu) / \sum_{\mu=0}^{\infty} \mathbf{d}^T(\mu)\mathbf{d}(\mu)} \qquad (37)$$

Figures 2 indicate the the ratio in (37), we could notice whether the error model is stable, knowing that the initial Condition are null, that indicate the little conservation of the method suggest in the paper.
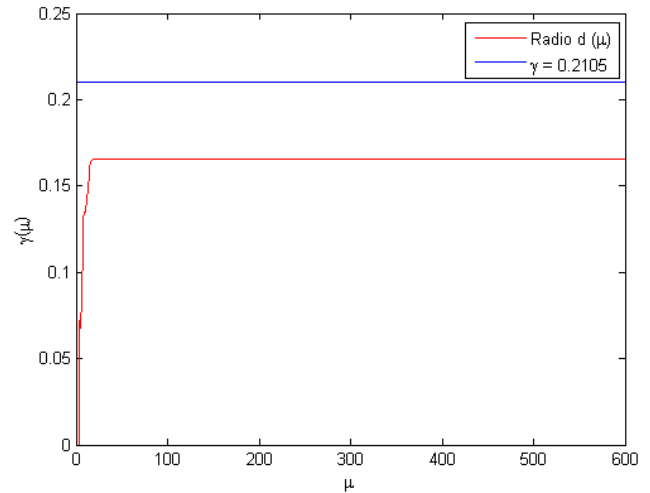
Figure 2: Value of $\mathbf{W}(\mu)$.

6 Conclusion

This work, we dealt with problematic for the FF model reduction design of nonlinear systems over FF ranges. We have suggested a model reduction process in order to minimize the conservatism design using the frequency information of the disturbances and we have assumed that the disturbances are known in a recognized FF domain. Also, systematic techniques have been suggested for the generation of a model reduction which ensures asymptotic stability and FF H_∞ index, at the basis of a more general linearization procedure.

References

[1] T. Takagi, M. Sugeno, " Fuzzy identification of systems and its applications to modeling and control," IEEE transactions on systems, man, and cybernetics, SMC-**15**(1), 116-132, 1985, doi: 10.1109/TSMC.1985.6313399

[2] H. Zhang, Y. Shi, A. S. Mehr, "On H_∞ filtering for discrete time Takagi Sugeno fuzzy Systems," IEEE Transactions on Fuzzy Systems, **20**(2), 396-401, 2011, doi: 10.1109/TFUZZ.2011.2175933

[3] A. Benzaouia, A. Hmamed, A. EL Hajjaji, " Stabilization of controlled positive discrete-time T-S fuzzy systems by state feedback control," International Journal of Adaptive Control and Signal Processing, **24**(12), 1091-1106, 2010, doi: 10.1002/acs.1185

[4] H. Gao, Y. Zhao, J. Lam, K. Chen, "H_∞ fuzzy filtering of nonlinear systems with intermittent measurements," IEEE Transactions on Fuzzy Systems, **17**(2), 289-299, 2009.

[5] L. Zhang, H. Yang, Q. Mao, "T S fuzzy system identification based on bayesian inference," ICIC Express Letter, **13**(1), 51-58, 2019.

[6] M. Nentwig, P. Mercorelli. "hrottle valve control using an inverse local linear model tree based on a fuzzy neural network," 7th International Conference on Cybernetic Intelligent Systems, IEEE, 1-6, 2008, doi: 10.1109/UKRI-CIS.2008.4798943

[7] M. Schimmack, F. Pöschke, H., Schulte, P. Mercorelli, "Observer Strategies for Virtual Sensing of Embroidered Metal-Polymer Heater Structure," 23rd International Conference on Methods & Models in Automation & Robotics (MMAR) IEEE, 887-892, 2018, doi: 10.1109/MMAR.2018.8486033

[8] A. C. Antoulas, "An overview of approximation methods for large-scale dynamical systems," Annual reviews in Control, **29**(2), 181-190, 2005, doi: 10.1016/j.arcontrol.2005.08.002

[9] K. Glover, "All optimal Hankel-norm approximations of linear multivariable systems and their L,∞-error bounds". International journal of control," **39**(6), 1115-1193, 1984, doi: 10.1080/00207178408933239

[10] G. Lastman, "Reduced-order aggregated models for bilinear time-invariant dynamical systems," IEEE transactions on automatic control, **29**(4), 359-361, 1984, doi: 10.1109/TAC.1984.1103534

[11] B. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," IEEE transactions on automatic control, **26**(1), 17-32, 1981, doi: 10.1109/TAC.1981.1102568

[12] L. Wu, X. Su, J. Qiu, "Model approximation for discrete-time state-delay systems in the T–S fuzzy framework," IEEE Transactions on Fuzzy Systems, **19**(2), 366-378, 2011, doi: 10.1109/TFUZZ.2011.2104363

[13] D. Wang, A. Zilouchian. "Model reduction of discrete linear systems via frequency-domain balanced structure," IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, **47**(6), 830-837, 2000, doi: 10.1109/81.852936

[14] X. Du, G. H. Yang, "H_∞ model reduction of linear continuous-time sys-tems over finite-frequency interval," IET control theory & applications, **4**(3), 499-508, 2020, doi : 10.1049/iet-cta.2008.0537

[15] X. Li, S. Yin, H. Gao, "Passivity-preserving model reduction with finite frequency H_∞ approximation performance," Automatica, **50**(9), 2294-2303, 2014, doi: 10.1016/j.automatica.2014.07.001

[16] D. W. Ding, X. J. Li, X. Du, X. Xie," Finite-frequency model reduction of Takagi-Sugeno fuzzy systems," IEEE Transactions on Fuzzy Systems, **24**(6), 1464-1474, 2016, doi: 10.1109/TFUZZ.2016.2540060

[17] A. El-Amrani, B. Boukili, A. El-Hajjaji, A. Hmamed, "H_∞ model reduc-tion for T-S fuzzy systems over finite frequency ranges," optimal control applications and methods, **39**(4), 1479-1496, 2018, doi : 10.1002/oca.2422

[18] D. W. Ding, X. Xie, X. Du, X. J. Li, "Finite-frequency model reduction of discrete-time T–S fuzzy state-delay systems," Neurocomputing, **203**, 121-128, 2016, doi: https://doi.org/10.1016/j.neucom.2016.03.053

[19] A. El-Amrani, A. El Hajjaji, I. Boumhidi, A. Hmamed, A. Aitouche, "Finite Frequency Approach for H_∞ model reduction of 2D continuous systems," 27th Mediterranean Conference on Control and Automation (MED) IEEE, 177-182, 2019, doi: 10.1109/MED.2019.8798569

[20] X. Su, L. Wu, P. Shi, Y. D. Song, "H_∞ model reduction of Takagi-Sugeno fuzzy stochastic systems", IEEE transactions on systems, man, and cybernet-ics. Part B, Cybernetics: a publication of the IEEE Systems, Man, and Cyber-netics Society, **42**(6), 1574-1585, 2012, doi: 10.1109/tsmcb.2012.2195723

[21] A. V. Oppenheim, A. S. Willsky, S. H. Nawab. "Signals and Systems", 2nd Edition,. Prentice Hall, Boston University, 1-987, 1996.

[22] M. J., Lacerda, R. C. F., Oliveira, P. L. D., Peres, "Robust H_2 and H_∞ filter design for uncertain linear systems via LMIs and polynomial matrices," Signal Processing, **91**(5), 1115-1122, 2011, doi: 10.1016/j.sigpro.2010.10.013

7 Appendix

Multiply equation (16) of $\begin{bmatrix} \nabla(\mu) \\ \mathbf{d}(\mu) \end{bmatrix}$ for right and by its transposition to the left, we get

$$\beta^2 \mathbf{d}(\mu)^T \mathbf{d}(\mu) + tr(C(\mathbf{g}^{j\tilde{\varrho}_3}\nabla(\mu)\nabla(\mu+1)^T$$
$$- e^{-j\tilde{\varrho}_3}\nabla(\mu+1)\nabla(\mu)^T - 2cos(\tilde{\varrho}_4)\nabla(\mu)\nabla(\mu)^T))$$
$$- \nabla^T(\mu+1)\mathcal{B}\nabla(\mu+1) - \nabla^T(\mu)\mathcal{B}\nabla(\mu) + \mathbf{g}^T(\mu)\mathbf{g}(\mu) \le 0 \tag{38}$$

By taking the sum of $\mu = 0$ in π,seeing that $\nabla(0) = 0$ and $\lim_{\mu\to pi}\nabla(\mu) = 0$ we get

$$\sum_{\mu=0}^{\pi}\mathbf{g}^T(\mu)\mathbf{g}(\mu) - \beta^2\sum_{\mu=0}^{\pi}\mathbf{d}(\mu)^T\mathbf{d}(\mu)$$
$$+ Tr[C\sum_{\mu=0}^{\pi}(e^{j\tilde{\varrho}_3}\nabla(\mu)\nabla(\mu+1)^T + e^{-j\tilde{\varrho}_3}\nabla(\mu+1)\nabla(\mu)^T$$
$$- 2\cos(\tilde{\varrho}_4)\nabla(\mu)\nabla(\mu)^T)] \le 0 \tag{39}$$

From Parseval function [21], we get :

$$\beta^2\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathbf{d}^T(\tilde{\varrho})\mathbf{d}(\tilde{\varrho})d\tilde{\varrho}$$
$$- Tr[\frac{\cos(\tilde{\varrho}-\tilde{\varrho}_c)-\cos\tilde{\varrho}_w}{\pi}\int_{-\pi}^{\pi}\mathcal{M}^T(\tilde{\varrho})C\mathcal{M}(\tilde{\varrho})d\tilde{\varrho}]$$
$$- \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathbf{g}^T(\tilde{\varrho})\mathbf{g}(\tilde{\varrho})d\tilde{\varrho} \ge 0 \tag{40}$$

Then :

$$\cos(\tilde{\varrho}-\tilde{\varrho}_3)-\cos(\tilde{\varrho}_4) \ge 0 \tag{41}$$

have toward all gifted $\tilde{\varrho}_1 \le \tilde{\varrho} \le \tilde{\varrho}_2$. From $C > 0$ and (41), it attends that

$$Tr[\frac{\cos(\tilde{\varrho}-\tilde{\varrho}_c)-\cos(\tilde{\varrho}_4)}{\pi}\int_{-\pi}^{\pi}\mathcal{M}^T(\tilde{\varrho})C\mathcal{M}(\tilde{\varrho})d\tilde{\varrho}] \ge 0 \tag{42}$$

seeing that, (40) and the FF of signal input $\tilde{\varrho}_1 \le \tilde{\varrho} \le \tilde{\varrho}_2$, thus (15) is fulfilled.

The Internal Reliability of a Questionnaire on the Impact of Enterprise Resource Planning on the Performance of Moroccan Companies

Maroua Barha[1,*], Soumaia Hmimou[2], Mounir Ait Kerroum[3], Hamid Ait Lemqeddem[1]

[1]*Research Laboratory in Organizational Management Sciences, National School of Business and Management, Ibn Tofail University, Kenitra, 14000, Morocco*

[2]*Biology and Health Laboratory, Faculty of Science, Ibn Tofail University, Kenitra, 14000, Morocco*

[3]*Laboratory of Informatics Research, Faculty of Sciences, Ibn Tofail University, Kenitra, 14000, Morocco*

A R T I C L E   I N F O

A B S T R A C T

Background: Today, Enterprise Resource Planning (ERP) software is a major tool for strengthening competitiveness. They are an asset that is changing work practices through the rapid circulation of information, the coordination of action and the development of new ways of doing things, rapid access to a wide range of knowledge and the opening up of new services.
The purpose of our study is to validate a reliable and reasonable questionnaire to measure the impact of ERP on the Performance of Moroccan companies.
Material and methods: This is a cross-sectional study that took place between January and April 2020. The study was based on a questionnaire. The statistical validation of the questionnaire is based on the calculation of the Cronbach's Coefficient Index and a factor analysis.
Results: The study involved 41 Moroccan companies, 27 of which have an ERP. The Cronbach's alpha value of all the items obtained is 0.96.
Conclusion: Results showed that questionnaire is a well-structured, objective, valid and reliable in the Moroccan context, the questionnaire could be an instrument to evaluate the impact of ERP on the performance of Moroccan companies

1. Introduction

The company, as an open system, pursues a certain number of objectives that are in the interest of the stakeholders such as: management, shareholders, customers, suppliers, etc. Therefore, it is a place where convergent and contradictory interests are located. To do this, it is led to produce information that responds to the various concerns it raises, hence the emergence of Information Technology (IT).

At a time of globalization, privatization and increased information, the use of New Technologies of Information and Communication (NTIC) has been proven and recognized as an essential investment solution to ensure the evolution and development of companies [1].

From this point, ERP is presented like an IT application that is part of production management, because it contributes to the simplification of certain constraints through computerization of information systems facilitating the exchange and posting of information between employees and different departments of the company.

From this point, ERP is presented like an IT application that is part of production management, because it contributes to the simplification of certain constraints through computerization of information systems facilitating the exchange and posting of information between employees and different departments of the company.

ERP is an evolution of Manufacturing Resource Planning 2 (MRP2) which appeared in the early 1970s in the United States.

MRP2 was intended for production planning from industrial and commercial planning to production monitoring.

It is a configurable software built around a single database, capable of covering the majority of the information processing needs of the company's various functions thanks to a set of compatible modules.

If it is well mastered, an ERP can be a source of important competitive advantage since it will allow an optimal management of all the company's information and a cross-referencing of data from the different departments that make up the company. On this point, the implementation of ERP in companies develops the control, effectiveness and efficiency of operations and strategic decision-making [2].

In Morocco, the first major ERP projects were initiated as early as 1995. Large companies and public bodies have made use of this technology, such as Office Cherifien of Phosphates (OCP), National Office of Electricity (ONE) and National Office of Drinking Water (ONEP). In the last decade, ERP have shown a significant increase [3]. Companies of all sizes are making investments in ERP implementations in order to anticipate the positive impacts on their organization. Studying the impact of ERP on business performance is becoming an important criterion to justify the associated investments.

In this sense, our objective is to study the internal reliability of a questionnaire that measures the impact of ERP on the performance of Moroccan companies.

2. Materials and methods

This cross-sectional study took place between January and April 2020. The study was mainly based on a questionnaire on the impact of ERP on the performance of Moroccan companies with the aim of validating it. The population studied includes Moroccan companies from different sectors of activity.

The literature review concerning the impact of ERP on business performance us to use a scale of 15 items [4]-[6]. The latter were submitted to the professors of higher education for their scientific advice and to a linguist to judge the comprehensibility of the items because this scale will be intended for an Arab-French speaking population. They are spread over 3 dimensions:

- Economic performance : 7 items
- Organizational performance : 5 items
- Human Performance : 3 items

The items corresponding to each dimension were measured using Likert scales with five degrees:

- The first degree is : Very low
- The second degree is : Low

- The third degree is: Medium
- The fourth degree is: High
- The fifth degree is: Very high

Similarly, we conducted a questionnaire containing questions on statutes, motivations and opinions. Data confidentiality was entrusted to the respondents by the anonymity of the participant.

In our study, the internal reliability of a questionnaire was assessed by measuring the internal consistency of the different scales. Internal consistency, which reflects how items on the same scale are interrelated, was assessed using Cronbach's alpha coefficient. A Cronbach's alpha coefficient greater than 0.70 indicated good internal consistency of the scales [7]. Since the number of items constituting a dimension influences the value of Cronbach's alpha. The standardized Cronbach's alpha was presented for each scale.

Cronbach's alpha was also used in the item reduction step [7]. This step allows the questionnaire to be as short as possible while retaining the items that contribute the most to the scales. The Cronbach's alpha was generated for each scale by considering all but one item. This procedure was carried out for each scale by excluding each item in turn. When the value of Cronbach's alpha increases with the deletion of an item, this indicates that this item correlates poorly with the other items of its scale. Such an item is then excluded from the questionnaire. Next, exploratory factor analyzes were performed. Statistical analysis were done by the Statistical Package for the Social Sciences (SPSS, version 21) [8]. The data collected were therefore the subject of exploratory analyzes to determine the dimensionality of the scale. Principal Component Factor Analysis (PCF) is the most efficient method for synthesizing information and uncovering the underlying structure of a concept since it is a method for analyzing multi-variate data. Which makes it possible to simultaneously explore the relationships that exist between several variables studied [8].

In our study we used the Kayser Meyer Olkin (KMO) test. A KMO of less than 0.5 is unacceptable, 0.5 is low, more than 0.6 is acceptable, 0.7 is average, 0.8 is meritorious and 0.9 is excellent [9] and Bartlett's sphericity test to assess the potential effectiveness of the PCR studied. For a factor analysis to be feasible the Bartlett test must be significant ($p < 0.05$) [10].

3. Results

3.1. Description of the study population

The socio-economic and demographic characteristics of the 41 enterprises that participated in the study are presented in Table 1. Most of the companies that participated in this study are service companies. Only 27 companies surveyed have an ERP and two companies are planning to implement it. The latter can be justified by the high cost of the ERP. (Figure 1)

Table 1: Description of Study Population

Variables	Number	Percentage
Sector of activity		
Agro-alimentaire	7	17%
Bank/Insurance	1	2%
Commerce/Distribution	5	12%
Industry	9	22%
Machine and equipment / Automobile	4	10%
Service	15	37%
Have an ERP		
Yes	27	66%
No	12	29%
No, but we're planning to implement	2	5%
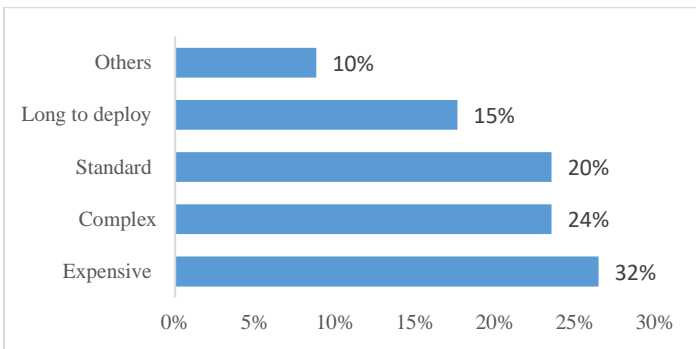
Figure 1: Critics of using ERP

3.2. Validation of the questionnaire

3.2.1. Internal reliability of items

The reliability of the measurement scale is checked by calculating the Cronbach's alpha for each dimension and for all items. (Table 2)

The reliability indices of the dimensions used in our study are well above 0.7. They are between 0.87 and 0.95. Cronbach's alpha is satisfactory, so there is homogeneity and correlation between the different dimensions of the questionnaire, they can be added together to form a scale score.

As for the total scale, the Cronbach's alpha value of all items was 0.96, which presents a good reliability.

As shows in Table 3, the results of the variables that best explain the impact of ERP on business performance.

After having completed the internal validity study in which we were able to exhaustively determine the variables most influencing the impact of ERP on business performance, we continue our study by processing the matrix of correlations between the 15 selected items.

Table 2: Cronbach alphas of items and dimensions based on the results of SPSS

	Cronbach's alpha	Items
Sets of items	0.96	15
Economic performance	0.95	7

Organizational performance	0.94	5
Human Performance	0.88	3

Table 3: Internal reliability based on the results of SPSS

	Scale average in case of deletion of an element	Scale variance when deleting an element	Cronbach's Alpha if the element is deleted
Time saving	44.47	207.263	.963
Saving Material Resources	44.95	207.164	.963
Human Resources Economics	44.95	199.608	.961
Improves customer satisfaction	44.84	203.585	.963
Improves the quality of services/products	44.63	202.579	.962
Increases Turnover	45.05	210.719	.965
Increases Market Share	45.11	211.655	.966
Facilitates access to information	44.47	208.374	.964
Improves the reliability of the information communicated	44.53	207.485	.963
Facilitates control tasks	44.47	207.708	.962
Improves coordination within the company	44.58	205.924	.962
Contributes to the decentralization of decisions	44.84	203.807	.963
Increases staff productivity	45.16	208.807	.963
Improves staff knowledge and skills	45.00	205.111	.963
Improves staff satisfaction	44.95	210.275	.964

3.2.2. Factor analysis

To make a factorial analysis of the questionnaire we took into consideration the value of the KMO index and the Barlett's test of sphericity.

For our study, the KMO index was 0.7 > 0.5 which shows an acceptable value to do the factor analysis, thus the Bartlett's sphericity test is highly significant. (Table 4)

To study the factorial structure of the collected data, we opted for Principal Component Analysis (PCA), which was carried out with varimax rotation on all the items of the questionnaire.

Table 4: Cronbach alphas of items and dimensions based on the results of SPSS

		Value
Indice KMO		**0.737**
Bartlett's sphericity test	Ci-square approx.	336.194
	Ddl	105
	Signification	**0.000**

3.2.3. Correlation matrix

We took into account 15 items used to assess the impact of ERP on business performance in order to design the correlation matrix as a quick way to estimate the internal consistency of the items. This will give 105 correlation coefficients.

The average correlation by averaging 105 coefficients is 0.655. More specifically, it turns out that the most important ration is 0.923 and the least important ratio is 0.318 (Table 5)

Table 5: Correlation Matrix based on the results of SPSS

	Time saving	Saving Material Resources	Human Resources Economics	Improves customer satisfaction	Improves the quality of services/products	Increases Turnover	Increases Market Share	Facilitates access to information	Improves the reliability of the information communicated	Facilitates control tasks	Improves coordination within the company	Contributes to the decentralization of decisions	Increases staff productivity	Improves staff knowledge and skills	Improves staff satisfaction
Time saving	1.000														
Saving Material Resources	.763	1.000													
Human Resources Economics	.801	.878	1.000												
Improves customer satisfaction	.804	.618	.773	1.000											
Improves the quality of services/products	.758	.630	.761	**.923**	1.000										
Increases Turnover	.504	.447	.588	.726	.757	1.000									
Increases Market Share	.626	.552	.613	.688	.658	.873	1.000								
Facilitates access to information	.738	.800	.735	.442	.554	**.318**	.369	1.000							
Improves the reliability of the information communicated	.683	.653	.704	.480	.635	.353	.328	.907	1,000						
Facilitates control tasks	.796	.739	.798	.688	.745	.488	.419	.838	,903	1.000					
Improves coordination within the company	.592	.768	.740	.552	.648	.426	.362	.779	,832	.885	1.000				
Contributes to the decentralization of decisions	.508	.618	.715	.595	.654	.530	.398	.606	,710	.801	.816	1.000			
Increases staff productivity	.635	.737	.754	.591	.633	.602	.667	.716	,669	.721	.746	.591	1.000		
Improves staff knowledge and skills	.530	.492	.679	.590	.673	.668	.485	.565	,706	.802	.778	.777	.731	1.000	
Improves staff satisfaction	.423	.480	.669	.598	.646	.686	.519	.504	,586	.664	.752	.740	.711	.877	1.000

3.3. Reasons for implementing an ERP system

Among the reasons for implementing an ERP, the reason to Harmonize their information system is in the lead with 32%, followed by the reason to Improve their performance and make their work easier with 28% (Figure 2).
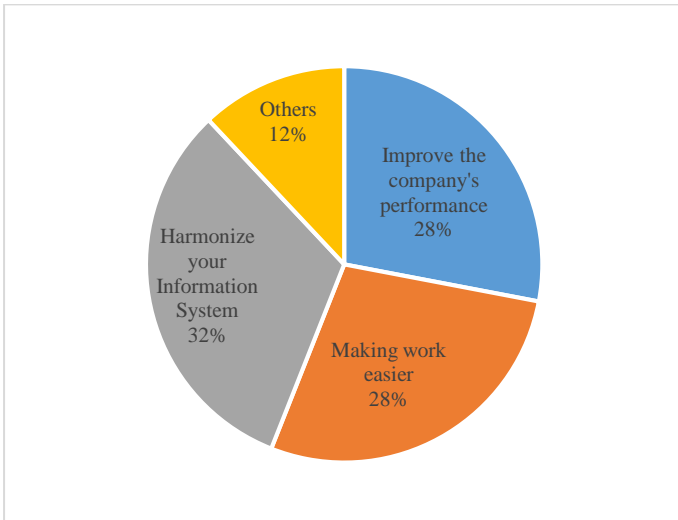
Figure 2: Main reasons for implementing an ERP system

4. Discussion

The objective of this study was to propose a scale that helps managers to measure the impact of ERP on the performance of Moroccan companies, which is reliable and has scientifically valid and satisfactory psychometric qualities, and which helps managers to measure the impact of ERP. Based on Churchill's paradigm [11]-[13].

The implementation of an ERP can generate benefits related to administrative productivity [14], such as: Optimization of the cost of ownership and elimination of reconciliation, correction and consolidation tasks [15]. For this reason, ERP can be a major tool to improve the overall functioning of the company and strengthen its performance [14]. According to our study 35% of the respondents used ERP to improve their performance.

The literature on performance measurement is broader than the literature on performance management [16]. The economic performance of an entity is measured on the basis of the change in business activity and the return on investments and sales. It can also come from an informed analysis of the income statement [17]. Organizational performance is the way in which the company is organized to achieve its objectives [18]. Human performance is analyzed through the results obtained by the employees, these results are expressed by the wealth created, the added value produced [19].

The implementation of ERP in Moroccan companies is recent, however, few studies that assess their impact on the performance of companies and this is what we notice in the literature and the

work already done. Therefore, our work can be considered as the first study carried out on the validation of the internal reliability of a questionnaire on the impact of ERP.

In order to test the internal reliability of a questionnaire, marketing researchers, especially those from other social sciences, are in the habit of comparing the value of the estimated Cronbach's alpha coefficient to a conventional threshold set at $\acute{\alpha} \geq 0.70$. Nevertheless, there are authors who have criticized this practice such as Peterson, 1994; Vos, Stem and Fotopoulos, 2000; Duhachek, Coughlan and Iacobucci, 2005 [20]. Indeed, the threshold $\acute{\alpha} \geq 0.70$ is only an empirical threshold because of the lack of statistical distributions to conclude whether Cronbach's alpha is acceptable or not [21].

However, thanks to the efforts of several psychometricians, the researchers were able to make statistical inferences about the distribution of Cronbach's alpha coefficient and its confidence interval [22]-[25].

Overall, the statistical analysis shows that the questionnaire on the impact of ERP on business performance meets several validity criteria. In our study, the internal reliability of the questionnaire was significant with a Cronbach's index of 0.969.

Although it was considered reasonable, we studied the internal validity in which we found that almost all the items retained their deletion leads to a decrease in Cronbach's alpha value, except for two items (Increases Sales and Increases Market Share) their deletion has no effect on Cronbach's alpha value [26], [27].

The correlation matrix shows that all the items are sufficiently correlated (r > 0.4) for us to consider that they measure the same construct, which is the impact of ERP on the company's performance.

5. Conclusions

In summary, we have demonstrated that the questionnaire on the impact of ERP on business performance has satisfactory reliability. Consistent with our expectation, all three dimensions show statistically significant correlations. Future validation studies will need to examine confidence interval and test-retest reliability.

Abbreviations

ERP: Enterprise Resource Planning;

IT: Information Technology;

KMO: Kayser Meyer Olkin;

MRP: Manufacturing Resource Planning;

NTIC: New Technologies of Information and Communication;

OCP: Office Cherifien of Phosphates;

ONE: National Office of Electricity;

ONEP: National Office of Drinking Water;

PCA: Principal Component Analysis;

PCF: Principal Component Factor

References

[1] M.E. Porter, V.E. Millar, "Strategic Management for Competitive Advantage How Information Gives You Competitive Advantage," 1985.

[2] A. Amzil, "Place Des Erp Dans Le Suivi De La Performance Des Filliales Par La Maison-Mere," 1–490, 2016.

[3] N. Belghazi, "La bataille des logiciels « fait tout »," L'Economiste, 1598, 2003.

[4] L. Gharsallah, Impact de l'ERP sur la performance : cas d'IGL, Université de Sfax, Tunisia, 2006.

[5] A. Chaabouni, "Implantation d'un ERP (Enterprise Resource Planning) : Antécédents et Conséquences," Conférence Internationale de Management Stratégique, Annecy / Genève 13-16, 1–26, 2006.

[6] H.R. Gopaul, A Systemic Analysis of Erp Implementation Challenges and Coping Mechanisms, Faculty of Commerce University of Cape Town, 2016.

[7] Peter M. Fayers and David Machin, Quality of Life - Assessment, Analysis & Interpretation, 2000, doi:10.1002/9780470024522.

[8] A. Field, Discovering Statistics Using Spss Andy Field.Pdf, Advances in Experimental Medicine and Biology, **622**, 153–167, 2008, doi:10.1177/014662168701100107.

[9] D.W. Stewar, "The Application and Misapplication of Factor Analysis in Marketing Research," Journal of Marketing Research, **18**(1), 51–62, 1981.

[10] V. Galtier, Proposition d'une échelle de mesure contextualisée de l'apprentissage d'équipe : Une analyse exploratoire, Université Paris IX Dauphine, 2003.

[11] G.A. Churchill, "A Paradigm for Developing Better Measures of Marketing Constructs," Journal of Marketing Research, **16**(1), 64, 1979, doi:10.2307/3150876.

[12] L.R. Fabrigar, R.C. MacCallum, D.T. Wegener, E.J. Strahan, "Evaluating the use of exploratory factor analysis in psychological research," Psychological Methods, **4**(3), 272–299, 1999, doi:10.1037/1082-989X.4.3.272.

[13] P. Roussel, F. Durrieu, E. Campoy, Méthodes d'équations structurelles : recherches et applications en gestion, Paris, 2002.

[14] J.-L. Deixonne, Piloter un projet ERP : transformer et dynamiser l'entreprise par un système d'information intégré et orienté métier, 2001.

[15] R. Marciniak, M. Carbonel, Management des projets informatiques : études de cas, Paris, 1996.

[16] B. Sogbossi Bocco, "Perception de la notion de performance par les dirigeants de petites entreprises en Afrique," La Revue Des Sciences de Gestion, **241**(1), 117, 2010, doi:10.3917/rsg.241.0117.

[17] C.Marmuse, Encyclopédie de la gestion, Edition de, Paris, 1996.

[18] C.Marmuse, Encyclopédie de la gestion, Edition de, Paris, 1997.

[19] D.C. Bernard Martory, Gestion des ressources humaines Pilotage social et performances, 5eme éditi, 2002.

[20] S. Moussa, "L ' alpha de Cronbach et l ' estimation de son intervalle de confiance : L ' étude de deux approches," in Entrepreneuriat et Entreprise dans la Nouvelle Economie : Contextes, Financement, Stratégie, Gouvernance et Performance, 11, 2008.

[21] B.P.Y. Evrard, Market : études et recherches en marketing : fondements, méthodes, 2ème, 1998.

[22] L.S. Feldt, D.J. Woodruff, F.A. Salih, "Statistical Inference for Coefficient Alpha," Applied Psychological Measurement, **11**(1), 93–103, 1987, doi:10.1177/014662168701100107.

[23] J.J. Barnette, "ScoreRel CI: An excel program for computing confidence intervals for commonly used score reliability coefficients," Educational and Psychological Measurement, **65**(6), 980–983, 2005, doi:10.1177/0013164405278577.

[24] J.M. Van Zyl, H. Neudecker, D.G. Nel, "On the distribution of the maximum likelihood estimator of Cronbach's alpha," Psychometrika, **65**(3), 271–280, 2000, doi:10.1007/BF02296146.

[25] D. Iacobucci, A. Duhachek, "Advancing Alpha: Measuring Reliability with Confidence," Journal of Consumer Psychology, **13**(4), 478–487, 2003, doi:10.1207/S15327663JCP1304_14.

[26] Jum C. Nunnally, Psychometric Theory, 2nd editio, New- York, 1978.

[27] N. Malhotra, Études marketing avec spss, 5e édition, Paris, 2007.

Application of Geographic Information Systems and Remote Sensing for Land Use/Cover Change Analysis in the Klip River Catchment, KwaZulu Natal, South Africa

Dunisani Thomas Chabalala[1,*], Julius Musyoka Ndambuki[1], Wanjala Ramadhan Salim[1], Sophia Sudi Rwanga[2]

[1]*Department of Civil Engineering, Tshwane University of Technology, Pretoria, 0001, South Africa*

[2]*Department of Civil Engineering, Vaal University of Technology, Vanderbijlpark, 1900, South Africa*

A B S T R A C T

Ladysmith is a major economic hub in the uThukela District Municipality. However, it has been experiencing floods almost every year which has resulted in the loss of lives and disruption of business activity within the Ladysmith Central Business District. The main objective of this study was to quantify the land use/cover changes before and after floods of 1994, 2006, and 2015 using Geographic Information System and remote sensing. Landsat images for the years 1990, 2000, 2010, and 2020 were used to prepare study area maps. The study revealed that the catchment has undergone drastic modifications in land use/cover in the past four decades. The results showed that agriculture, barren land, and built-up increased by 0.09 %, 63.95 %, and 34.19 %, while vegetation and water bodies drastically declined by 45.88 % and 60 % respectively. In conclusion, the Klip river catchment is at high risk of continuous flooding because of the rapid decrease in natural vegetation and water bodies. Therefore, the study recommends that government should give a greater focus on protecting, preserving, and regenerating natural vegetation as well as water bodies. This information will be useful to planners and policymakers in the planning and development of land use management strategies needed to reduce flooding in the study area.

1. Introduction

Identifying changes in land use/cover (LU/C) becomes an essential tool for policy formulation, planning as well as the development of targeted land use management strategy [1]. Land use/cover may be used interchangeably, land cover is defined as the physical characteristics found at the earth's surface [2], while land use describes the way in which human beings use the land. It occurs in two different forms which are conversion and modification. A change from one land-use class to another is called conversion, while a change in the physical characteristic of the land without a change in the overall classification is called modification [3]. In [4], the author indicated that changes in catchment characteristics may influence the frequency and magnitude of the flood regime, therefore, it is important to assess changes in land use/cover that have occurred each year [1]. There is evidence to prove the effectiveness of the Geographic Information System (GIS) in collecting, capturing, storing, retrieving, manipulating, analysing, and displaying digital data that is required for land use/cover change detection [5 - 7]. On the other hand, remote sensing has proved its potential for providing accurate and up-to-date information [8], whereas GIS is used to detect and monitor changes in land use/cover [9]. Integration of GIS and remote sensing has been proven to be the effective tools to be used to detect changes in land use/cover that have taken place in the catchment [10]. Several studies have previously focused on the changes in land use/cover using GIS and remote sensing. For example, in [11], the author applied GIS and remote sensing to investigate changes that have occurred in the Dedza district in Malawi. The study results showed that between 1991 and 2015, there was a significant decrease in agricultural land, water bodies, forest land, and wetlands while built-up areas and barren land significantly increased. The long-term annual rate of change in water bodies dropped from 5.54% ha^{-1} to 1.74% ha^{-1} during the study period. In contrast, the increase in the annual rate of change in forest land, agricultural land as well as built-up area was from 1.71% ha^{-1} to 1.94% ha^{-1}, 0.02% ha^{-1} to 0.11% ha^{-1}, and 7.22% ha^{-1} to 9.80% ha^{-1} respectively. In [12], the author used remote sensing and GIS to investigate change along the River Kaduna floodplain. Results from their study showed a substantial change in the land use/cover classes in the study area. The study

further revealed that there has been a progressive conversion of natural vegetation areas into farmlands and turning into built-up areas in the end. They concluded that the floodplain areas have undergone considerable changes because of human activities, and it has led to disastrous consequences. In [13], the author investigated the changes that happened before and after the floods in District Charsadda, Pakistan using GIS and remote sensing techniques. After assessing the percentage change of different classes, agricultural land, barren land, and water bodies was standing at 68.5%, 22.5 %, and 8.8 % before the flood, while, after the flood, water bodies, barren land, and agricultural land had grown up to 16.4 %, 26.30 %, and 57 %, respectively. The study also pointed out that the settlements along River Kabul were severely damaged by the flood. In [14], the author examined the hydrological effects of land use/cover change in the Harts catchment from 1990 to 2010 using GIS and remote sensing. From the study, it was found that the vegetation cover decreased from 758,345 ha in 1990 to 736,879.4 ha in 2008, while barren land increased by 97,651.49 ha from 226,670 ha to 324,321.49 ha in the same year. The study further pointed out that a decrease in natural vegetation cover has led to an increase in a runoff production and consequently changing flow regimes in the Harts catchment. In [15], the author quantified changes in land use/cover in the Western Cape Province, South Africa. The study used GIS and remote sensing to create land use/cover maps. The results indicated a significant change in the period from 1990 to 2014, with a decrease in grasslands, barren land, forest plantations as well as wetlands, and an increase in mines and quarries, water bodies, built-up areas, woodlands, thicket as well as shrubland classes. The study also found that mines and quarries had the highest increase of 52.82%, followed by plantation which decreased by 32.41%. The study cited the high demand of sand for construction purposes as the cause of the decline in mines and quarries and the government's exit policy as the cause of a decline in plantations. From the reviewed literature, it appears that land use/cover change analysis was conducted in different parts of the world, including South Africa. As far as the author's knowledge is concerned, there is still a lack of information on land use/cover change in the Klip river catchment, KwaZulu Natal. Therefore, this study aims to assess the land use/cover change using GIS and remote sensing. Findings from this study will be helpful to planners and policymakers in their planning development of land use management strategies aimed at reducing flooding in the study area.

2. Materials and Methods

2.1. Study area

The study site is located in Ladysmith which is a major town in Alfred Duma's local municipality (Figure 1). Ladysmith town lies between the latitudes 28°39′53″ and 28°39′53″ south and between longitudes 29°57′42″ and 29°57′42″ east, downstream of the Klip River catchment [16]. Klip River originates from the Drakensberg mountains approximately forty kilometers to the western side of Ladysmith, and its area covers about 1670 km² [16]. The study area receives rainfall of about 700 mm to 1000 mm/year, which is above the country's average annual rainfall of about 450 mm, with the highest amount of rainfall received in the months between October to March, while May to August received

low rainfall [16]. Ladysmith serves as a major commercial, financial and administrative area of both Alfred Duma municipality and uThukela District Municipality [17], however, it has always been subjected to flooding, and the situation has become more and more severe almost every year since 1884 [18]. In particular, the floods that occurred in 1994, 2011, 2012 and 2015, resulting in the loss of lives and disruption of business activity within the Ladysmith Central Business District [17 - 19]. In order to minimize floods occurrences, the local municipality has constructed two dams, constructed levees, installed stormwater drains but floods are continuing to cause problems within the town and its surrounding areas [19]. The study area's population continue to increase rapidly over the past few decades, it was 174 535 in 1996 and increased to 237437 in 2011, while the provincial population increased from 8.57 million in 1996 to 11.29 million in 2019 [20].
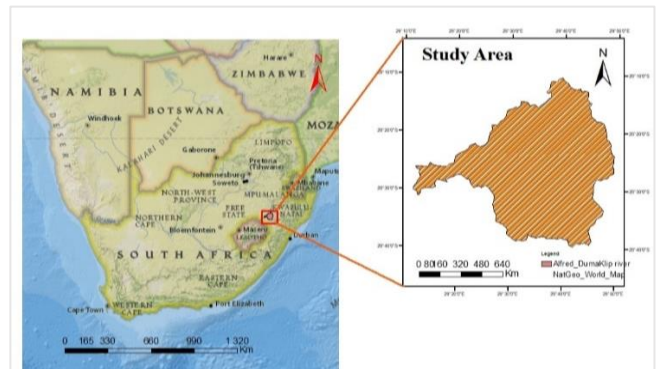
Figure 1: Location map of the study area

2.2. Data collection

In this paper, satellite imagery data has been used to investigate the historical changes in land use/cover that have taken place in the study area over 30 years starting from 1990 to 2020. Landsat images of four separate years (1990, 2000, 2010, and 2020) and digital elevation model (DEM) were freely downloaded from the USGS (United States Geological Survey) earth explorer website, https://earthexplorer.usgs.gov/. In order to avoid errors in the classification process, this study considered a cloud cover of less than ten (10) percent [21]. The physical characteristics of the Landsat images are shown in Table 1.

Table 1: Physical characteristics of the Landsat images used in this study

Satellite Data	Path/row	Date of acquisition	Grid cell (m)
Landsat 4-5 TM	169/80	22 October 1990	30
Landsat 7 ETM+	169/80	09 October 2000	30
Landsat 5 TM	169/80	06 May 2010	30
Landsat 8 OLI/TIS	169/80	25 September 2020	30

3. Methodology

3.1. Image pre-processing and classification

In this study, all images were pre-processed and classified using the ERDAS image software version 2015 and ArcGIS software version 10.7.1. Landsat images were georeferenced

(WGS_84, Zone 35 North Coordinate System – Universal Transverse Mercator), and adopted a supervised and maximum likelihood algorithm for LU/C classification [21]. The process began by creating training sites and a total of 375 signatures were created using the maximum likelihood tool, and for consistency, 65 signatures were created for each one of the five classes, namely, agriculture, barren land, built-up areas, vegetation, and water bodies [22]. After classification, LU/C maps were produced and used for further analysis in square kilometers [23]. The percentage area covered by each class was computed using Equation (1) [7; 24; 25]

$$P_A(\%) = \left(\frac{A_{LU/C}}{T_A}\right) * 100 \tag{1}$$

where P_A (%) represents the percentage area covered by LU/C, $A_{LU/C}$ stands for the area of land use/cover, and T_A is the total catchment area.

3.2. Land use/land cover change detection

LU/C change detection was carried out to indicate the forms of changes that have taken place between the two periods in the study area [7; 24; 26]. In this paper, the ten-year interval of LU/C maps of 1990, 2000, 2010 and 2020 was used to calculate the rate of change in between the two periods, and also the 30 years (net) changes in LU/C considering 1990 to 2020. Therefore, the rate of change between the two periods is calculated using Equation (2), [7; 23; 25; 27; 28].

$$RC\ (\%) = \left(\frac{B_2 - B_1}{B_1}\right) * 100 \tag{2}$$

where RC (%) represents the rate change in LU/C between two periods $B_1\ and\ B_2$, where B_1 denotes the area of LU/C at the initial year, and B_2 is the area of LU/C in the final year.

3.3. Accuracy assessment

In [29; 30], authors indicated that accuracy assessment is the final phase of classification and is used to measure the map accuracy to be able to use the information that is correct and accurate. This study adopted the Kappa coefficient statistical tool which was developed by Cohen in 1960 in order to measure the agreement between observed and chance agreement [31;32]. The classification error matrix information resulting from producer's accuracy, user's accuracy as well as overall accuracy was used as an input to Kappa coefficient, detailed information and concepts are reported in studies conducted by [32 - 36]. The advantage of using the Kappa coefficient is its capability to correct chance agreement between the predicted and observed values [37]. In order to achieve accurate results in image classification, a total of 50 random sample points were considered as recommended by [38], and for consistency, this study used the same number of points among the five considered categories, namely, agriculture, barren land, built-up areas, vegetation, and water bodies [39]. The sensitivity, specificity, error of commission and omission, as well as positivity and negativity predictive power were computed as follows: [21; 32; 34; 35].

$$Sensintivity = \frac{NA}{NA + NB} \tag{3}$$

$$Specificity = \frac{ND}{NB + ND} \tag{4}$$

$$Error\ of\ commision = 1 - \frac{ND}{NB + ND} \tag{5}$$

$$Error\ of\ ommision = 1 - \frac{NA}{NA + NB} \tag{6}$$

$$Positive\ Predictive\ Power = \frac{NA}{NA + NB} \tag{7}$$

$$Negative\ Predictive\ Power = \frac{ND}{NC + ND} \tag{8}$$

where:

NA stands for the number of times a classification agreed with the observed value.

NB represents the number of times a point was classified as X when it was observed to not be X.

NC is the number of times a point was not classified as X when it was observed to be X. ND is the number of times a point was not classified as X when it was not observed to be X.

$$Total\ points\ (TP) = (NA + NB + NC + ND) \tag{9}$$

Then, the overall accuracy was computed using Equation (10), while the Kappa coefficient was computed using Equation (11), [7; 21; 24; 26; 34; 38].

$$OA = \left(\frac{NCP}{TNRP}\right) * 100 \tag{10}$$

Where OA stands for Overall Assessment; NCP represents the number of correct points; $TNRP$ is the total number of reference points.

$$(KC) = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+} * x_{+i})} \tag{11}$$

where, KC (%) = Kappa coefficient, r = number of rows in the error matrix, x_{ii} = number of observations in rows i and column i, x_{i+} = marginal totals of rows i, x_{+i} = marginal totals of column i, N = total number of observations.

According to [12; 21; 24; 32; 40], the computed kappa coefficient results are guided by five rating scores. Table 2 presents the recommended Kappa coefficient rating score.

Table 2: Kappa statistics rating criteria

Score (%)	Descriptions
less than 0.2	Poor or very poor
0.2 to 0.4	Fair
0.4 to 0.6	Good
0.6 to 0.8	Very good
0.8 to 1.0	Excellent

4. Results and Discussion

4.1. Land use/land cover classification

In this study, the findings are presented in the form of maps, pie charts, and tables. These include LU/C classes, LU/C classification maps, and change detection over the period between 1990 and 2020. Four satellite images of 1990, 2000, 2010 and 2020 were used to create four different LU/C maps (Figure 2 a, b, c, and d). The LU/C were classified into five namely, agriculture, built-up areas, barren-lands, vegetation, and water bodies, and were described in Table 3, whereas the magnitude of changes in LU/C are shown in Figure 3 a, b, c, and d.
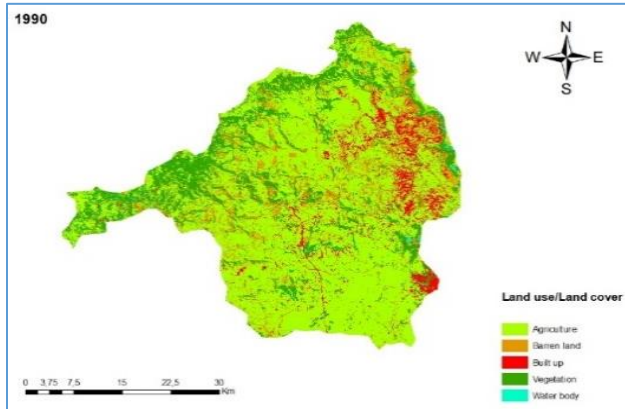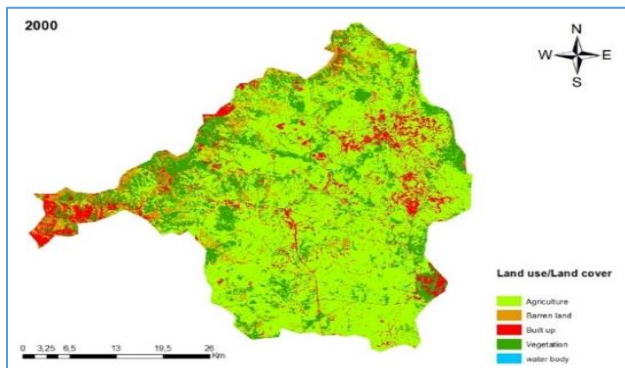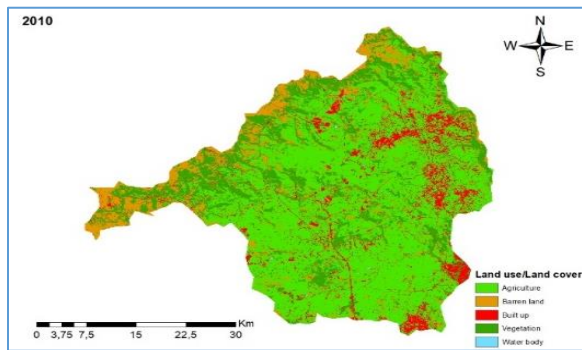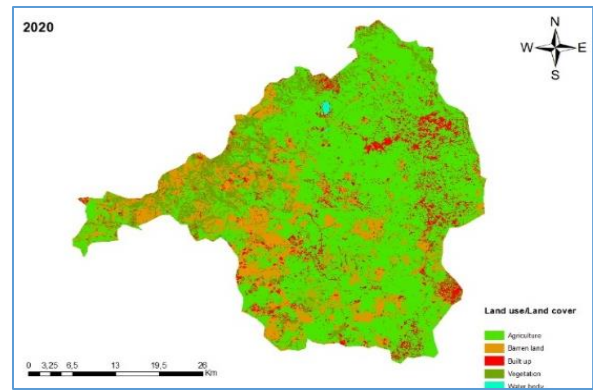
(d)

Figure 2 (a), (b), (c), and (d): Land use/cover maps of 1990, 2000, 2010 and 2020

(a)

(b)

(c)

(a)

(b)

68
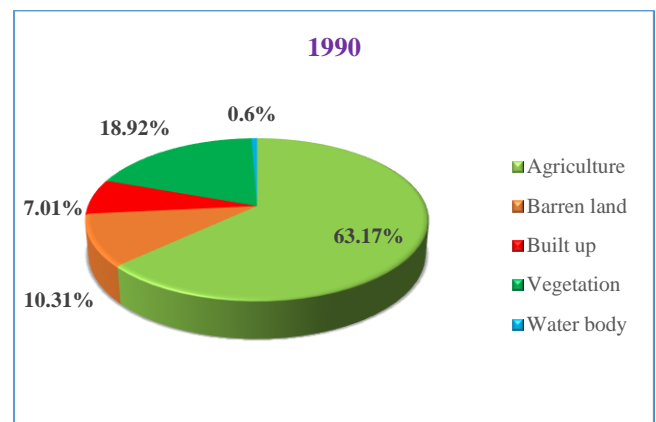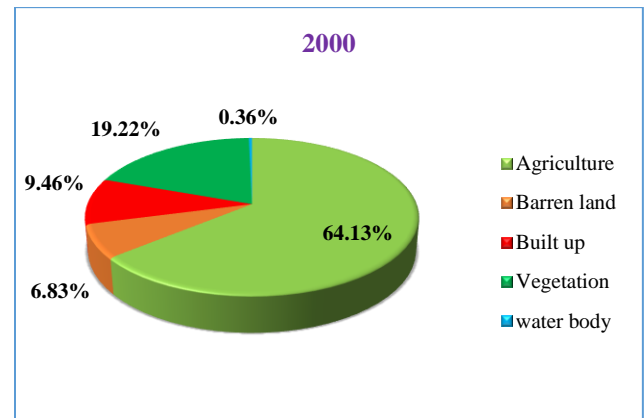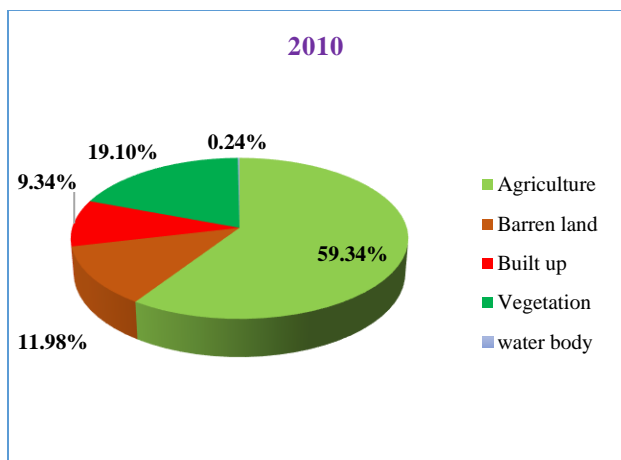
(c)
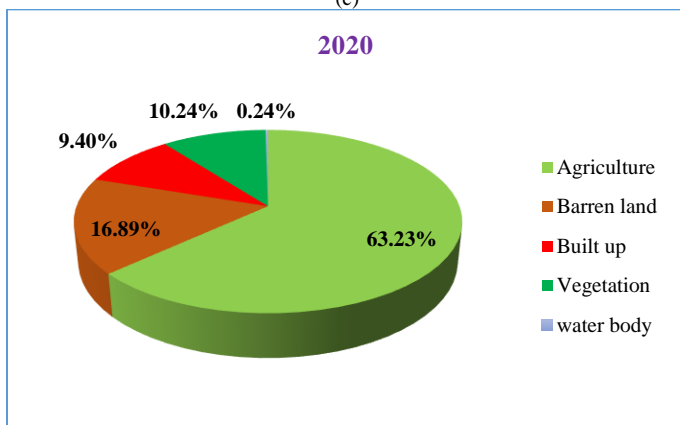
(d)

Figure 3(a), (b), (c), and (d): Pie charts showing land use/cover areas (1990, 2000, 2010 and 2020)

Table 3: Descriptions of land use/cover classes

Code	Classes	Descriptions
1	Agriculture	Areas used for agricultural activities, including crop fields and pastures.
2	Barren land	Areas with exposed soil including rocky or sandy areas, areas of severe erosion.
3	Built-up areas	Areas covered by buildings or structures including residential, industrial, and commercial services in both rural and urban.
4	Vegetation	Areas with perennial trees mostly grow up in the preserved land, mixed woodlands, and forest plantations near rivers and on top of the mountains.
5	Waterbody	Land covered by water, including dams and other bodies containing clear open water.

4.2. Land use/land cover change detection

This section describes the changes that have taken place before and after the floods of 1994, 2006, and 2011, and it has been done using LU/C maps of the Klip river catchment produced. The change statistics for the periods 1990-2000, 2000-2010, 2010-2020, and 1990-2020 produced from the comparison of classified maps (1990, 2000, 2010, and 2020) are shown in Tables 4, 5, 6, and 7. Table 4 displays the results of the LU/C change that occurred between the years 1990 and 2000. In 1990, agricultural land, barren land, built-up areas, vegetation and water bodies covers 63.17 % (1055 km^2), 10.30 % (172 km^2), 7.01 % (117 km^2), 18.92 % (316 km^2), and 0.6 % (10 km^2) of the total area, respectively. Then, barren land and water bodies significantly decreased by 33.72 % (58 km^2) and 40 % (4 km^2), from 172 km^2 and 10 km^2 in 1990 to 114 km^2 and 6 km^2 in 2000, respectively. This means that 58 km^2 of barren land and 4 km^2 of water bodies have been converted to agricultural land, built-up areas, and vegetation, which increased by 1.49 % (16 km^2), 35.04 % (41 km^2), and 1.58 % (5 km^2), respectively.

Table 4: Land use/cover change results (1990 – 2000)

Land use/Land cover type	1990 [Before Flood]		2000 [After Flood]		Rate of change km^2 (%)
	Area (km^2)	% of area	Area (km^2)	% of area	
Agriculture	1055	63,17	1071	64,13	16 (1,49)
Barren land	172	10,30	114	6,83	58 (-33,72)
Built up areas	117	7,01	158	9,46	41 (35,04)
Vegetation	316	18,92	321	19,22	5 (1,58)
Water body	10	0,60	6	0,36	4 (-40,00)
Total	1670	100	1670	100	

Table 5 presents the results of the LU/C change that has occurred within the period 2000 and 2010. In this period, barren land was found to be the only class that recorded an increase, it significantly increased by 75.44 % (86 km^2) from 114 km^2 in 2000 to 200 km^2 in 2010. Contrary to an increase, agricultural land, built-up areas, vegetation, and water bodies decreased from 63.14 % (1071 km^2), 9.46 % (158 km^2), 19.22 % (321 km^2), and 0.36 % (6 km^2) in 2000 to 59.34 % (991 km^2), 9.34 % (156 km^2), 19.10 % (319 km^2), and 0.24 % (4 km^2) in 2010, which represents decreasing annual rate 7.47 %, 1.28 %, 0.62 %, and 33.33 % respectively. The reason behind the decrease in built-up areas was political violence that had occurred in the province of KwaZulu Natal in the 1980s and 1990s, and continued for several years till 2011, with over 500 000 people being displaced from their homes and communities [41- 43].

Table 5: Land use/cover change results (2000 – 2010)

Land use/Land cover type	2000 [Before flood]		2010 [After flood]		Rate of change km^2 (%)
	Area (km^2)	% of area	Area (km^2)	% of area	
Agriculture	1071	64,13	991	59,34	80 (-7,47)
Barren land	114	6,83	200	11,98	86 (75,44)
Built up areas	158	9,46	156	9,34	2 (-1,28)
Vegetation	321	19,22	319	19,10	2 (-0,62)

Water body	6	0,36	4	0,24	2 (-33,33)
Total	1670	100	1670	100	

Table 6 shows the LU/C change results that occurred in the period between 2010 to 2020. In 2010, agriculture constituted the major part of the five classes, covering 59.34 % (991 km^2) of the total area, followed by vegetation 19.10 % (319 km^2), barren land 11.98 % (200 km^2), built-up areas 9.34 % (156 km^2), whereas water bodies occupied the smallest area 0.24 % (319 km^2). In this period, vegetation significantly decreased by 46.39 %, this indicates that 148 km^2 of natural vegetation has been converted to agricultural land, barren land, and built-up areas, which increased from 991 km^2, 200 km^2, and 156 km^2 in 2010 to 1056 km^2, 282 km^2, and 157 km^2 in 2020 with an annual increasing rate of 6.56%, 41%, and 0.637%, respectively. The results further showed that 4 km^2 of water bodies did not change in 2020.

Table 6: Land use/cover change results (2010 – 2020)

Land use/Land cover type	2010 [Before flood]		2020 [After flood]		Rate of change km^2 (%)
	Area (km^2)	% of area	Area (km^2)	% of area	
Agriculture	991	59,34	1056	63,23	65 (6,56)
Barren land	200	11,98	282	16,89	82 (41,00)
Built up areas	156	9,34	157	9,40	1 (0,637)
Vegetation	319	19,10	171	10,24	148 (-46,39)
Water body	4	0,24	4	0,24	0 (0,00)
Total	1670	100	1670	100	

Table 7 presents the 30-year changes (net) in LU/C for the period between 1990 and 2020. In this period, agriculture, barren land as well as built-up areas recorded an increase from 63.17 %, 10.30 %, and 7.01 % in 1990 to 63.23 %, 16.89 %, and 9.40 % in 2020, with an annual increasing rate of 0.09 %, 63.95 %, and 34.19 % respectively. On the other hand, vegetation and water bodies significantly decreased from 18.92 % and 0.6 % in 1990 to 10.24 % and 0.4 % in 2020, respectively, with an annual decreasing rate of 45.88 % and -60 %, respectively.

Table 7: Land use/cover change results (1990 – 2020)

Land use/Land cover type	1990 [Before flood]		2020 [After flood]		Rate of change km^2 (%)
	Area (km^2)	% of area	Area (km^2)	% of area	
Agriculture	1055	63,17	1056	63,23	1 (0,09)
Barren land	172	10,30	282	16,89	110 (63,95)
Built up areas	117	7,01	157	9,40	40 (34,19)
Vegetation	316	18,92	171	10,24	145 (-45,88)
Water body	10	0,6	4	0,24	6 (-60,00)
Total	1670	100	1670	100	

4.3. Accuracy assessment

The figures below show the accuracy assessment results for classified images for the years 1990, 2000, 2010 and 2020. The overall accuracy (OA) results were 81.0 %, 81.2 %, 84.0 %, and 88.0 %, while the calculated kappa coefficient (KC) results were 75.3 %, 76.0 %, 79.8 %, and 84.8 %, respectively (Figure 4). On the other hand, user's accuracy (UA) for all classes was above 70% except only vegetation which was 66.7% in 1990 and 62.1% in 2000. Similarly, producer's accuracy (PA) was also above 70% for all classes, except only water bodies which stood at 65.8% in 2020 (Figure 5). In general, the land use/cover maps accuracy achieved high accuracy results, and this means that both overall and kappa accuracy satisfy the minimum requirement as stipulated by [12; 21; 24; 32; 40].
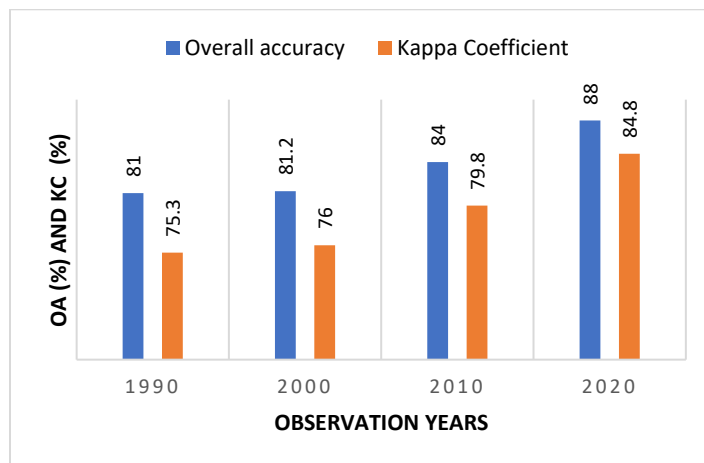
Figure 4: The overall and kappa coefficient accuracy assessment results (1990 – 2020)
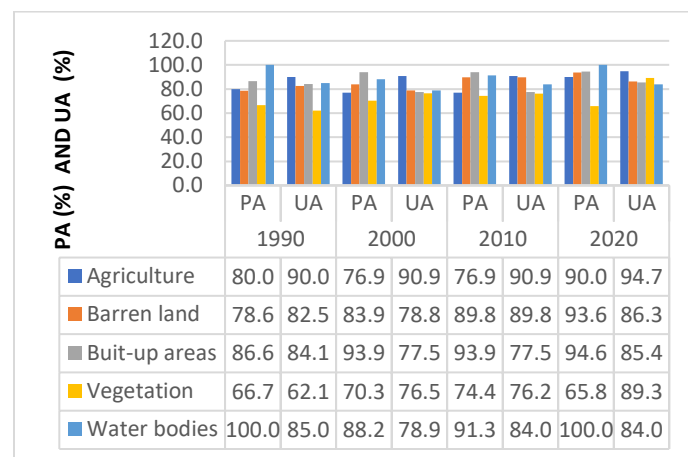
Figure 5: Producer and user's accuracy assessment results (1990 – 2020)

5. Conclusion

The objective of this study was to quantify the land use/cover changes before and after floods of 1994, 2006, and 2015 using Geographic Information System and remote sensing. The study results showed an increasing trend in agriculture, barren land, and built-up areas, while vegetation and water bodies showed a decreasing trend over the study period (1990 – 2020). In general, the study showed that the catchment has undergone drastic modifications over the past thirty years. Results further revealed that both vegetation and water bodies drastically decreased by 45.88 % and 60 %, from 18.92 % (316 km^2) and 0.6 % (10 km^2) in 1990 to 10.24 % (171 km^2) and 0.24 % (4 km^2) in 2020, respectively. The drastic decline has led to the expansion of agriculture, barren land as well as built-up areas, which has increased from 63.17 % (1055 km^2), 7.01 % (117 km^2), 10.30 %

(172 km^2) in 1990 to 63.23 % (1056 km^2), 9.40 % (157 km^2) and 16.89 % (282 km^2) in 2020, respectively. Therefore, it can be concluded that a drastic decline in natural vegetation and water bodies could be one of the contributing factors to the recurring floods in the Ladysmith Central Business District and its surrounding areas. This manuscript suggests that government should give greater focus on protecting, preserving and regenerating natural vegetation as well as water bodies. This information will be useful to planners and policymakers in the planning and development of land use management strategies needed to reduce flooding in the study area.

Acknowledgements

References

[1] M. H. Saputra, H. S. Lee, "Prediction of land use and land cover changes for north Sumatra, Indonesia, using an artificial-neural-network-based cellular automaton," Sustainability, **11**(11), 3024, 2019.

[2] C. Liping, S. Yujun, S, Saeed, "Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques: A case study of a hilly area," Jiangle, China, PloS one, **13**, (7), 0200493, 2018, doi.org/10.1371/journal.prone.0200493.

[3] R. Choudhary, D. Pathak, "Land use/land cover change detection through temporal imageries and its implications in water-induced disaster in Triyuga watershed, east Nepal," Journal of Nepal Geological Society, **51**, 49-54, 2016.

[4] M. P. A. Tali, T. A. Kanth, "Land use/land cover change and its impact on flood occurrence: a case study of upper Jhelum floodplain, M. Phil dissertation," University of Kashmir, 2011.

[5] E. Zadbagher, K. Becek, S. Berberoglu, "Modeling land use/land cover change using remote sensing and geographic information systems: a case study of the Seyhan Basin, Turkey," Environmental monitoring and assessment, **190**(8), 1-15, 2018, doi.org/10.1007/s10661-018-6877-y.

[6] H. Chen, L. Fleskens, J. Baartman, F. Wang, S. Moolenaar, C. Ritsema, "Impacts of land-use change and climatic effects on streamflow in the Chinese Loess Plateau: A meta-analysis," Science of the Total Environment, **703**, 134989, 2020, doi.org/10.1016/j.scitotenv.2019.134989.

[7] B. Bufebo, E. Elias, "Land Use/Land Cover Change and Its Driving Forces in Shenkolla Watershed, South Central Ethiopia," The Scientific World Journal, **2021**, 2021, doi.org/10.1155/2021/9470918.

[8] D. L. Peterson, S. L. Egbert, K. P. Price, E. A. Martinko, "Identifying historical and recent land-cover changes in Kansas using post-classification change detection techniques," Transactions of the Kansas Academy of Science, **107**(3), 105-118, 2004.

[9] J. Maina, S. Wandiga, B. Gyampoh, C. Kkg, "Assessment of Land Use and Land Cover Change Using GIS and Remote Sensing: A Case Study of Kieni, Central Kenya," Journal of Remote Sensing & GIS, **9**(1), 1-5, 2020, doi:10.35248/2469-4134.20.9.270.

[10] S. Poongothai, N. Sridhar, R. A. Shourie, "Change detection of land use/land cover of a watershed using remote sensing and GIS," International Journal of Engineering and Advanced Technology, **3**(6), 226-230, 2014.

[11] M. G. Munthali, J. O. Botai, N. Davis, A. M. Adeola, "Multi-temporal analysis of land use and land cover change detection for Dedza district of Malawi using geospatial techniques," International Journal of Applied Engineering Research, **14**(5), 2019.

[12] A. J. Abubakar, E. A. Olowolafe, Y. M. Obansa, "Land Use and Land Cover Change Analysis along River Kaduna Floodplain Using Geospatial Techniques," IIARD International Journal of Geography and Environmental Management, **1**(8), 2012.

[13] M. Fida, I. Hussain, W. Tao, A. Rashid, S. A. Ali Shah, "Land use and land cover change analysis of District Charsadda, Pakistan along Kabul River in 2010 flood: using an advance geographic information system and remote sensing techniques," Natural Hazards and Earth System Sciences Discussions, 1-16, 2020, doi:10.5194/nhess-2020-255.

[14] T. H. Kabanda, L. G. Palamuleni, "Land use/cover changes and vulnerability to flooding in the Harts catchment, South Africa," South African

[15] P. Tizora, A. Le Roux, G. G. Mans, A. K. Cooper, "Land use and land cover change in the Western Cape Province: quantification of changes & understanding of driving factors," 2016.

[16] D. T. Chabalala, J. M. Ndambuki, R. W. Salim, S. S. Rwanga, "Impact of climate change on the rainfall pattern of Klip River catchment in Ladysmith, KwaZulu Natal, South Africa," In IOP Conference Series: Materials Science and Engineering, IOP Publishing, **640** (1), 012088, November 2019, doi:10.1088/1757-899X/640/1/012088.

[17] N. Ngcobo, "Stakeholder's perceptions of using ecological infrastructure in addressing flood risks: the case study of the Emnambithi/Ladysmith Local Municipality," Master's Thesis, University of KwaZulu-Natal, 2016.

[18] F. G. Bell, T. R. Mason, "The problem of flooding in Ladysmith, Natal, South Africa," Geological Society, London, Engineering Geology Special Publications, **15**(1), 3-10, 1998.

[19] B. A. Schep, A. M. Roor, M. E. Rottink, P. Spielmann, A. E. Waqué, B. J. Van de Zande, "Project Ladysmith," 2016.

[20] South African Institute of Race Relation, "SA changing demographics-implications for policy," Institute of race relation", **2**(22), 1-10, 2019.

[21] S. S. Rwanga, J. M. Ndambuki, "Accuracy assessment of land use/land cover classification using remote sensing and GIS," International Journal of Geosciences, 8(4), 611-622, 2017, doi:10.4236/ijg.2017.84033.

[22] Y. Hailu, B. Tilahun, H. Kerebeh, T.Tafese, "Land use land cover change detection in Gibe Sheleko National Park, Southwestern Ethiopia," Agricultural and Resource Economics: International Scientific E-Journal, (**4**)4, 2018.

[23] J. S. Alawamy, S. K. Balasundram, C. T. Boon Sung, "Detecting and analyzing land use and land cover changes in the region of Al-Jabal Al-Akhdar, Libya using time-series Landsat data from 1985 to 2017," Sustainability, **12**(11), 4490, 2020, doi:10.3390/su12114490.

[24] T. Gashaw, T. Tulu, M. Argaw, A. W. Worqlul, "Evaluation and prediction of land use/land cover changes in the Andassa watershed, Blue Nile Basin, Ethiopia," Environmental Systems Research, **6**(1), 1-15, 2017, DOI10.1186/s40068-017-0094-5.

[25] K. Alkaradaghi, S. S. Ali, N. Al-Ansari, J. Laue, "Evaluation of land use & land cover change using multi-temporal Landsat imagery: a case study Sulaimaniyah Governorate, Iraq," Journal of Geographic Information System, **10**(6), 247-260, 2018, DOI:10.4236/ijis.2018.103013.

[26] M. A. Giraldo, L. S. Chaudhari, L. O. Schulz, "Land-use and land-cover assessment for the study of lifestyle changes in a rural Mexican community: The Maycoba Project," International Journal of health geographics, **11**(1), 1-9, 2012.

[27] D. C. Phan, T. H. Trung, T. Sasagawa, T. P. T. Vu, D. T. Bui, M. Hayashi, K. N. Nasahara, "First comprehensive quantification of annual land use/cover from 1990 to 2020 across mainland Vietnam," Scientific reports, **11**(1), 1-20, 2021, doi.org/10.1038/s41598-021-89034-5.

[28] A. S. Minale, "Retrospective analysis of land cover and use dynamics in Gilgel Abbay Watershed by using GIS and remote sensing techniques, Northwestern Ethiopia," International Journal of Geosciences, **4**(07), 1003, 2013, dx.doi.org/10.4236/ijg.2013.47093.

[29] R. Manandhar, I. O. Odeh, T. Ancev, "Improving the accuracy of land use and land cover classification of Landsat data using post-classification enhancement," Remote Sensing, **1**(3), 330-344, 2009, doi:103390/rs1030330.

[30] G. M. Foody, "Estimation of land coverage from a land cover classification derived from remotely sensed data," GeoJournal, **36**(4), 361-370, 1995.

[31] G. H. Rosenfield, K. Fitzpatrick-Lins, "A coefficient of agreement as a measure of thematic classification accuracy," Photogrammetric engineering and remote sensing, **52**(2), 223-227, 1986.

[32] S. Talukdar, P. Singha, S. Mahato, S. Pal, Y. A. Liou, A. Rahman, "Land-use land-cover classification by machine learning classifiers for satellite observations—A review," Remote Sensing, **12**(7), 1135, 2020, doi:10.3390/rs12071135.

[33] K. S. Cheng, J. Y. Ling, T. W. Lin, Y. T. Liu, Y. C., Shen, Y. Kono, "Quantifying Uncertainty in Land-Use/Land-Cover Classification Accuracy: A Stochastic Simulation Approach," Frontiers in Environmental Science, **9**, 46, 2021, doi:10.3389/fenvs.2021.628214.

[34] G. Banko, "A review of assessing the accuracy of classifications of remotely sensed data and methods including remote sensing data in forest inventory," 1998.

[35] Á. Barsi, Z. Kugler, I. László, G. Szabó, H. M. Abdulmutalib, "Accuracy Dimensions in Remote Sensing". International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, **42**(3). 2018, doi.org/10.5194/isprs-archives-XLII-3-61-2018.

[14] ... Geographical Journal, **95**(1), 105-116, 2013, doi: 10.1080/03736245.2013.806165.

[36] R. C. Weih Jr, D. White Jr, "Land-Use/Land-Cover Characterization Using an Object-Based Classifier for the Buffalo River Sub-Basin in North-Central Arkansas," Journal of the Arkansas Academy of Science, **62**(1), 125-132. 2008.

[37] J. Jenness, J. J. Wynne, "Cohen's Kappa and classification table metrics 2.0: An ArcView 3. x extension for accuracy assessment of spatially explicit models," Open-File Report OF   2005-1363. Flagstaff, AZ: US Geological Survey, Southwest Biological Science Center, 86p, 2005.

[38] R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," Remote sensing of environment, **37**(1), 35-46. 1991.

[39] K. Kitada, K. Fukuyama, "Land-use and land-cover mapping using a gradable classification method," Remote Sensing, **4**(6), 1544-1558, 2012, doi:10.3390/rs40615544.

[40] M. A. Ismail, A. N. M. Ludin, N. Hosni, "Comparative Assessment of the Unsupervised   Land Use Classification by Using Proprietary GIS and Open Source Software," In IOP   Conference Series: Earth and Environmental Science, IOP Publishing, 540 (1), 012020), July 2020, doi:10.1088/1755-1315/540/1/012050.

[41] M. Schuld, "Voting and violence in KwaZulu-Natal's no-go areas: Coercive mobilisation and territorial control in post-conflict elections," African Journal on Conflict Resolution, **13**(1), 2013.

[42] M. De Haas, "The killing fields of KZN." South African Crime Quarterly, **57**, 43-53, 2016, doi.org/10.17159/2413-3108/2016/v0n57a456.

[43] C. Higson-Smith, "Supporting communities affected by violence: A casebook from South Africa". Oxfam, 2002.

Survey on Novelty Detection using Machine Learning Techniques

Baida Ouafae[*], Louzar Oumaima, Ramdi Mariam, Lyhyaoui Abdelouahid

National School of Applied Sciences, Abdelmalek Essaadi University, Tangier, 90060, Morocco

A R T I C L E I N F O	A B S T R A C T
	Novelty detection affords to identify data patterns that stray strikingly from the normal behavior. it allows a good identification and classification of objects which were not known during the learning phase of the model. In this article, we will introduce an organized and comprehensive review of the study on novelty detection. We have grouped existing methods into three classes. Statistical Based techniques, Machine Learning Based techniques and Deep Learning Based techniques. In addition, we provide a discussion on application domains of novelty detection, and for each category, we have defined the novelty, cited the most used dataset, as well as a description and perspectives of the latest work carried out in this domain. Our article is developed with the aim of facilitating to researchers a better understanding of the interest of using novelty detection in the various fields mentioned in the article, as well as to clarify the different existing novelty detection methods.

1. Introduction

Novelty detection refers to the issue of deciding whether a data is in or outside the standard norm. This means identifying that the test data diverge in some way from the data was accessible over training.

Novelty detection methods covering a huge number of applications including fraud detection [1] , medical diagnosis [2], video surveillance [3], mobile robotics [4], social media [5] and among many others.

As a rule, novelty detection methods and techniques are applied when data is accessible from a particular pattern, which is typically called the "normal" class, as opposed to "abnormal" class which contains the new data.

In the literature, Novelty detection has many synonymous terms, such as anomaly, outlier or event detection [6]. In general, the expression event detection is generally utilized to indicate a diversity of detection types.

The problem of detecting abnormal forms in a data set which is not compatible with the data has drawn a lot of attention in a vast range of fields.

There are two types of classification, One-class classification and conventional multi-class classification, and the distinction between these two is that One-class classifiers during training, require nothing else than the target data, and for new instances,

perform a decision of target or unknown. In the case of absence of non-target data in the training dataset, it's favorable to use a multi-class classifier able to employ in its evaluation the negative data.

One class classification is closely related to rare event detection, outlier detection/removal, and anomaly detection. It has been applied to a huge range of learning problems like text classification [7], medical analysis [8], machine fault detection [9] etc. In addition, diverse business domains like financial credit scoring [10] and supplier selection [11] of which one class classification was applied.

The remainder of this paper is organized as follows. In Section 2, the related applications of novelty detection in the recent years. In Section 3, we present the novelty detection methods including the statistical based techniques, machine learning based techniques and deep learning based techniques. The novelty detection techniques measurements are presented in section 4. And the application domains for the novelty detection methods are discussed in Section 5. In Section 6, we provide an overall conclusion for this review and highlights the main directions of the future work.

2. Background

Distinct machine learning algorithms are used in several research works to perform novelty detection. This section mainly presents summaries of novelty detection research work carried out by a number of researchers and authors.

The authors in [6] introduced a totally new architecture which considers the transient progression of data streams in social media for detecting novelties. This novel architecture involves three new improvements. They proposed in the first a novel concept dependent on temporal windows for novelty definition. Second, to decide the quality of novelty, they proposed an expression. In addition, by using the COCO dataset and the MASK-RCNN convolutional neural network, they acquainted a novel methodology with the combination of heterogeneous data (image + text) which makes the data ready to be recognized by machine learning algorithms by transforming the data to a single data format. They used unsupervised algorithms like kNN, HBOS, FBagging, IForesting and Autoencoders since novelty detection is a task where tagged samples are scant or inexistent. The database created contains 27,494 tweets collected from Twitter, and by using this database, the results show that merging the data gives a higher classification than using only text or images as an input value.

The authors in [12] proposed a new methodology for novelty detection that utilizes kernel dependence techniques in order to make references based on the characterizations of statistical dependencies of random variables (RV). When the statistical dependencies of the RV are strong, their ignorance can lead to inaccurate inferences which are in the form of high false positive rates. In novelty detection, taking these dependencies in multivariate issues is an important challenge. The proposed method can handle problems with arbitrarily high dimensional data. For this it uses a kernel method to encode dependencies instead of using density estimation methods which are not robust with respect to dimensionality. This approach can be applied to general novelty detection problems since it does not require any prior information about the dependency structure of RVs.

In order to identify unusual patents that may give a new idea of potential opportunities, in [13], the authors developed a new methodology which allows to extract patents based on the novelty detection statistical strategy. In order to reduce the problem of vocabulary mismatch and alleviate the tedious efforts of expert keyword selection, this natural language processing technique is applied to extract hidden relations between words in patent documents. The proposed method can be applied in the telehealth industry and even can help telehealth companies to formulate their technological strategies for determining in a large data space the outliers patents which are distinct from most of the gathered patent documents collected. The authors in [14] demonstrated a methodology which includes both the classical classification of several scenarios known a priori and the innovative detection capacity of new operating modes not already accessible. It is a conditional technique applied to industrial machines and based on novelty detection. In order to optimize the performance of classification and novelty detection, a double feature reduction stage of characteristic reduction applied based on principal component analysis and linear discriminant analysis. This method is based on the temporal segmentation of the available physical quantities and on the estimation of a set of temporal statistical characteristics.

3. Novelty Detection Methods

One-class classification task is considered as a novelty detection problem that attempts to detect situations in which out of the ordinary events occur. Usually, in the database, the amount of normal data takes up most of the data set, knowing that the normal class is detected with precision and acoustic events are considered as new events located outside of the class [15]. Based on the latest research and articles on the subject of novelty detection, we can have grouped novelty detection approaches into three different categories: Statistical based techniques, Machine Learning based techniques and Deep Learning based techniques as shown in Figure 1.

3.1. Statistical Based Methods

Usually statistical approaches using statistical properties based on modelling data to appraise if a test sample comes from a similar distribution or not. To flag novelty, the probability estimate can be thresholder [16]. An additional model, can from a threshold that is based on the number of standard deviations and from the class mean, can easily find the sample distance [17], [18]. Statistical tests are used to differentiate all selections of deviations of the current behavior from the normal behavior.

In general, when designing a statistical novelty detector and to the assessment of the probability density function, there are two essential models to follow, parametric and non-parametric. The parametric approach guesses that the data comes from a group of known distributions which considering as normal distribution, the typical dissemination and certain parameters are determined to fit this distribution. Conversely in non-parametric methods, the general form of the density function is obtained from the data just as the parameters of the model, it makes no suppositions about a parametric distribution when displaying the data.

3.1.1. Parametric Approaches

In parametric technique, the ordinary information is thought to be made by the score for the anomaly of the data and parameters. Parametric approaches make a supposition that statistically as a function of the covariance and the data means, the distributions can be modeled. And they consider these distributions are Gaussian in nature.

Gaussian mixture models have been generally applied in speech and image classification, just as other comparable example acknowledgment issues [19]. GMM models general distributions assessing the density utilizing less kernels than the quantity of patterns in the training set. By maximizing the log likelihood of the training data with respect to the model, the parameters of the model are chosen. In the event that the dimensionality of the data is high, an extremely enormous number of samples are expected to train the model. In this sense, GMM experiences the scourge of dimensionality.

There are a variety of advanced parametric methods available, some of them include Gaussian Mixture Model (GMM) and Hidden Markov Models (HMM).

Hidden Markov models were proposed without precedent in the late 1960's mid 1970's [20]. What's more, are stochastic models for successive data [21]. A specific HMM regroups a finite number of unobservable (hidden) states and in each state, some state-dependent events can be observed by a probability distribution. To appraise the parameters of a HMM for displaying normal system behavior, arrangements of ordinary events
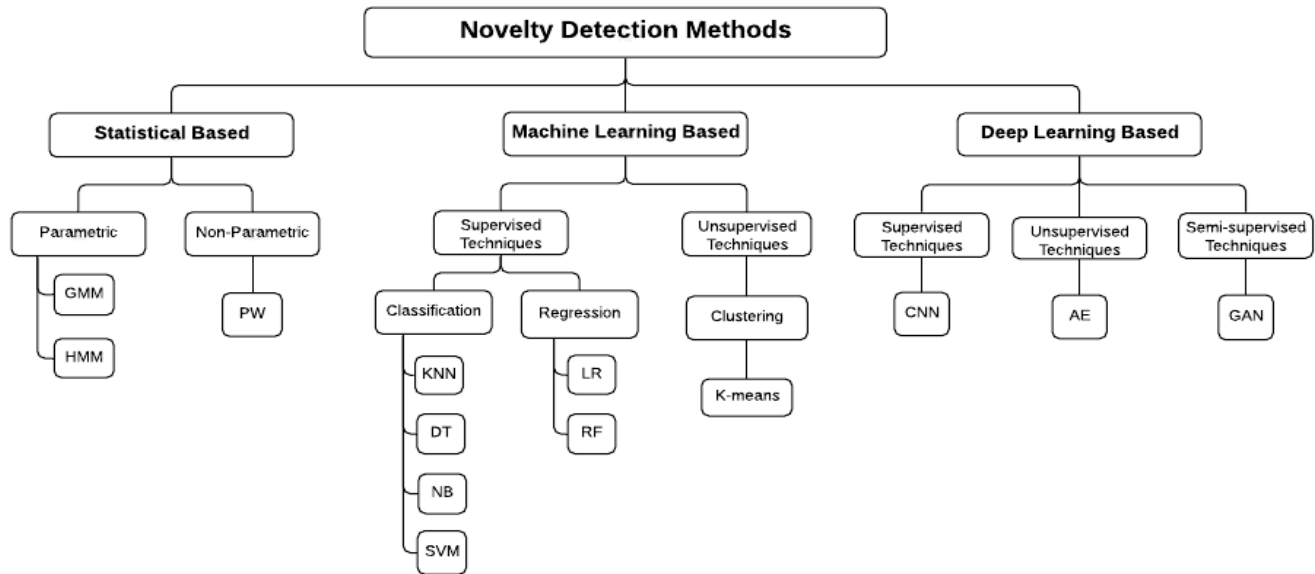
Figure 1: Classification of Novelty Detection Methods

gathered from normal system activity are utilized as training data. HMM is a specific probabilistic model broadly utilized in the field of acknowledgment particularly in speech, imagery, manually written writings and as of late in the analysis of biological sequences (DNA, proteins) [22].

3.1.2. Non parametric approaches

Non-parametric methods do not require making an assumption for the distribution of the data or the population being studied, but instead it is necessary to rely on the estimation of a distribution based on the data themselves. Frequently, Non-parametric methods solve from training data the problems that involve adaptability by modeling the system based on learned patterns [23]. The Parzen Windows approaches are among the statistical techniques used in a non-parametric model.

Parzen Window (PW) approach is generally used to estimate class-conditional probability densities for a given pattern. At the point when the size of the training dataset is during the several thousands and L is additionally enormous (a couple at least hundred), nonetheless, the PW plan can turn out to be computationally impractical. To improve the computational drawback of the PW approach, different schemes have been recommended in over years.

3.2. Machine Learning Based Methods

Generally, experimental studies use machine learning methods that integrate artificial intelligence systems. These methods have two stages, for prediction or classification, the first stage involves the study of the problem then the selection of the variables and the relevant model. And the second step involves applying the selected model to the test data and then measuring the performance of that model.

We can categorize machine learning techniques for novelty detection into two ctegories depending on the availability of the labels, supervised (Sup) which include classification and

regression algorithms and unsupervised (Unsup) which include clustering algorithms.

3.2.1. Supervised Methods

The classifier in supervised learning learns the models of which combination of characteristics lead to which labeled data whose functionality is associated with their labels [24]. Supervised learning problems can be further grouped into classification and regression problems.

3.2.1.1. Classification algorithms

A classification problem is the point at which the output variable is considered as a category. It's a technique where we categorize data into a given number of classes. The principal objective of classification-based technique is to attribute each data to the corresponding category/class. Classification can be categorized into k-Nearest Neighbors (kNN), Decision Tree (DT), Naive-Bayes (NB) and Support Vector Machine (SVM).

In conjunction with the Large Margin Nearest Neighbor (LMNN) metric learning algorithm, in [25], the authors have successfully adopted a kernelization technique in order to exploited non-linear distances learned from multi-class training data to detect novelties. They proposed a variation of the KNN classifier that utilizes the learned ideal distance which attempts to keep each known class' examples together while pushing cases from various realized classes to remain sensibly distant. The experiments show that learned measurements are powerful in improving contrasts with different measurements.

3.2.1.2. Regression algorithms

From the data entered in the system, the regression algorithms according to the input features predict the output values. The result in this case is based on learning the model during the training phase. Some of the popular types of regression

algorithms used for novelty detection are Logistic Regression (LR) and Random Forest (RF).

In order to detect anomalies of web traffic, in [26], the authors developed a LR model based on machine learning. They complete the feature extraction of the URL and train the logistic regression model by using the TF-IDF algorithm. The test results show that the model can separate abnormal traffic well for newly generated web traffic.

3.2.2. Unsupervised Methods

The training model in unsupervised techniques distinguishes unusual classes from a dataset without anyone else, whose dataset is only labelled with one label as ordinary [27].

3.2.2.1. Clustering Algorithms

Clustering based techniques are used mostly as an unsupervised learning method, they are considered as useful tools for data mining, compression, probability density estimation, and many other important tasks [28]. Clustering algorithms make it possible to recognize similar training data instance classes. Anomalies do not fit into any cluster; they can model a sparse cluster [24]. Clustering groups data instances into subsets in such a manner that similar instances are grouped together, while different instances belong to different groups [29]. And among the clustering algorithms, it is the K-means algorithm.

K-means clustering algorithm [30] serves automatically to partition a data set into k groups. It makes it possible to select k initial cluster centers and then iteratively refining them. For diverse practical applications, the k-means method is effective in producing good clustering results [31].

In the aim to improve the intrusion detection rate and decrease the false alarm rate. In [26], the authors focused on a detection method based on a K-means algorithm called MDKM to detect anomaly activities and corresponding simulation. They obtained the high-density parameters and cluster-partition parameters by calculating the distances between all sample data points and they gated the k clustering center accurately by using dynamic iterative process. The system achieves expectant aim since the results showed a higher detection rate and a lower false alarm rate.

3.3. Deep Learning Based Methods

Deep Learning is a new arising field of exploration in data science. It's basically a blend of artificial intelligence and machine learning. Deep Learning is a machine learning procedure that learns computers to do what works out naturally to humans [32]. It's a computerized reasoning capacity that emulates the activities of the human mind in processing data and making patterns for use in decision making [33]. There are diverse deep learning approaches that can uphold the objective of performing novelty detection or anomaly detection. Deep learning techniques can be comprehensively sorted into three categories: Supervised (Sup), unsupervised (Unsup) and semi-supervised (Semi-sup).

3.3.1. Supervised Methods
3.3.1.1. Convolutional Neural Network

Convolutional Neural Network (CNN) is the most representative model of deep learning, it's the neural network of choice for computer vision (image recognition) and video recognition [34].

As mentioned in [35], the CNN architecture offers the possibility of optimizing together several related tasks. By benefitting from the great learning capability of deep CNNs, the problems transformed from the classic computer vision challenges that have high dimensional data have been solved from a variety of viewpoints. Due to the advantages of CNN, it has been widely applied into many research fields among them the detection of novelty and anomaly.

3.3.2. Unsupervised Methods
3.3.2.1. Auto Encoder

Auto-encoder (AE) is the nonlinear generalization of PCA that uses an adaptive, multilayer "encoder" network to transform the high-dimensional data into a low-dimensional code and a similar "decoder" network to recover the data from the code [36].

The unsupervised technique AE attempts to understand an identity function to make the input signal and the output signal as reliable as could be expected. It's a particular neural network architecture with the objective of building as output the similar values utilized in input. As such, its goal is to get a proficient coding that considers the conditions among input values, performing a dimensionality reduction [37].

AE can potentially improve classification results by improving the capacity of learning features during training, decreasing dimensionality and extracting from original data the advanced features of abstract features. The authors in [38] proposed a deep learning method dependent on hybrid AE model formed by the blend of CAE (Contractive Autoencoder) and customary neural organization SRBM (Sparse Restricted Boltzmann Machine).

3.3.3. Semi-Supervised Methods
3.3.3.1. Generative Adversarial Network

Generative Adversarial Networks (GANs) by using alternative training techniques, estimates over a distribution density function. And it's considered as a type of deep generative model.

The authors in [39] introduced a utilization of GAN based models to distinguish anomalies in an industrial software system to forestall system outages that can weaken system availability using semi-supervised one-class learning. For the selected industrial system, the results showed that when contrasted with the traditional GAN based anomaly detection, the utilization of a variant of GAN known as bidirectional GAN (BiGAN) efficacy.

4. Novelty Detection Techniques Measurements

In this section, we give an overview of the novelty detection techniques measurements used to evaluate techniques of novelty detection. In order to evaluate this, standard metrics such as detection rate, accuracy, precision, recall, F-score and the Area Under the Curve (AUC) [40].

4.1. Detection Rate

Detection rate is a normally used measurement to check novelty detection approaches or techniques. The result of

recognition in detection rate can be grouped into four groups as following:

- True Positive is the proportion of positives that are effectively predicted.

- True Negative is the proportion of negatives that are effectively predicted.

- False Positive is the proportion of legal transaction predicted as illegal.

- False Negative is the proportion of illegal transaction predicted as legal.

4.2. Accuracy

Accuracy metric alludes to the proportion of all effectively arranged instances either typical or abnormal. The equation beneath shows how we can have determined accuracy as a percentage:

$$Accuracy = 100 \times \frac{(TP + TN)}{(TP + FP + FN + TN)} \qquad (1)$$

4.3. Precision

Precision is a measure that computes the number of positive forecasts that are accurately recognized as positive. It is defined as follows:

$$Precision = 100 \times \frac{(TP)}{(TP + FP)} \qquad (2)$$

4.4. Recall

Recall computes the number of positive instances are effectively anticipated are positive. It is otherwise called sensitivity or true positive rate. It is detailed as:

$$Recall = 100 \times \frac{(TP)}{(TP + FN)} \qquad (3)$$

4.5. F-score

The F-score is a measurement of accuracy that computes the weighted harmonic mean of recall and the precision. It is defined as:

$$F - score = 2 \times \frac{(Precision \times Recall)}{(Prcision + Recall)} \qquad (4)$$

4.6. Area Under the Curve

The AUC is acquired as a graph of the rate of true positives versus false positive rates for various thresholds. It is for the most part used to show their ability in classification in slanted and covering informational data sets.

5. Applications Domains and Discussion

Novelty detection essentially based on two elements. The first element is dedicated to normal activities and present activities in order to form behavior profiles for these activities, and the second element allows to recognize any form of deviation for the profiles that are tuned in the light of various techniques.

Novelty detection is applied in different areas of applications, and it is of great importance in applications that include huge data sets gained from critical systems. Different uses of novelty detection are enumerated below:

5.1. Financial Domain

Recently, there has been a renewed interest in research on this domain. We can define fraud as unjustified deception aimed to result in financial or personal gain. Fraud prevention and fraud detection systems are considered two main mechanisms to avoid frauds and losses due to fraudulent activities.

5.1.1. Identify Novelty

A general definition of a novelty within the context of fraud detection domain is fraudulent transactions while the fraudulent transactions are partitioned into two sorts according to the article [40]: online and offline fraud. Realizing that the online misrepresentation is makes by utilizing a taken actual card at the market store e.g., while the offline fraud is submitted by taking casualty identities, for example, name of the visa holder, MasterCard numbers, expiry date, and secret phrase.

5.1.2. Datasets

Each credit card memorizes information about transaction details like account numbers, client name, exchange area, time and size of transaction, MasterCard, buy types and merchant code. To examine the anomalies that can introduce a suspect case, these details are capable of deciding if a transaction is fraudulent or legitimate by using them in various algorithms.

Among the databases used by articles located in the following part we have "European data" [41]. As a feature of an association between Worldline company specialized in money transfer service and Libre Brussels University (LBU), the "European data" was initially collected by these two research teams based on transactions from European cardholders who performed within two days in September 2013. We have also "PaySim data" [42]. it has a very low fraudulent transaction rate (< 0.2 %). By using PaySim it is built of 1048575 Credit Card Transactions CCT, and 11 ascribes (highlights) with 1142 examples addressing fraudulent transactions and 1047433 different transactions addressed as real. There are several works that used a real dataset obtained from a private bank e.g.

5.1.3. Discussion

Abdelkbir ARMEL and Dounia ZAIDOUNI, in their work [43], they managed credit card fraud detection by using the

machine learning library (MLlib) of Apache Spark. And in order to distinguish between anomalies and valid transactions, they randomly generated the data used in their simulation following a normal distribution. They applied their data on four algorithms by giving a comparative analysis which shows that the RF algorithm gives the best accuracy with 98.18% among the four algorithms: Simple Anomaly detection algorithm, Decision Tree algorithm, Random Forest algorithm and Naïve Bayes algorithm.

Naoufal Rtayli and Nourddine Enneya have developed a hybrid approach for credit card fraud detection (CCFD) in their article [40]. The capacity of the proposed model to identify fraudulent transactions has been well shown in their article compared to the various studies already proposed. The Recursive Feature Elimination (RFE) for choosing the most valuable predictive feature, the GridSearchCV for Hyper-Parameters Optimization (HPO), and the synthetic Minority Oversampling (SMOTE) to defeat the imbalanced data problem are three sub-methods which build the secret of the robustness of the proposed model. To perform the model, they considered three real datasets, European data (DB1), PaySim data (DB2) and Data set 03 (DB3). In terms of accuracy, by using the datasets DB1 and DB2, the RFC (HPO, RFE) accomplishes 100%, and 99% by using the DB3 dataset. Regardless of the size of the data, this model maintains good precision performance, so it can be considered as a stable classifier model.

Table 1: Examples of application of novelty detection in financial domain

Ref	Techniques		Dataset	Metrics of evaluation		Year
[42]	Sup	NB	Generated randomly	Accuracy (%)	91.24	2019
		DT			96.96	
		RF			98.18	
		Simple Anomaly detection algorithm			77.04	
[40]	Sup + Unsup	RFE + HPO + SMOTE Synthetic	European data	Accuracy (%)	100	2020
			PaySim data		100	
			Data set 03		99	

5.2. Video Surveillance

Video Surveillance is an application area which gives an easy checking movement in a given climate. Video Surveillance systems are utilized in different areas like military domain, it is giving an incorporated view to the environment to react to the progressing exercises.

5.2.1. Identify Novelty

Novelty can be identified by various characteristics and challenges depending on the application domain. In the field of video surveillance, novelty can be defined by several objects. It can be a pedestrians' abnormal behavior, or abnormal movement

of a vehicle or an unexpected movement of a human. The novelty differs according to the desired problem solved.

5.2.2. Dataset

Due to the limitations of real datasets of novelty detection for video surveillance, many researchers construct the data manually to evaluate their methods. In the work [44] They have collected the dataset from a collection of videos with frames labeled as "normal" or "novelty". It is composed of three datasets which are traditional benchmarks for novelty detection problems in this domain, UCSD pedestrian dataset [45] which is composed of two subsets (Ped1 and Ped2) and Avenue Dataset [46].

The UCSD pedestrian is a video dataset was acquired by monitoring a pedestrian walkway using a stationary camera. The pedestrians are considered as normal frames and on the other hand bicycles, vehicles, skateboarders and wheelchairs passing through pedestrians are examined as anomalies. The Ped1 subassembly contains almost 8900 images with a resolution of 238×158 pixels of which 5500 are normal and 3400 are anomalous frames. The Ped2 subset includes 1998 images with a resolution of 360×240 pixels of which 364 are normal and 1652 are anomalous frames. The Avenue dataset is a video dataset captured at the Chinese University of Hong Kong campus avenue which regroup 16 video clip comprising 15,328 frames for training and 21 video comprising 15,324 frames for testing where the resolution is 640×360 pixels. The individuals walking by towards various directions are considered in the normal frames and the abnormal frames contain individuals running, tossing objects, and dallying.

5.2.3. Discussion

The authors in [44] used a Convolutional Auto encoder (CAE) in the novelty detection context. This method catches the 2D construction in image successions in the learning phase by applying in each frame the reconstruction error as an anomaly score. The input data was the blends of original frames and movement highlights and appearance. In order to gauge the "novelty level" of frames, they have utilized the regularized reconstruction error (RRE) and to evaluate the classification performance they have used AUC at various thresholds. The UCSD Ped, Avenue dataset and UCSD Ped2 acquired respectively the following AUC values 0.895, 0.754 and 0.547.

In the approach [47], in multiple instance learning (MIL), normal and abnormal videos were considered as bags, and instances were video segments, and to predict high anomaly scores, they learned a deep anomaly ranking model. This proposed method works better than the basic methods according to the experimental results applied to the data set which is made up of a variety of real world anomalies. He acquired an AUC measurement of 75.41.

In [48], the author proposed an approach which makes it possible to distinguish several objects of different sizes by pixel-wise foreground investigating framework. It is a technique which depends on two algorithms: Multi-layer perception recurrent

neural network (MLP-RNN) and Maximally Stable Extremal Region (MSER). The proposed technique produces a segmentation guide for the same spatial objects by taking as information a reference for which the anomaly does not exist and an objective edge. Recognized anomalies are represented by featured pixels. The classification accuracy, specificity and sensitivity were reached at rates of 98.56%, 96.05% and 98.21%, respectively. These results reflect improved precision, better classification and a reduced rate of calculation errors.

Table 2: Examples of application of novelty detection in video Surveillance domain

Ref	Techniques		Dataset	Metrics of evaluation		Year
[44]	Unsup	CAE	UCSD Ped1	AUC	0.895	2018
			Avenue		0.754	
			UCSD Ped2		0.547	
[47]	Sup	MIL	Collected manually	AUC	75.41	2018
[48]	Sup	MLP-RNN	Video surveillance system	Accuracy (%)	98.56	2020

5.3. Robotics

The use of mobile robots for several tasks is an attractive idea. Each robot is trained to identify sensory perceptions that signify potential or actual problems without requiring human intervention and can travel through environments that humans cannot. However, the technology aims to ensure that the robot does not miss any possible appearance of the problem (false negatives) while several methods and techniques of novelty detection can resolve this problem i.e. virtually impossible using conventional methods.

5.3.1. Identify Novelty

Regarding a robot's status, robotic events can predict notable amounts of information which can be extrapolated to detect productivity, anomalies, malfunctions and used for monetization [49].

In robotic systems, the successful detection of anomalies is a key issue in order to improve their maintenance and consequently reducing economic costs and downtime. Novelty or abnormalities in this area can be identified by abnormal robot gestures or out-of-vocabulary gestures as discussed in the discussion and perspectives section.

5.3.2. Dataset

Each model or methodology must be tested on a database that corresponds to a specific domain. Among the articles discussed in the next section, there is the paper [50], The UC2017 Static and

Dynamic Hand Gestures data set [51] and UC2018 DualMyo data set [52] are used to try the system proposed.

- The UC2017 data set presents the static and dynamic motion tests caught with an information glove and an attractive tracker. There are 2400 samples in the library which represent 24 classes of static gestures with a total of 100 repetitions for each of 24 classes. 19 classes among them are used to train the classifier and the 5 others are used as new models.
- The UC2018 DualMyo dataset contains 880 transactions equivalent to 8 classes of which each class includes 110 repetitions. The classifier was trained on 7 classes of which the 8th class is not banally separated from the others in an unsupervised way and was set aside to be the new model class.

5.3.3. Discussion

To addressing the issue of characterization of out-of-vocabulary motions, and by training in the Generative Adversarial Network (GAN) framework, the authors implemented in [50] a new technique utilizing Artificial Neural Networks (ANNs) in order to increase the data set online with novel created tests and the utilization of stochastic target vectors to diminish the normal prediction score. As we mentioned before, the UC2017 SG and UC2018 DualMyo data sets are used to test the model, to decide the impact of the proposed changes on the separation capacity of a neural network. According to the dataset, the classification precision was high 95.4% and 90.2% with a loss accuracy of 5% on the trained classes, this shows that the accuracy of novelty detection has been improved by the utilization of stochastic target vectors.

Table 3: Examples of application of novelty detection in robotics domain

Ref	Techniques		Dataset	Metrics of evaluation		Year
[50]	Semi-sup	ANNs	UC2017 SG	Accuracy (%)	90.2	2019
			UC2018 DualMyo		95.4	
[53]	Unsup	AE	Collected from the Gazebo simulations	Accuracy (%)	94.5	2019

In order to recreate the robot's sensor data and distinguish anomalies and based on an AE architecture that use both the spatial and temporal characteristics of the sensor data, and in order to detect sensor spoofing attacks against robots, authors in [53] have proposed a method for detecting anomalies. The collection of data started when the robots began moving and closed when the entirety of the robots were certain that they had completely investigated the region and accurately merged the map. The dataset used for the test contains many unique geometries for

Light Detection and Ranging LiDAR which are collected from Gazebo recreations and stored in ROSbag files [54] and also contain multiple moving objects traversing in and out of. vision. The approach has been tested against various types of attacks and the auto encoder has reached an accuracy of 0.945.

5.4. Social Media

Over the past twenty years, social media such as Facebook, Twitter or LinkedIn have seen a higher level of use by millions of users around the world, and their interaction with these social networks has affected their lives. Everyone discusses this latest news, or new technology. This information is presented on social networks in the form of text, links, images and videos.

5.4.1. Identify Novelty

Novelty detection generally presents an abnormal event. In fact, novelty is a general term used to refer to a variety of detection types according to the field of application of novelty detection. Novelty in social media domains may present sexual predators, malicious individuals, organized criminal behavior, community, fake accounts or illegal user activities.

5.4.2. Dataset

To make experiments and build the dataset for detecting fake accounts among twitter accounts, you need to get data from Twitter profiles e.g. The most used datasets in this kind of project get from 'The fake project' the Italian CNR dataset and Twitter API.

- The dataset @TheFakeProject is extracted from a twitter account which was created on December 12, 2012 and was used in the article [55]. This dataset contains two categories of profiles, 1950 of real accounts and 3351 of fake accounts. As mentioned in the technical report [56], the real accounts are collected from various sources like the #elezioni2013 dataset which comprised of 1481 confirmed profiles, in addition, the "Fake Project" team collected 469 verified human accounts. On the other hand, as mentioned in [57], the fake profiles were extracted from three sources http://fastfollowerz.com, http://twittertechnology.com and http://twittertechnology.com at a price of $19, $14 and $13 respectively.
- Using this API, Twitter allows us to interplay with all the attributes of these tweets. The results of requests sent to the Twitter API using a server-side scripting language are easily read since they are in JSON format. The Twitter API mainly consists of the following four objects: Tweets, Users, Entities and Places each of which contains many attributes. They have chosen some of them to add their dataset as indicated by their accessibility and appropriateness to the dataset [57]. Tweet objects that cannot be accessed by protected accounts are considered as the basic atomic building of all things.

5.4.3. Discussion

The authors in [58] introduce technique comprises two essential steps, the initial one is deciding the principle factors that impact a right detection of fake accounts, and the other step is for discovering the fake accounts by using the factors determined in the first step in a classification algorithm on Twitter accounts. This technique makes it possible to detect fake accounts that exist on Twitter social networks with maximum precision but using the minimum possible set of attributes depending on the requirement of the extended task of extraction, planning and examining features. They showed that the proposed approach allows to detect fake accounts with seven attributes from 22 attributes, even for other social networks such as Facebook with minor changes depending on the nature of the database. Among the best classification algorithms, they have applied Random Forest with accuracy equal to 91.40%, Decision Tree with accuracy equal to 90.08 %, Naïve Bayes with accuracy equal to 93.98%, Neural Network with accuracy equal to 93.80% and Support Vector Machine with the best accuracy equal to 94.26%.

Table 5: Examples of evaluation techniques

Ref	Techniques		Dataset	Metrics of evaluation		Year
[58]	Semi-sup	RF	TFP UC2018 DualMyo	Accuracy (%)	91.40	2019
		DT			90.08	
		NB			93.98	
		NN			93.80	
		SVM			94.26	
[59]	Semi-sup	AE	TFP	Accuracy (%)	98.44	2017
[60]	Semi-sup	CNMF	Collected from Sina Weibo	Precision	95	2017

6. Conclusion

Novelty, anomaly, outlier or exception are synonyms which present a pattern in the data that does not conform to the expected behavior. Since a machine learning system can never be trained on all possible sample classes, it must be able to recognize in the test phase the difference between known and unknown objects. Essentially for this reason, novelty detection has attracted significant attention within the research community and it's considered as an important learning paradigm.

This review provided an updated and curated survey of research articles of the novelty detection that exists in the literature during the last decade in machine learning and deep learning literature. We have presented a background of the related topics and applications done by a number of authors in novelty detection. We observed that novelty detection algorithms can extensively be classified into three principal classes depending on the nature of the dataset, statistical based techniques, Machine learning based techniques and Deep learning-based techniques.

In addition, we have given an overview of the novelty detection techniques measurements used to evaluate techniques of novelty detection such as detection rate, accuracy, precision, recall, F-score and the Area Under the Curve (AUC).

As discussed, novelty detection methods can be applied in various application areas such as financial domain, medical diagnosis, robotics, social media, and video surveillance. In the last section of this review, for each of these areas, we have defined the novelty, cited the most used dataset, as well as a description and perspectives of the latest work carried out in this field. The future work will be a proposal for a new approach using deep learning methods for social media domain.

References

[1] M. Amorim, F.D. Bortoloti, P.M. Ciarelli, E.O.T. Salles, D.C. Cavalieri, "Novelty Detection in Social Media by Fusing Text and Image Into a Single Structure," IEEE Access, **7**, 132786–132802, 2019, doi:10.1109/ACCESS.2019.2939736.

[2] E. Angola, "Novelty detection of rotating machinery using a non-parametric machine learning approach," in 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), 48–54, 2017, doi:10.1109/ICPHM.2017.7998304.

[3] A. Armel, D. Zaidouni, "Fraud Detection Using Apache Spark," in 2019 5th International Conference on Optimization and Applications (ICOA), IEEE, Kenitra, Morocco: 1–6, 2019, doi:10.1109/ICOA.2019.8727610.

[4] L.E. Baum, T. Petrie, G. Soules, N. Weiss, "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," The Annals of Mathematical Statistics, **41**(1), 164–171, 1970.

[5] J.A. Carino, M. Delgado-Prieto, D. Zurita, M. Millan, J.A.O. Redondo, R. Romero-Troncoso, "Enhanced Industrial Machinery Condition Monitoring Methodology Based on Novelty Detection and Multi-Modal Analysis," IEEE Access, **4**, 7594–7604, 2016, doi:10.1109/ACCESS.2016.2619382.

[6] J. Castellini, V. Poggioni, G. Sorbi, "Fake Twitter followers detection by denoising autoencoder," in Proceedings of the International Conference on Web Intelligence, Association for Computing Machinery, Leipzig, Germany: 195–202, 2017, doi:10.1145/3106426.3106489.

[7] D. Chen, P. Wang, L. Yue, Y. Zhang, T. Jia, "Anomaly detection in surveillance video based on bidirectional prediction," Image and Vision Computing, **98**, 103915, 2020, doi:10.1016/j.imavis.2020.103915.

[8] Credit card Fraud data - dataset by raghu543, Data.World, 2020.

[9] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, M. Tesconi, "Fame for sale: Efficient detection of fake Twitter followers," Decision Support Systems, **80**, 56–71, 2015, doi:10.1016/j.dss.2015.09.003.

[10] W. Dou, X. Wang, W. Ribarsky, M. Zhou, "Event Detection in Social Media Data," 4.

[11] R.O. Duda, P.E. Hart, D.G. Stork, "Part 1: Pattern Classi cation," 69, 1995.

[12] E. Marchi, F. Vesperini, F. Eyben, S. Squartini, and B. Schuller, "A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional LSTM neural networks," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015.

[13] A. ElAzab, A. M. Idrees, M. A. Mahmoud, H. Hefny, "Fake Account Detection in Twitter Based on Minimum Weighted Feature set," World Academy of Science Engineering and Technology, Johannesburg, South Africa, 2016.

[14] B. Erşahin, Ö. Aktaş, D. Kılınç, C. Akyol, "Twitter fake account detection," in 2017 International Conference on Computer Science and Engineering (UBMK), 388–392, 2017, doi:10.1109/UBMK.2017.8093420.

[15] A.B. Gardner, A.M. Krieger, G. Vachtsevanos, B. Litt, "One-Class Novelty Detection for Seizure Analysis from Intracranial EEG," Journal of Machine Learning Research, **7**(37), 1025–1044, 2006.

[16] M. Markou and S. Singh, "Novelty detection: a review—part 1: statistical approaches," 2003.

[17] M. Graeme, P. Gareth, W. Keith, "On the Long-Term Stability of Normal Condition for Damage Detection in a Composite Panel," Cardiff, UK, 2001.

[18] G. Manson, S.G. Pierce, K. Worden, T. Monnier, P. Guy, K. Atherton, "Long-term stability of normal condition data for novelty detection," in Smart Structures and Materials 2000: Smart Structures and Integrated Systems, International Society for Optics and Photonics: 323–334, 2000, doi:10.1117/12.388835.

[19] X. Guo, Z. Yuan, B. Tian, "Supplier selection based on hierarchical potential

[20] M. Hajabdollahi, R. Esfandiarpoor, E. Sabeti, N. Karimi, S.M.R. Soroushmehr, S. Samavi, "Multiple abnormality detection for automatic medical image diagnosis using bifurcated convolutional neural network," Biomedical Signal Processing and Control, **57**, 101792, 2020, doi:10.1016/j.bspc.2019.101792.

[21] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A. Ng, B. Liu, P.S. Yu, Z.-H. Zhou, M. Steinbach, D.J. Hand, D. Steinberg, "Top 10 algorithms in data mining," Knowledge and Information Systems, **14**(1), 1–37, 2008, doi:10.1007/s10115-007-0114-2.

[22] M. Hargrave, How Deep Learning Can Help Prevent Financial Fraud, Investopedia, 2020.

[23] M. Lasfar, H. Bouden, "A method of data mining using Hidden Markov Models (HMMs) for protein secondary structure prediction," Procedia Computer Science, **127**, 42–51, 2018, doi:10.1016/j.procs.2018.01.096.

[24] A. ARMEL, D. ZAIDOUNI, "Fraud Detection Using Apache Spark," in 2019 5th International Conference on Optimization and Applications (ICOA), 1–6, 2019, doi:10.1109/ICOA.2019.8727610.

[25] S.R. Silva, T. Vieira, D. Martínez, A. Paiva, "On novelty detection for multi-class classification using non-linear metric learning," Expert Systems with Applications, 114193, 2020, doi:10.1016/j.eswa.2020.114193.

[26] T. Liu, L. Zhang, "Application of Logistic Regression in WEB Vulnerability Scanning," in 2018 International Conference on Sensor Networks and Signal Processing (SNSP), 486–490, 2018, doi:10.1109/SNSP.2018.00097.

[27] B. Liu, W.S. Lee, P.S. Yu, X. Li, "Partially Supervised Classification of Text Documents," 13, 2002.

[28] V. Lopes, L.A. Alexandre, "Detecting Robotic Anomalies using RobotChain," in 2019 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), 1–6, 2019, doi:10.1109/ICARSC.2019.8733618.

[29] C. Lu, J. Shi, J. Jia, "Abnormal Event Detection at 150 FPS in MATLAB," 2720–2727, 2013.

[30] L.M. Le Cam, J. Neyman, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Weather modification, University of California Press, 1967.

[31] K. Alsabti, S. Ranka, V. Singh, "An efficient k-means clustering algorithm," Electrical Engineering and Computer Science, 1997.

[32] I.G. Maglogiannis, Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies, IOS Press, 2007.

[33] V. Mahadevan, W. Li, V. Bhalodia, N. Vasconcelos, "Anomaly detection in crowded scenes," in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1975–1981, 2010, doi:10.1109/CVPR.2010.5539872.

[34] R. Mohammadi-Ghazi, R.E. Welsch, O. Büyüköztürk, "Kernel dependence analysis and graph structure morphing for novelty detection with high-dimensional small size data set," Mechanical Systems and Signal Processing, **143**, 106775, 2020, doi:10.1016/j.ymssp.2020.106775.

[35] Z. Zhao, P. Zheng, S. Xu, X. Wu, "Object Detection With Deep Learning: A Review," IEEE Transactions on Neural Networks and Learning Systems, **30**(11), 3212–3232, 2019, doi:10.1109/TNNLS.2018.2876865.

[36] M. Murugesan, S. Thilagamani, "Efficient anomaly detection in surveillance videos based on multi layer perception recurrent neural network," Microprocessors and Microsystems, **79**, 103303, 2020, doi:10.1016/j.micpro.2020.103303.

[37] M.A. Russo, L. Kurnianggoro, K. Jo, "Classification of sports videos with combination of deep learning models and transfer learning," in 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 1–5, 2019, doi:10.1109/ECACE.2019.8679371.

[38] Z. Yang, H. Jing, "A deep learning method based on hybrid auto-encoder model," in 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 1100–1104, 2017, doi:10.1109/ITNEC.2017.8284911.

[39] T. Kumarage, S. Ranathunga, C. Kuruppu, N.D. Silva, M. Ranawaka, "Generative Adversarial Networks (GAN) based Anomaly Detection in Industrial Software Systems," in 2019 Moratuwa Engineering Research Conference (MERCon), 43–48, 2019, doi:10.1109/MERCon.2019.8818750.

[40] G. Muruti, F.A. Rahim, Z. bin Ibrahim, "A Survey on Anomalies Detection Techniques and Measurement Methods," in 2018 IEEE Conference on Application, Information and Network Security (AINS), 81–86, 2018, doi:10.1109/AINS.2018.8631436.

[41] E. Özbilge, "Experiments in online expectation-based novelty-detection using 3D shape and colour perceptions for mobile robot inspection," Robotics and Autonomous Systems, **117**, 68–79, 2019,

doi:10.1016/j.robot.2019.04.003.

[42] K. Pahwa, N. Agarwal, "Stock Market Analysis using Supervised Machine Learning," in 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 197–200, 2019, doi:10.1109/COMITCon.2019.8862225.

[43] T. Pourhabibi, K.-L. Ong, B.H. Kam, Y.L. Boo, "Fraud detection: A systematic literature review of graph-based anomaly detection approaches," Decision Support Systems, **133**, 113303, 2020, doi:10.1016/j.dss.2020.113303.

[44] M. Ribeiro, A.E. Lazzaretti, H.S. Lopes, "A study of deep convolutional auto-encoders for anomaly detection in videos," Pattern Recognition Letters, **105**, 13–22, 2018, doi:10.1016/j.patrec.2017.07.016.

[45] S. Rivera, S. Lagraa, A.K. Iannillo, R. State, "Auto-Encoding Robot State Against Sensor Spoofing Attacks," in 2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), 252–257, 2019, doi:10.1109/ISSREW.2019.00080.

[46] rosbag, accessed: 2018-03-27 - ROS Wiki, 2020.

[47] N. Rtayli, N. Enneya, "Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization," Journal of Information Security and Applications, **55**, 102596, 2020, doi:10.1016/j.jisa.2020.102596.

[48] S. Cresci, R. D. Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "A Fake Follower Story: improving fake accounts detection on Twitter," 2014.

[49] H.J. Shin, D.-H. Eom, S.-S. Kim, "One-class support vector machines—an application in machine fault detection and classification," Computers \& Industrial Engineering, **48**(2), 395–408, 2005, doi:10.1016/j.cie.2005.01.009.

[50] A. Shrestha, A. Mahmood, "Review of Deep Learning Algorithms and Architectures," IEEE Access, **7**, 53040–53065, 2019, doi:10.1109/ACCESS.2019.2912200.

[51] M. Simão, P. Neto, O. Gibaru, "Improving novelty detection with generative adversarial networks on hand gesture data," Neurocomputing, **358**, 437–445, 2019, doi:10.1016/j.neucom.2019.05.064.

[52] M. Simão, P. Neto, O. Gibaru, UC2017 Static and Dynamic Hand Gestures, 2018, doi:10.5281/zenodo.1319659.

[53] M. Simão, P. Neto, O. Gibaru, UC2018 DualMyo Hand Gesture Dataset, 2018, doi:10.5281/zenodo.1320922.

[54] W. Sultani, C. Chen, M. Shah, "Real-World Anomaly Detection in Surveillance Videos," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Salt Lake City, UT: 6479–6488, 2018, doi:10.1109/CVPR.2018.00678.

[55] Synthetic Financial Datasets For Fraud Detection, 2020.

[56] J. Wang, Y.-J. Chen, "A novelty detection patent mining approach for analyzing technological opportunities," Advanced Engineering Informatics, **42**, 100941, 2019, doi:10.1016/j.aei.2019.100941.

[57] What Is Deep Learning? \textbar How It Works, Techniques \& Applications, 2020.

[58] Y.Y. Yang, M. Mahfouf, G. Panoutsos, "Confidence Interval Assessment for Charpy Impact Energy Predictions – A Gaussian Mixture Model (GMM)-Based Approach," IFAC Proceedings Volumes, **44**(1), 11738–11743, 2011, doi:10.3182/20110828-6-IT-1002.01881.

[59] Yongqiao Wang, Shouyang Wang, K.K. Lai, "A new fuzzy support vector machine to evaluate credit risk," IEEE Transactions on Fuzzy Systems, **13**(6), 820–831, 2005, doi:10.1109/TFUZZ.2005.859320.

[60] D. Yu, N. Chen, F. Jiang, B. Fu, A. Qin, "Constrained NMF-based semi-supervised learning for social media spammer detection," Knowledge-Based Systems, **125**, 64–73, 2017, doi:10.1016/j.knosys.2017.03.025.

A S T E S

The Effect of Myocardial Fat's Thickness and Myocardial Impedance on Bipolar Radiofrequency Catheter Ablation Using Computer Simulation

Yao Sun[1], Keijiro Nakamura[2], Xin Zhu[*,1]

[1]*Biomedical Information Engineering Lab, The University of Aizu, Aizu-Wakamatsu, Fukushima, 960-0000, Japan*

[2]*Division of Cardiovascular Medicine, Toho University Ohashi Medical Center, Meguro, Tokyo, 100-003, Japan*

A R T I C L E   I N F O

A B S T R A C T

Radiofrequency catheter ablation is routinely used for the therapy of cardiac arrhythmias. Compared with the traditional unipolar ablation, bipolar ablation may improve the controllability of treatment, and prevent side effects and complications caused by catheter ablation. In addition, the variations of myocardial fat's thickness and myocardial impedance may have significant influence on the performance of bipolar ablation. In this study, computer simulation was performed to study the effects of myocardium fat's thickness and myocardial impedance on unipolar and bipolar ablation. The simulation demonstrates similar results with experimental ones using a swine heart. We observed that when the myocardial fat's thickness increases, bipolar ablation's heating effect and controllability may decrease. However, the final heating effect of bipolar ablation is invariably better than that of unipolar ablation. The ablation effects of unipolar and bipolar ablation are both reduced when myocardial impedance increases, while the heating effects of bipolar ablation are more sensitive to the variation of myocardial impedance and fat layers' thickness compared with unipolar ablation. The unipolar ablation is more stable in terms of fat, impedance and ablation time.

1. Introduction

Ventricular tachycardia (VT) is generally related to severe structural heart diseases (SHD) and accounts for an elevated risk of sudden cardiac death in patients with the heart failure [1]. It is also reported that for patients hospitalized with AMI, 5% to 10% of them have ventricular fibrillation (VF) or sustained VT prior to hospital presentation, and another 5% have VF or sustained VT after hospital arrival, most within 48 hours of admission [2]. A report about the incidence and case-fatality of ventricular events in hospital mentioned that the hospital case-fatality rates (CFRs) of VT declined from 27.7% to 6.9% and the incidence rates (IRs) declined from 14.3% to 10.5% from 1986 to 2009. This report concluded that the CFRs and IRs of VT are decreased, but it's still necessary to timely identify and treat patients with these serious ventricular arrhythmias [3]. Implantable cardioverter defibrillators (ICDs) remain the main method for the primary and secondary prevention of sudden cardiac death associated with these arrhythmias [4]. However, ICDs can only be used for treatment, not for the prediction or prevention of ventricular

arrhythmia [4]. ICDs are effective but their shocks are painful and can lead to post-traumatic stress disorders [5]. As a first line drug for arrhythmias, amiodarone can reduce arrhythmias for some patients but have limited efficacy for long-term management [6].

Radiofrequency (RF) catheter ablation has been proved as an effective treatment method for ventricular arrhythmias [7]. Particularly, the use of RF to treat atrial fibrillation (AF) has a history of more than 20 years [8]. Catheter ablation causes partial necrosis of lesion and therefore restore normal heart rhythms. Unipolar ablation (UA) is the standard configuration of catheter ablation, i.e., a singular electrode and a sheet ground on body surfaces [9]. UA is routinely used in clinical practice, but the effect and the treatment process of UA are expected to be improved [10][11]. Bipolar ablation (BA) can deliver concentrated energy between two electrodes. BA's focused energy delivery shortens ablation durations and minimizes lesion width; therefore, this reduces the potential of adjacent tissue injury. Although BA is still under study, it has been shown to be more effective than UA [12].

[*]Corresponding Author: Xin Zhu, Email: zhuxin@u-aizu.ac.jp

Computer simulation is an additional method for the study on the mechanisms of cardiac arrhythmias and the procedure of catheter ablation [13]. In [14], the author reported a systematic study on the effectiveness of different ablation line patterns based on a biophysical model. Analysis with a two-dimensional finite element method can also predict temperature distributions in ablation or hyperthermia [15], [16].

Due to the difference in density and other properties, the thermal and electrical conductivities of myocardial fat are significantly different in myocardium, blood and pericardial fluid [10]. In view of the low electrical conductivity of fat, the therapeutic effects of UA and BA may vary in different myocardial fat's thickness and myocardial impedance. Theoretically, when the thickness of cardiac fat increases, the voltage of myocardium would decrease because the impedance of fat is larger, and the electric current decreases as the impedance increases when the ablation power keeps the same value, and *vice versa*. As a result, the performance of ablation is reduced correspondingly. At present, the main research on radiofrequency ablation is the occurrence of complications and improvement of catheter ablation. However, there are few studies about the effect of myocardial fat's thickness and myocardial impedance on catheter ablation.

In this study, we conducted computer simulation to investigate the heating effects of UA and BA at different values of myocardial impedance and fat layer's thickness. To validate the results of computer simulation, we performed an *in vitro* experiment by using a swine's heart for comparison.

2. Methods and Materials

2.1. Configurations of Myocardial Fat Thickness and Impedance

We created a two-dimensional (2D) myocardial model based on the cross-section of the heart with a fat layer between the pericardial fluid and the myocardium as illustrated in Figure 1. To reflect most clinical conditions, fat thickness of 0-2mm was set [17]. Considering the pressure of the electrode on myocardium, the fat layer under the electrode invades has a distortion as the semidiameter of the electrode [18]. In this model, we set up a thin layer of pericardial fluid and blood with a thickness of 3 mm, and for myocardium, the thickness is 10 mm [10], [19]. The ablation
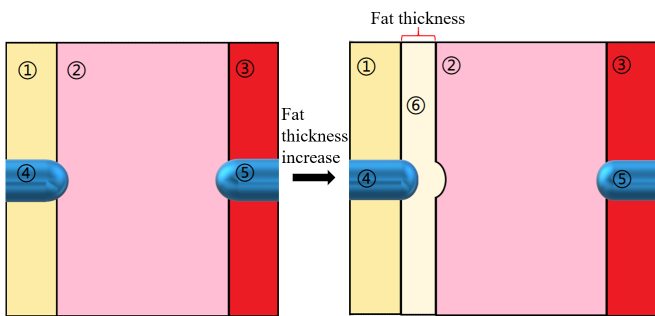
Figure 1: 2D model. ① Pericardial fluid, ②Myocardium, ③Blood, ④ Positive electrode, ⑤ Negative electrode, ⑥ Fat.

electrodes have a radius of 1.155 mm (7 Fr). The influence from blood and water flow was not considered to simplify this study.

In general, as the heating temperature of radiofrequency ablation does not exceed 100°C, the conductivity of body tissue in the model increases with temperature up to 100° C as follows,

$$z = z_0[1 + max \times (T - 37)/(100 - 37)], \quad (1)$$

where z is impedance, max the increasing percentage of z at 100°C, l the length and s the area. As the conductivity is proportional to impedance [18], we adopted the function of electricity in radiofrequency ablation, and impedance was considered increasing by 5%, 10%, 15% and 20% during the heating period, *i.e.*, max = 5%, 10%, 15%, and 20%. By

$$p = U^2/Re(z) \quad (2)$$

$$\frac{U^2}{Re(z)} = U^2 \frac{1}{Re(z_0)}(1 - \frac{\Delta Re(z)}{Re(z_0)}) \quad (3)$$

$$\frac{\Delta Re(z)}{Re(z_0)} = max \times (T - 37)/(100 - 37) \quad (4)$$

where p is power, and U voltage [20]. We calculated p from z_0 and U, and we also hypothesized that the rate of increase of impedance approximately equals to the decreasing rate of power at a constant voltage. In this study, $Re(z_0) = 100\Omega$ was set to calculate power. Figure 2 showed the variation of power with impedance.
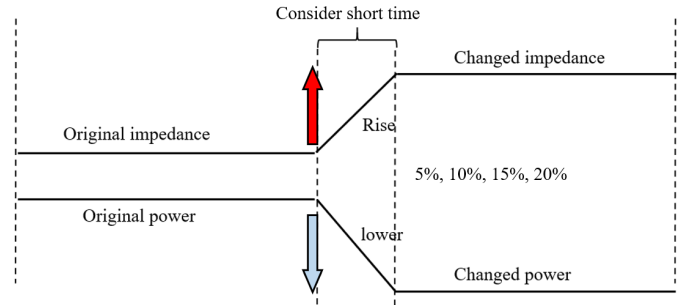
Figure 2: The power changed by increased myocardial impedance.

2.2. Numerical Analysis

In this study, we used the Finite Element Toolbox in MATLAB™ R2020b (Mathworks Inc., USA) to construct computer models and perform computer simulation. The simulation is converted to a procedure resolving partial differential equations, i.e., parabolic and elliptic functions in partial differential equation (PDE). Boundary conditions have been set to create electric and thermal fields [21]. The electrical frequency in radio frequency ablation is about 500 kHz. The biological medium can be considered almost totally resistive [10][15] because the human's body scales is about 0.5 m, which is much less than the wavelength of radio frequency wave. This means we can use a quasi-static approach to solve the electrical problem. Laplacian equation serves as the governing equation for calculating the electric potential within a 2D heterogeneous model [10].

$$\nabla(-\varepsilon\nabla V) = 0, \quad (5)$$

where V is the potential, and ε the permittivity varying spatially due to the heterogeneity of tissues. From the solution of (5), the

electric field E is got, and the specific absorption rate (SAR) Q can be expressed as follows,

$$E = -\nabla V \qquad (6)$$

$$Q = \frac{1}{2}\sigma|E|^2, \qquad (7)$$

where σ is the electric conductivity. Based on SAR, i.e., Q obtained above, and Penne's bioheat transfer equation as follows[15][22], we can obtain temperature distributions.

$$\rho c_p \frac{\partial T}{\partial t} - \nabla(\kappa \nabla T) = Q - \omega \qquad (8)$$

$$\omega = (F\rho)_t (\rho c_p)_b (T - T_b) \qquad (9)$$

In equation (8), ρ is the density, c_p the constant pressure specific heat capacity, κ thermal conductivity, T temperature, and Q SAR calculated by equation (7). In equation (9), ω is the heat taken away by the blood flow rate F, and T_b blood temperature, with an initial temperature of myocardium (37°C). 2-D FEM can be used to solve equations (5)-(9) numerically. Spatially and temporally variable physical parameters like current, potential and heat are mapped to first-order triangular elements [23]. Then,

equations (5)-(8) are discretized, converted into a matrix notation and solved with finite element method (FEM). Equation (8) is discretized spatially and temporally; therefore, a transient state is hypothesized to resolve this problem. Finally we got the temperature distribution [15].

We employed two boundary conditions including Dirichlet and Neumann boundary conditions. For the configurations of electric field, Dirichlet boundary condition is used to constrain the potential of electrode, and Neumann boundary condition is used to calculate electric currents, mainly setting the outer boundaries to be of air insulation. Figure 3 shows the boundary conditions in computer simulation, including the initial temperature and electrodes boundary conditions.

The initial temperature in BA and UA is 37°C. It is hypothesized that the temperature in the perpendicular direction of the model is the same, so that the temperature conduction is only affected by the electrode and the initial temperature in the 2D model. For the bipolar model in Figure 3(b), the left and right electrodes are positive and grounded (U=0V), respectively; for the unipolar model shown in Figure 3(a), the positive electrode is the same as that of the bipolar model, but the negative electrode is insulated (I=0A). Table 1 listed the parameters used in the computer simulation [10], [18].

Table 1: Parameters used in the computer simulation

Parameters	σ (S/m)	ε	κ (W/m/K)	ρ (kg/m³)	c (J/kg/K)
Pericardial effusion	1.35	118.00	0.63	1.02×10^3	4.18×10^3
Myocardial	0.61	113.00	0.60	1.02×10^3	8.30×10^{-6}
Fat	0.05	20.00	0.22	0.90×10^3	5.00×10^{-7}
Blood	1.10	118.00	0.54	1.06×10^3	4.18×10^3

σ, Conductivity; ε, electric constant; κ, Thermal Conductivity; ρ, Density; c, Specific heat capacity

2.3. Experiment

We conducted an *in vitro* experiment to perform catheter ablation using a swine's heart, thus we can validate the results of computer simulation. A glass tank filled with water at normal temperature(22℃) was employed to simulate the environment of human tissue, and a swine's myocardium was used to perform UA and BA. The power of 30W lasts for 30s and 50W for 10s. Figure 4 shows the experiment results of swine's myocardium after ablation.

Table 2 lists the experiment results for the models with 2 mm's fat layer and without fat, including the lesion depth and width in swine's myocardium under different parameters of ablation durations and powers. The power and duration of heating commonly used in clinical practice were used in the experiment. Based on the experimental results, even though the cooling effect of water may reduce the efficiency of ablation, the transmural damage was obtained in the myocardium with a thickness of 8 mm. The heating effects of BA are better than those of UA in any ablation configurations. The existence of myocardial fat has a significantly stronger effect on the results of BA, and we usually

Figure 3: The boundary conditions of electric and thermal fields for UA (a) and BA (b).

don't use 50W power in BA heating for 30s because the power is too strong, and may put patients at risk if the heating time is too long.

Table 2: Experiment Result Compared with Simulation Result

Experiment Result (environmental temp.: 22℃)				
Power/ time	**Lesion thickness (without fat)**			
	Uni. Diameter(mm)	Uni. Depth(mm)	Bi. Diameter(mm)	Bi. Depth(mm)
30W/ 30s	6.0	3.0	6.0	8.0
50W/ 10s	7.0	5.0	—	—
Power/ time	**Lesion thickness (with 2mm fat)**			
	Uni. diameter(mm)	Uni. depth(mm)	Bi. diameter(mm)	Bi. depth(mm)
30W/ 30s	6.5	3.0	5.5	5.0
50W/ 10s	7.5	4.5	—	—
Simulation Result (environmental temp.: 22℃)				
Power/ time	**Lesion thickness (without fat)**			
	Uni. diameter(mm)	Uni. depth(mm)	Bi. diameter(mm)	Bi. depth(mm)
30W/ 30s	6.5	3.5	6.5	10.0
50W/ 10s	7.0	5.5	—	—
Power/ time	**Lesion thickness (with 2mm fat)**			
	Uni. diameter(mm)	Uni. depth(mm)	Bi. diameter(mm)	Bi. depth(mm)
30W/ 30s	6.5	3.5	6.0	5.5
50W/ 10s	7.5	5.0	—	—

Figure 4: The results of a 30s 30w unipolar ablation experiment performed by a physician using an in vitro swine's myocardium. The lesion has a diameter of 6.00mm, and a depth of 3.00mm.

3. Results

Table 2 lists the simulation results obtained under the same configurations. Through comparison, the lesion thicknesses in the simulation are nearly the same but better than those in experiments. This may attribute to that the ablation parameters were more ideal in simulation. As the lesion thickness is the main therapeutic effect

index in catheter ablation, we may believe the results are basically the same.
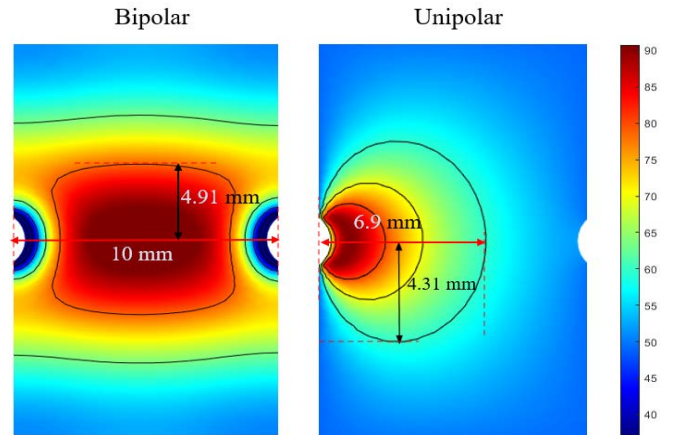
Figure 5: Simulation under 16W ablation for 120 s.

Simultaneously, in order to further ensure the accuracy of the simulation results, we also set the same ablation parameters based on a research report published in 2014 to validate the simulation [10]. We set an ablation configuration with a power of 16W ablation for 120 s. The results of the study are shown in Figure 5. Our simulation results are similar to those in [10], so we can ensure that our simulation results are basically consistent with the experiment and related research.

In this study, we selected central regions from the experimental results for calculating transmural lesion ratios (TLR), a 10 mm × 5 mm red rectangle shown in Figure 6. Thus we have:

$$TLR = \frac{area\ of\ temperature\ over\ 60\ ℃\ in\ 10\ mm×5\ mm}{the\ area\ of\ 10\ mm×5\ mm}. \quad (10)$$

We also visualized the temperature distribution in order to better observe heating areas as shown in Figure 6. The colormap was changed from colorful to gray levels from 0 to 255, and we found the appropriate threshold value to differentiate the temperature after testing, regions at 37℃ to 59℃ corresponded to the gray levels from 0 to 93 are marked in black, 60 ℃ to 100℃ corresponded to the gray levels from 94 to 254 marked in gray, and more than 100℃ corresponded to the gray level 255 marked in white.
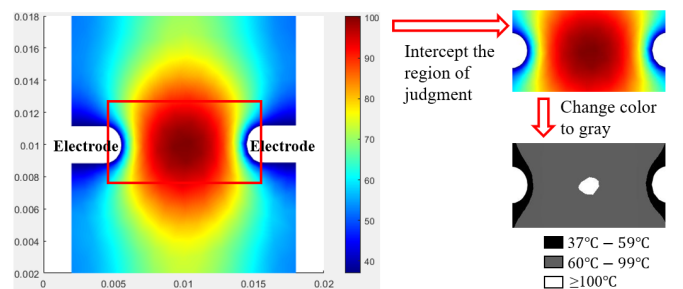
Figure 6: The visualization of temperature distribution.

At regions above 100℃, the blood vaporizes and clots to form a thrombus. This is a situation that needs to be avoided during

therapy. In the study, the simulation temperature may be too high due to the absence of cooling liquid, and we treat this as a tissue necrosis.
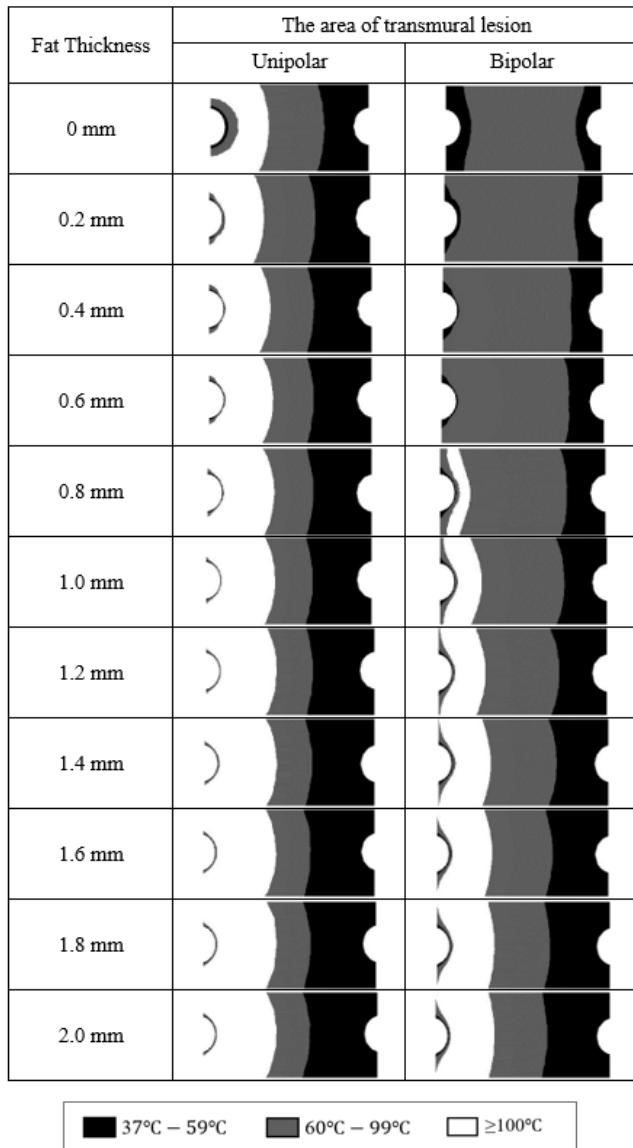
Figure 7: The area of transmural lesion of UA and BA under different fat thickness.

Figure 7 shows simulated heating results under the heating power of 30 W in UA and BA with ablation time of 60s. We observed that when the fat thickness is less than 1 mm, BA has obvious advantages over UA. BA can effectively control the high temperature range and maintain a wall penetration effect. When the fat thickness is greater than 1 mm, the controllability of BA gradually decreases. This heating results of UA are almost unaffected by fat thickness, but the local temperature is too high and transmural effect is not ideal in the UA.

Figure 8 shows the TLR of simulation under UA and BA with the variation of myocardial fat's thickness. Both UA and BA's heating effects show an overall reduced trend. The TLR of BA are better than that of UA, but with the increasing of the fat's

thickness, the TLR of BA decline slightly faster. When the fat's thickness is 0.2 mm, as heat begins to transfer to the fat, the temperature around the fat rises, leading to a sudden improvement in transmural heating effects. The TLR of UA also tends to be stable when the fat's thickness is over 0.2 mm.
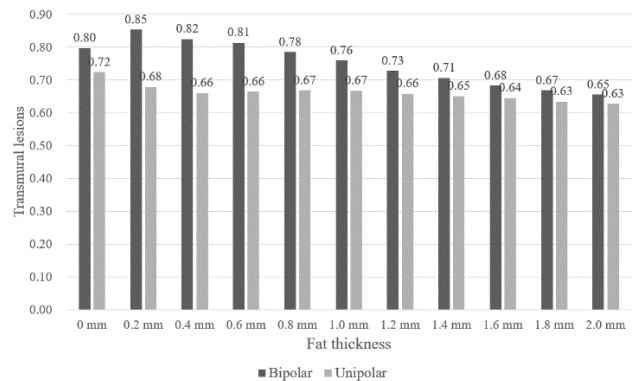
Figure 8: TLR of UA and BA under different fat thickness.

In this study, we considered the myocardial impedance increased 5%, 10%, 15% and 20% from the original impedance when the ablation power is 30W and sustains for 60 s. Figure 8 shows the TLR of simulated UA and BA under different heating time with increased impedance.

Figure 9 shows that the transmural lesions of both UA and BA generally decrease as the impedance increases. For BA, heating effect is poor at the beginning because the heat is still in the accumulation stage. When the heating duration is 30s to 40s, the transmural lesion increases rapidly and stabilizes after more than 40s. The final transmural effect of BA is stronger than that of UA. For UA, transmural lesion increases slowly and steadily. This indicates UA did not change drastically when impedance changed. As a result, the ablation effect of BA is more sensitive to myocardial impedance.

4. Discussion

This study compared the impact of myocardial fat's thickness and impedance changes on the heating effect of catheter ablation. From the simulation, we roughly identified a range of fat thickness, i.e., a fat thickness less than 1 mm, that BA has obvious advantages, including controllable temperature and a well transmural effect over UA. Because the UA has a radial range of action, fat does not change the direction of current conduction and UA is basically stable from the variation of fat's thickness and myocardial impedance. Therefore, UA is nearly unaffected by external conditions in case of excessive myocardial fat or impedance uncertainty.

In terms of ablation effect, we found BA demonstrated better results. The heat diffuses from the middle to ensure higher transmural ability, and the dispersed heat also reduces the burning temperature and ensures safety. Thus, for patients with lower BMI, BA might be a better choice. Less myocardial fat can reduce heat accumulation and lead to lower myocardial impedance. On the other hand, since the conductivity of fat is lower than that of

	10	20	30	40	50	60
original Uni	0.2872	0.4325	0.5269	0.6028	0.667	0.7239
original Bi	0	0	0.4213	0.6894	0.764	0.7972
Uni increase 5%	0.2749	0.4204	0.5169	0.5915	0.6495	0.7121
Bi increase 5%	0	0	0.3135	0.6599	0.7398	0.7801
Uni increase 10%	0.2601	0.4091	0.5078	0.5802	0.6403	0.689
Bi increase 10%	0	0	0.1816	0.6066	0.7076	0.7624
Uni increase 15%	0.2455	0.3979	0.4959	0.5706	0.6299	0.6784
Bi increase 15%	0	0	0.0805	0.5528	0.6856	0.7372
Uni increase 20%	0.2354	0.3853	0.482	0.5578	0.6193	0.6706
Bi increase 20%	0	0	0	0.4644	0.6475	0.7146

ABLATION TIME (s)

Figure 9: TLR of UA and BA under different impedance during ablation time.

myocardium, pericardial effusion and blood, fat thickness may be positively correlated with myocardial impedance, but their relationship remains to be determined by further studies.

In addition to the influence from myocardial fat's thickness and myocardial impedance, we also studied the ablation time to further understand the heating duration and characteristics of UA and BA. We found that BA is more sensitive to time variations. As far as the simulation conclusion is concerned, UA has a better ablation than effect 30s ago, and the BA effect is significantly improved, which is better than UA after 30 s. Time may be an important variable that affects BA. High-power and short-term ablation is the goal pursued by catheter ablation, but cooling water is necessary under UA in case of increasing power. Otherwise, blood steam evaporation may occur. Thus BA may avoid blood steam evaporation.

Using a 2D model for biological research has some advantages over 3D model because the calculation time of 2D model is greatly shortened. In contrast, the calculation time of a 3D model is longer, but the simulation results are closer to actual situations. Furthermore, flow of blood and cooling water can be easily introduced in a 3D model [24], [25]. These are hard to be set in a 2D model [26]; therefore, they are not considered in this study.

5. Conclusions

We perform a study about the effect of myocardial fat's thickness and myocardial impedance on catheter ablation using computer simulation. The results indicate that the thickness of fat layer and myocardial impedance change may have influence on the heating performance of BA to a certain degree. Therefore, for patients with lower BMI, BA might be a better choice. The effect of UA is more stable for patients with higher BMI. We also proved

that the BA has better ablation effects compared with those of UA when there are small variations of myocardial fat's thickness and myocardial impedance.

6. Future Work

In the future, we will extend the 2D cardiac model to a 3D cardiac model and consider the influence of blood flow and electrode types on the performance of UA and BA.

Acknowledgements

References

[1] W.G. Stevenson, K. Soejima, "Catheter Ablation for Ventricular Tachycardia," Circulation, **115**(21), 2750–2760, 2007, doi:10.1161/CIRCULATIONAHA.106.655720.

[2] S.M. Al-Khatib, W.G. Stevenson, M.J. Ackerman, W.J. Bryant, D.J. Callans, A.B. Curtis, B.J. Deal, T. Dickfeld, M.E. Field, G.C. Fonarow, A.M. Gillis, C.B. Granger, S.C. Hammill, M.A. Hlatky, J.A. Joglar, G.N. Kay, D.D. Matlock, R.J. Myerburg, R.L. Page, "2017 AHA/ACC/HRS Guideline for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society," Journal of the American College of Cardiology, **72**(14), e91–e220, 2018, doi:10.1016/J.JACC.2017.10.054.

[3] H. V. Tran, A.S. Ash, J.M. Gore, C.E. Darling, C.I. Kiefe, R.J. Goldberg, "Twenty-five year trends (1986-2011) in hospital incidence and case-fatality rates of ventricular tachycardia and ventricular fibrillation complicating acute myocardial infarction," American Heart Journal, 208, 1–10, 2019, doi:10.1016/J.AHJ.2018.10.007.

[4] K. Shivkumar, "Catheter Ablation of Ventricular Arrhythmias," Https://Doi.Org/10.1056/NEJMra1615244, **380**(16), 1555–1564, 2019, doi:10.1056/NEJMRA1615244.

[5] M.O. Sweeney, "The Contradiction of Appropriate Shocks in Primary Prevention ICDs," Circulation, **122**(25), 2638–2641, 2010,

doi:10.1161/CIRCULATIONAHA.110.000208.

[6] G.H. Bardy, K.L. Lee, D.B. Mark, J.E. Poole, D.L. Packer, R. Boineau, M. Domanski, C. Troutman, J. Anderson, G. Johnson, S.E. McNulty, N. Clapp-Channing, L.D. Davidson-Ray, E.S. Fraulo, D.P. Fishbein, R.M. Luceri, J.H. Ip, "Amiodarone or an Implantable Cardioverter–Defibrillator for Congestive Heart Failure," Http://Dx.Doi.Org/10.1056/NEJMoa043399, **352**(3), 225–237, 2009, doi:10.1056/NEJMOA043399.

[7] L.S. Klein, H.T. Shih, F.K. Hackett, D.P. Zipes, W.M. Miles, "Radiofrequency catheter ablation of ventricular tachycardia in patients without structural heart disease.," Circulation, **85**(5), 1666–1674, 1992, doi:10.1161/01.CIR.85.5.1666.

[8] M. Haïssaguerre, P. Jaïs, D.C. Shah, A. Takahashi, M. Hocini, G. Quiniou, S. Garrigue, A. Le Mouroux, P. Le Métayer, J. Clémenty, "Spontaneous Initiation of Atrial Fibrillation by Ectopic Beats Originating in the Pulmonary Veins," Http://Dx.Doi.Org/10.1056/NEJM199809033391003, **339**(10), 659–666, 2009, doi:10.1056/NEJM199809033391003.

[9] X. Zhu, D. Yang, W. Lu, W. Chen, D. Wei, K. Fukuda, H. Shimokawa, "Computer simulation of cathode ablation for atrial fibrillation," Proceedings - 2014 IEEE International Conference on Computer and Information Technology, CIT 2014, 721–725, 2014, doi:10.1109/CIT.2014.37.

[10] A. González-Suárez, M. Trujillo, J. Koruth, A. d'Avila, E. Berjano, "Radiofrequency cardiac ablation with catheters placed on opposing sides of the ventricular wall: Computer modelling comparing bipolar and unipolar modes," Http://Dx.Doi.Org/10.3109/02656736.2014.949878, **30**(6), 372–384, 2014, doi:10.3109/02656736.2014.949878.

[11] R.R. Tilz, A. Rillig, A.M. Thum, A. Arya, P. Wohlmuth, A. Metzner, S. Mathew, Y. Yoshiga, E. Wissner, K.H. Kuck, F. Ouyang, "Catheter ablation of long-standing persistent atrial fibrillation: 5-year outcomes of the hamburg sequential ablation strategy," Journal of the American College of Cardiology, **60**(19), 1921–1929, 2012, doi:10.1016/J.JACC.2012.04.060.

[12] G. SIVAGANGABALAN, M.A. BARRY, K. HUANG, J. LU, J. POULIOPOULOS, S.P. THOMAS, D.L. ROSS, A. THIAGALINGAM, P. KOVOOR, "Bipolar Ablation of the Interventricular Septum is More Efficient at Creating a Transmural Line than Sequential Unipolar Ablation," Pacing and Clinical Electrophysiology, **33**(1), 16–26, 2010, doi:10.1111/J.1540-8159.2009.02602.X.

[13] B. Avitall, "Computer simulation of ventricular tachyarrhythmias during coronary artery ligation and release," Journal of Electrocardiology, **12**(1), 17–22, 1979, doi:10.1016/S0022-0736(79)80040-0.

[14] A.F. Members, A.J. Camm, G.Y.H. Lip, R. De Caterina, I. Savelieva, D. Atar, S.H. Hohnloser, G. Hindricks, P. Kirchhof, E.C. for P.G. (CPG), J.J. Bax, H. Baumgartner, C. Ceconi, V. Dean, C. Deaton, R. Fagard, C. Funck-Brentano, D. Hasdai, A. Hoes, P. Kirchhof, J. Knuuti, P. Kolh, T. McDonagh, C. Moulin, B.A. Popescu, Ž. Reiner, U. Sechtem, P.A. Sirnes, M. Tendera, et al., "2012 focused update of the ESC Guidelines for the management of atrial fibrillationAn update of the 2010 ESC Guidelines for the management of atrial fibrillationDeveloped with the special contribution of the European Heart Rhythm Association," European Heart Journal, **33**(21), 2719–2747, 2012, doi:10.1093/EURHEARTJ/EHS253.

[15] N. Tsuda, K. Kuroda, Y. Suzuki, "An inverse method to optimize heating conditions in RF-capacitive hyperthermia," IEEE Transactions on Biomedical Engineering, **43**(10), 1029–1037, 1996, doi:10.1109/10.536904.

[16] P.M. Van Den Berg, A.T. De Hoop, A. Segal, N. Praagman, "A Computational Model of the Electromagnetic Heating of Biological Tissue with Application to Hyperthermic Cancer Therapy," IEEE Transactions on Biomedical Engineering, BME-**30**(12), 797–805, 1983, doi:10.1109/TBME.1983.325081.

[17] J.-W. Jeong, M.H. Jeong, K.H. Yun, S.K. Oh, E.M. Park, Y.K. Kim, S.J. Rhee, E.M. Lee, J. Lee, N.J. Yoo, N.-H. Kim, J.C. Park, "Echocardiographic Epicardial Fat Thickness and Coronary Artery Disease," Circulation Journal, **71**(4), 536–539, 2007, doi:10.1253/CIRCJ.71.536.

[18] M. Trujillo, E.B.-I.J. of Hyperthermia, undefined 2013, "Review of the mathematical functions used to model the temperature dependence of electrical and thermal conductivities of biological tissue in radiofrequency," Taylor & Francis, **29**(6), 590–597, 2013, doi:10.3109/02656736.2013.807438.

[19] C.E. Hamner, D.D. Potter, K.R. Cho, A. Lutterman, D. Francischelli, T.M. Sundt, H. V. Schaff, "Irrigated radiofrequency ablation with transmurality feedback reliably produces cox maze lesions in vivo," Annals of Thoracic Surgery, **80**(6), 2263–2270, 2005, doi:10.1016/J.ATHORACSUR.2005.06.017.

[20] Callegaro, Luca. Electrical impedance: principles, measurement, and applications. CRC Press, 2012.

[21] Pinder, George F., and William G. Gray. Finite element simulation in surface and subsurface hydrology. Elsevier, 2013.

[22] M. Ezzat, N. AlSowayan, … Z.A.-M.-H. and M., undefined 2014, "Fractional modelling of Pennes' bioheat transfer equation," Springer, **50**(7), 907–914, 2014, doi:10.1007/s00231-014-1300-x.

[23] T. Nakata, N. Takahashi, "Direct finite element analysis of flux and current distributions under specified conditions," IEEE Transactions on Magnetics, **18**(2), 325–330, 1982, doi:10.1109/TMAG.1982.1061870.

[24] A. González-Suárez, E. Berjano, J.M. Guerra, L. Gerardo-Giorda, "Computational Modeling of Open-Irrigated Electrodes for Radiofrequency Cardiac Ablation Including Blood Motion-Saline Flow Interaction," PLOS ONE, **11**(3), e0150356, 2016, doi:10.1371/JOURNAL.PONE.0150356.

[25] J.M. GUERRA, E. JORGE, S. RAGA, C. GÁLVEZ-MONTÓN, C. ALONSO-MARTÍN, E. RODRÍGUEZ-FONT, J. CINCA, X. VIÑOLAS, "Effects of Open-Irrigated Radiofrequency Ablation Catheter Design on Lesion Formation and Complications: In Vitro Comparison of 6 Different Devices," Journal of Cardiovascular Electrophysiology, **24**(10), 1157–1162, 2013, doi:10.1111/JCE.12175.

[26] L. Zou, L. Jing, V. Cvetkovic, "Shear-enhanced nonlinear flow in rough-walled rock fractures," International Journal of Rock Mechanics and Mining Sciences, **97**, 33–45, 2017, doi:10.1016/J.IJRMMS.2017.06.001.

A S T E S

Emotion Mining from Speech in Collaborative Learning

Nasrin Dehbozorgi[*,1], Mary Lou Maher[2], Mohsen Dorodchi[3]

[1]Department of Software Engineering, Kennesaw State University, GA, 30004, USA

[2]Department of Software and Information Systems, UNC Charlotte, NC, 27012, USA

[3]Department of Computer Science, UNC Charlotte, NC, 27012, USA

A R T I C L E I N F O

A B S T R A C T

Affective states, a dimension of attitude, have a critical role in the learning process. In the educational setting, affective states are commonly captured by self-report tools or based on sentiment analysis on asynchronous textual chats, discussions, or students' journals. Drawbacks of such tools include: distracting the learning process, demanding time and commitment from students to provide answers, and lack of emotional self-awareness which reduces the reliability. Research suggests speech is one of the most reliable modalities to capture emotion and affective states in real-time since it captures sentiments directly. This research, which is an extension of the work originally presented in FIE conference'20, analyses students' emotions during teamwork and explores the correlation of emotional states with students' overall performance. The novelty of this research is using speech as the source of emotion mining in a learning context. We record students' conversations as they work in low-stake teams in an introductory programming course (CS1) taught in active learning format and apply natural language processing algorithms on the speech transcription to extract different emotions from conversations. The result of our data analysis shows a strong positive correlation between students' positive emotions as they work in teams and their overall performance in the course. We conduct aspect-based sentiment analysis to explore the themes of the positive emotions and conclude that the student's positive feelings were mostly centered around course-related topics. The result of this analysis contributes to future development of predictive models to identify low-performing students based on the emotions they express in teams at earlier stages of the semester in order to provide timely feedback or pedagogical interventions to improve their learning experience.

1. Introduction

Social constructs have an essential role in the learning process which highlights the importance of teamwork and improving students' social skills in educational settings [1], [2]. Evaluating individual student's performance in teams is a complex process, especially in low-stake teams in which the product of teamwork does not have a large contribution to the students' grade [3]. In such teams, the main goal of teamwork is peer learning and developing interpersonal skills [1], [4]. Practicing low-stake teams is common in collaborative active learning classes [3].

In collaborative active learning, class time is utilized to create an engaging learning experience for students and improve their social skills [5]. There are multiple types of pedagogy that could

be classified as active learning, such as team-based learning (TBL) [6], cooperative learning [7], [8], collaborative learning [5], problem-based learning [5], or studio-based learning [9]. Most forms of active learning emphasize collaboration and the social construction of knowledge. During the past decade, active learning has been applied more often in higher education, and it highlights the importance of having structured protocols for team formation, team size, and assessment [10], [11]. The inherently complex nature of teamwork in active learning calls for methods to assess students' performance at both individual and team levels [12], [13]. This is more critical in low-stake teams which are used in introductory-level courses. To address this evaluation issue theories on team performance, converge in identifying attitude components such as emotional states that influence team performance [13]-[18].

*Corresponding Author: Nasrin Dehbozorgi, Dnasrin@kennesaw.edu

Here the question is how to operationalize attitude constructs as factors to be measured. Research shows that a survey is commonly used for this purpose [13]. Attitude has been traditionally measured by having students fill out self-report forms, typically using a Likert scale to express their feelings towards a specific subject [13]. However, the self-report tools have some drawbacks such as: lack of student commitment to fill them out, not providing precise answers, and lack of awareness about their emotion at the moment. More advanced tools exist to allow researchers to retrieve emotional information by observing facial expressions, gestures, posture, and periods of silence [19]-[22]. Although these tools may provide more precise results on capturing emotion compared to self-reports, still they are not error-free, and deploying such tools requires both technical and human resources which may not be cost-effective. These drawbacks as well as the potential distraction of the learning process can make such tools less practical in educational settings and classrooms. Researchers suggest speech is a very good source to identify emotion, however because of challenges such as environmental noise level it is not commonly practiced [23].

This paper focuses on emotion analysis as a subset of affect and measures the correlation between the emotions that students express in low-stake teams with their performance in the course. We identify students' emotions as they work in dyad teams by recording their verbal conversations and analyze how emotional states correlate with their performance. Identifying these relationships can help us in developing tools to monitor students' emotions and identify at-risk individuals at earlier stages of the semester.

In the next section, we discuss different dimensions of teamwork and methods of evaluating students' performance at both team and individual levels in educational settings and introduce emotion as an attitude construct that is associated with team performance. In section 3, we present our methodology, study design, and data collection protocol followed by the results of data analysis. Finally, we conclude by discussing the takeaways and the plan for future work.

2. Background

Evaluating students' involvement in teams and providing timely feedback are key components to a good team experience. Researchers have proposed different tools and insights on evaluating team performance over the past 30 years. The review of the literature of teamwork assessment shows there are four primary ways to collect data on team performance: 1) self-report, 2) peer assessment, 3) observation, and 4) objective outcomes. For optimal results, it is suggested to combine both qualitative and quantitative data collection [13]. For a comprehensive teamwork evaluation, we need to collect information from different sources [13]. Gathering the necessary information from all perspectives of teamwork requires a group of team observers [24]. In general, it is not easy for instructors to evaluate team dynamics inside and outside the classroom to verify a fair and successful team experience for every individual. For this reason, peer and self-evaluation are common methods of assessing individuals' contributions in teamwork. Surveys are employed for this kind of assessment and these surveys can include Likert scales, partner ranking, descriptive word matching, short answers about peers,

and journaling about their effort and experiences [25]. Finally, the weight of the grade is either provided by the instructor in the form of a standard rubric, or the weight of each component is negotiated between the instructor and the students [26]. Teams have outcomes at both team and individual levels, therefore measuring teamwork at the individual level is important. The team-level outcome is the final product which is the result of the effort of all team members, while at the individual level the outcome can be the team member's attitude or contribution to the teamwork process [13].

In low-stake teams, no significant team-level final product is produced to be evaluated as a team performance metric. In these types of teams, students do not have assigned roles, and since teamwork activity has a low contribution to final grades, there is a high chance that students rely on their partners and don't participate in team activity as expected. This situation has severe consequences in collaborative active learning since students will not utilize the class time to learn course material through team activities. In such cases, the emphasis of the team evaluation should be more at the student level and their attitude towards teamwork, in order to provide timely feedback to them.

Lack of quantitative and objective measures of teamwork at the individual level is a barrier for evaluating teams' performance and assembling effective teams [27]. Most metrics to evaluate individuals in teams rely on experts who observe and rate teams by using rubrics or based on qualitative dimensions like leadership and team structure [27]. Here the question is what data needs to be collected and what factors should be used for evaluating individuals in low-stake teams. Recent research shows that attitudinal factors in teamwork can determine team effectiveness [28]. In this research, the first step is to identify the attitude that individuals pose in teams [29]. In the following section, we discuss how the affect (combination of emotion and mood) as attitudinal dimension is associated with performance in the educational setting and discuss the methods for measuring emotional states.

2.1. Affect and Emotion

Affect as an attitudinal attribute plays an important role in students' learning in educational settings [30]. The affective domain which includes both emotions and moods shows how much a person values the learning process, their willingness to contribute to learning new things, their ability to make a decision, and how generally one behaves in different situations [31], [32]. Emotional obstacles can negatively impact students' learning process while positive feelings of joy, happiness, and satisfaction positively influence students' learning [33]. Research shows that students who experience emotional difficulties that are not identified early may not receive appropriate feedback and can result in lower performance [34]. Furthermore, affective states not only impact the performance but also, influence the interpersonal relationships in teams in educational settings [32]. Therefore, considering affective states and emotional awareness is important for both students and instructors [35]. For students, emotional awareness empowers them with the required skills to manage their emotions and establish positive relationships with peers and handle challenging situations effectively [32]. This is critical in collaborative active learning in which the learning and cognitive

process are integrated into teamwork [35]. Identifying students' emotional state is important for instructors since it can lead to both cognitive and affective scaffolding, which can improve social knowledge construction [35].

Psychologists claim a person can be known based on three domains called the ABCs of attitude which is: Affect, Behavior, and Cognition [36]. Cognition can be measured based on the learning outcome and behavior can be observed, but affect is not directly observable and is more challenging to measure [31]. There is a wide range of methods for measuring affective states. These methods vary from psychological measures such as heart rate, diagnostic tasks, self-reports, facial expressions, and knowledge-based methods to derive affective state in a given context like time of day and length of the task and individual journals [23]. The most common form of measuring affect is using self-report surveys [13] which has certain drawbacks as discussed in the previous section. More advanced tools such as EEG technology have been applied to capture emotional states through brain signals or by observing facial expressions, gestures, posture, and periods of silence [37]-[27]. Although these tools provide better results on capturing emotion compared to individual surveys and self-reports, still they are not error-free, and deploying these tools and using experts as observers is demanding in the educational domain for a large number of students.

To date, most of the existing tools to identify learners' emotional states use self-report instruments such as the Achievement Emotions Questionnaire (AEQ) [35], [40]. Fewer researchers recognize emotional states by doing sentiment analysis on students' journals and learning diaries or from their conversations in forums or other asynchronous textual data [41]. These methods are applied to either identify the polarity of students' emotions or the expressions of six fundamental emotions of anger, trust, surprise, sadness, joy, fear, disgust, and anticipation [41]. Although sentiment analysis is more promising to evaluate emotion, it has not been widely applied in the educational domain compared to the social media or e-commerce review corpora due to data limitations [33], [41].

Researchers claim speech is one of the best tools to extract information about emotion, however because of challenges in data collection it is not widely used in the educational domain [23]. The selection of a suitable method for emotion analysis depends on different factors such as the type of emotions to be identified. the resources required to collect data, and the context in which the task is performed [23]. For example, in some contexts using self-reports may be more appropriate than using sensors, since sensors can cause interference with the task being performed by the students [23].

One of the innovations of this research is assessing students' emotions based on their verbal speech in teams. The advantage of this approach is that emotion is captured directly, it minimizes distraction during the learning activity, and data collection can be done at a large scale. In the next section, we introduce our methodology to operationalize emotional states and elaborate on how we design the study.

3. Methodology

In this research, we identify correlations between emotion and students' performance in low-stake teams. We further conduct aspect-based sentiment analysis to identify the topics in which the students expressed the most positive or negative emotion. The result of this analysis can help us identify the areas in which students have more challenges and apply pedagogical interventions. We formulate the following null hypothesis:

H_0: There is no correlation between a student's positive sentiment in low-stake teams and their individual performance in the course.

In our study, speech is the modality of data collection for measuring and analyzing students' emotions. We recorded students' speech as they talked in teams during the class activity throughout the semester. The recorded audio files were then transcribed for text mining and emotion analysis. The data processing phase includes the following steps: 1) noise reduction by applying filters on the audio to improve the quality, 2) audio transcription, 3) assigning IDs to each individual based on the voice recognition, and 4) removing the speech utterances from the third person (i.e., speakers other than the team members). As a result, the transcription of all speech utterances of every participant was stored as one data point. This resulted in 28 datasets to feed into the text mining and sentiment analysis algorithms. This process is illustrated in Figure 1.
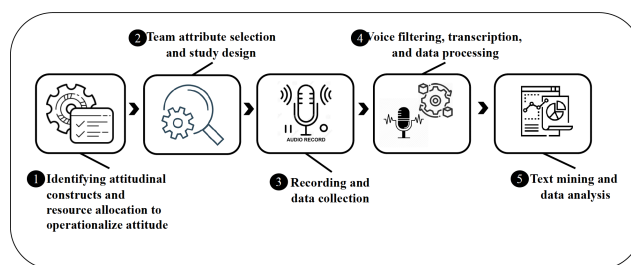
Figure 1: Data Collection and Analysis Process

We used different methods to analyze students' sentiments in speech and identified their correlation with performance. First, was measuring the polarity, frequency, and intensity of the sentiments in three classes of positive, negative, and neutral. Next, different classes of emotion were extracted.

Finally, we performed thematic analysis of the sentiments (aspect-based sentiment analysis) to explore the context and themes in which students expressed more positive sentiment as they spoke in the class. In the following, we explain how each method for sentiment analysis was conducted in this study.

3.1. Polarity Sentiment Analysis

The first step for polarity sentiment analysis was the segmentation of the datasets so that each vector is the speech initiation point, meaning each vector consists of the speech of a participant until it is finished or interrupted by the other teammate. As a result, the number of vectors in each dataset denotes the number of times a participant initiated the speech either in an active or reactive mode. The visualization of the text-mining algorithm for polarity and intensity sentiment analysis is presented in Figure 2.
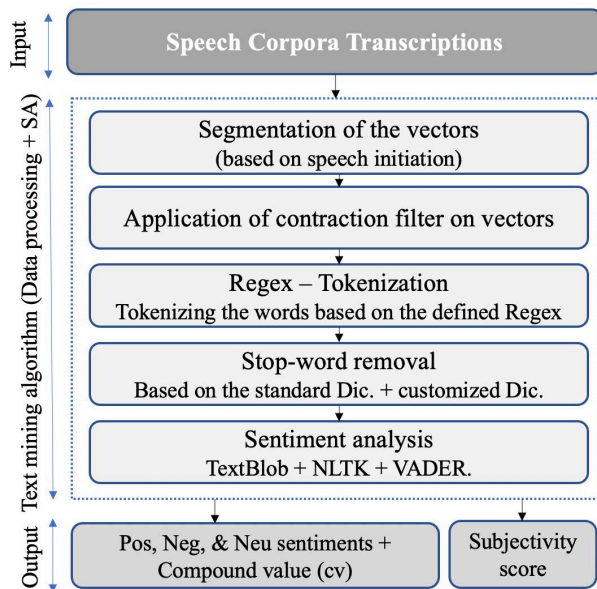
Figure 2: Algorithm for Polarity and Intensity Sentiment Analysis

After applying the contraction filtering on the vectors, we developed a regular expression for tokenizing the vectors and eliminating extra characters which did not impact the sentiment score. We used both the standard English dictionary and also created a custom dictionary to remove the stop words and as well as the redundant words that speakers used habitually and didn't impact the sentiment score. For developing the customized dictionary, we measured the frequency of unigrams in each dataset, and the most frequent common unigrams across all datasets were included in the dictionary to be removed from the corpora. Examples of such words are 'yeah', 'ok' that students uttered repeatedly.

Finally, we analyzed the sentiments from the processed data by applying lexicon-based and rule-based NLP python tools; TextBlob, NLTK, and VADER. The VADER algorithm was applied due to its higher precision and accuracy on short tokens at string-level compared to other sentiment analysis tools [42]-[44]. Moreover, many sentiment analysis tools are either polarity-based or valence-based but VADER measures both the polarity and valence of the input. VADER outputs four classes of 'Negative', 'Neutral', 'Positive', and 'Compound' with values between -1 to 1. The compound score is the adjusted normalized value of the sum of valence scores of each word in the lexicon. The equation of compound value is presented in Equation (1) [45]:

$$Compound\ value = \frac{sum\_val}{\sqrt{(sum\_val)^2 + 15}} \quad (1)$$

where sum_val is the sum of the sentiment arguments passed to the score_valence() function in the VADER algorithm.

We used compound value (cv) as a metric for the unidimensional measure of sentiment. The threshold for classifying positive, negative, and neutral sentiments can vary based on the context. The typical thresholds are [12][46]:

Positive sentiment: cv >= 0.05

Neutral sentiment: -0.05 < cv < 0.05

Negative sentiment: cv <= - 0.05

In this study, we conducted the k-means clustering on the compound values to determine the right threshold for the data. The Elbow method was used to identify the optimum number of clusters. The Elbow method explains the percentage of the variance as a function of the number of clusters meaning the optimum number of clusters is defined such that adding one more cluster does not provide better modeling of the data [47]. The best number of clusters is chosen based on the 'Elbow criterion', which is shown as an angle in the graph (Figure 3.a) [47]. The Elbow criterion determines three as the optimum number of clusters. The 3-means clustering of data shows the most density of compound values in the cluster with cv= 0 as shown in Figure 3.b.
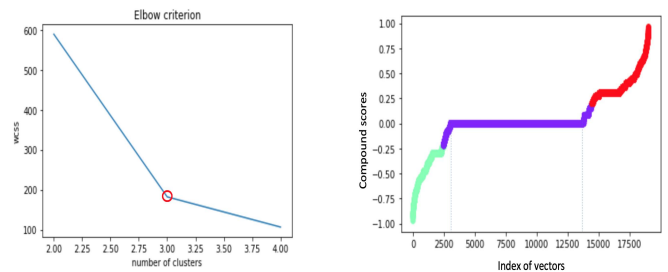
3.a. Elbow Criterion   3.b. 3-Means Clustering

Figure 3: Elbow Criterion and 3-Means Clustering of cv

As a result, we considered zero as the threshold for classifying cv into positive, neutral, and negative sentiments. This means any vector with cv < 0 was labeled as negative, vectors with cv > 0 were considered as positive, and vectors with cv = 0 were treated as neutral sentiment classes.

Next, we identified multiple classes of sentiments from speech corpora. The methodology for doing so is described in the following section.

3.2. Multi-class Sentiment Analysis

In addition to analyzing the polarity of the sentiments, we analyzed how different types of emotions contributed to students' performance. For extracting multiple classes of emotions, we applied the LIWC (Linguistic Inquiry and Word Count) which is an efficient text analysis tool for analyzing diverse emotional and cognitive components in verbal or written speech corpora [48]. The core of the LIWC is the LIWC2015 dictionary which includes

almost 6400 word-stems and selected emotions. In this tool, a single word such as "cry" that carries emotion can belong to different categories such as sadness, negative emotion, and overall affect [48]. Each category of LIWC2015 includes a list of dictionary words that define the relevant scale. The complete list of scales and scale words can be found in the LIWC2015 Development manual [48].

In this study, we selected specific dictionary categories to identify which emotions more contribute to individuals' performance. The total number of 93 dimensions in all 4 sub-dictionaries exist in the LIWC2015, out of which we identified 63 dimensions to be most relevant to the context of this study. These

dimensions exist in the four sub-dictionaries of 1) psychological processes, 2) Linguistic processes 3) punctuation, 4) other grammar. The proportion of selected dimensions in each sub-dictionary were 67% psychological processes, 24% linguistic processes, 6% other grammar, and 3% punctuations. The scale score of each 63 dimensions is a numerical value measured by using the LIWC framework.

In the next step, we applied Principal Component Analysis (PCA) technique to conduct factor analysis. PCA is an unsupervised statistical method to reduce the dimensionality of large and complex data [49]. It identifies the interrelation between features and reduces the number of features by creating a new feature (component) based on the linear combination of initial features. The main steps of PCA are 1) removal of the target feature and class labeling by creating a d-dimensional matrix, 2) calculating the covariance matrix of the whole dataset, 3) computing eigenvectors (e_1, e_2, ...e_d) and corresponding eigenvalues (λ_1, λ_2, ..., λ_d), 4) sorting the eigenvectors by decreasing eigenvalues and choose k eighteen vectors with the larger eigenvalues to form a d*k dimensional matrix (in this study, we considered the threshold of 1 and chose the number of components (k) with eigenvalues > 1), 5) using the d*k eigenvector matrix to transform the samples onto the new subset.

The PCA method is widely applied in developing predictive models. After dimensionality reduction, the dataset is divided to test and train datasets to fit the data into the training algorithm. In this study, the goal is to identify the correlation of different emotions with performance and identify the ones that most contribute to performance. However, the development of the predictive models is not the focus of this study and will be done in future research. The next step of analyzing emotions is aspect-based sentiment analysis, which is described in the following section.

3.3. Aspect-based Sentiment Analysis

The last step of sentiment analysis in this study was the aspect-based analysis and exploring the themes of the positive sentiments. We applied word frequency analysis and measured the frequency of the tokens both unigrams and bi-grams in the positive sentiment corpora. After identifying the most frequent word tokens we identified the proportion of course-related tokens among them.

3.4. Data Collection

The data for this study was collected from an active learning CS1 class in which students worked in low-stake teams to do the class activities. Based on the research [13] and our empirical evidence, in order to improve students' engagement, we decided to form teams of two students. Team formation was done subjectively, and they were mixed of both males and females with diverse backgrounds. We formed the teams at the beginning of the semester by a gamified activity in which students were paired based on their month of birth. Since students were mostly involved in the team formation game, we noticed high levels of satisfaction between teammates throughout the semester.

The class included 65 students among which 48 participated in this study. We collected the recorded speech of 48 participants

throughout the whole semester, however, due to some technical issues or absence of students during some sessions, the recording of 28 individuals (14 teams of two students) was considered in this study due to their consistent pattern of attendance in the class.

The class was held for 75 minutes during two days of the week. In each class, the first 15-20 minutes was dedicated to resolving student's misconceptions from the previous session followed by a low-stake clicker quiz and a 15-minute mini-lecture. The remaining class time (about 40 minutes) was dedicated to peer learning class activities in the form of low-stake teamwork.

During data collection, we faced multiple environmental and technical issues or human errors which made the data unavailable for analysis. For example, sometimes students pressed the stop button on the recorder accidentally, or sometimes it was the battery or hardware issue which led to the loss of data. On top of that, environmental noise level and cross talk in class when all team members were talking at the same time was a challenge in data collection. We employed some protocols to overcome these challenges such as training specific TAs for the recording-related tasks or encouraging students to sit in certain locations to minimize the noise level and improving the quality of the recording. To overcome the hardware-related challenges, we identified recording devices that had features like bidirectional paired microphones, built-in noise cancellation, lasting battery, high-level user-friendliness, and low cost. In the next section, the result of the data analysis is presented, and the null hypothesis is evaluated.

4. Data Analysis

The main goal of data analysis is to identify the correlation between positive emotions and students' performance and identifying emotions that can serve as predictive metrics.

The performance metric is considered to be students' grades in the course. Students' performance was assessed in a formative style during the semester by three major assignments, three lecture tests, and three lab tests. Each milestone had a certain contribution to the final grade: assignment 20%, lecture test 30%, and lab test 30%. Participants' grade distribution is presented in Figure 4.
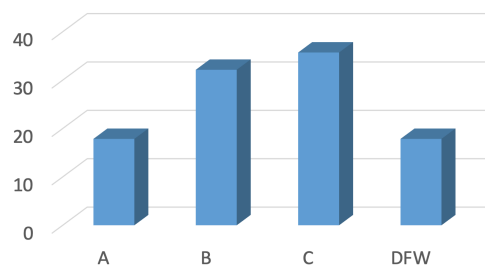
Figure 4: Grade Distribution of the Participants

The linear regression analysis of the performance scores over time determines the slope of trendlines (i.e., the ratio of change over time). Accordingly, the participants were categorized into three groups; upward trend (slope >0): 57%, downward trend (slope <0): 28%, no trend (slope = 0): 14%.

The data shows that most of the participants had an upward trend which means their performance increased during the semester. In the rest of this section, we present the analysis result.

4.1. Polarity and Intensity of Sentiments

To analyze the polarity of the sentiments NLTK and VADER algorithms were applied on the 28 datapoints to classify each vector into four classes of Positive (Pos), Neutral (Neu). Negative (Neg) and Compound (Comp). As described in the methodology section, the vectors with cv > 0 are considered to be positive sentiments. The intensity and frequency of positive sentiments were normalized based on the amount of speech. Frequency refers to the number of vectors with cv > 0 and intensity refers to the actual score of cv in each vector. Equations (2) and (3) show how the mean frequency and intensity scores are measured.

$$Frequency = \frac{m}{n} \tag{2}$$

$$Intensity = \frac{\sum_{x=1}^{m}(comp\_value_x)}{n} \tag{3}$$

where:

n= total number of vectors in each dataset

Vectors with positive comp_value= {1,2, …m}

We identified a homogeneous pattern between frequency and intensity of positive sentiments (cv > 0), based on the linear regression analysis. This means the students with higher frequency of positive sentiments had higher levels of intensity in their positive emotions. The linear regression plot of the frequency and intensity of positive sentiments is presented in Figure 5.

The Spearman's rank correlation coefficient was applied to identify if there is a positive relationship between students' positive sentiments (intensity and frequency) and their performance. Spearman's rank correlation coefficient is a nonparametric (distribution-free) rank statistic for measuring the strength of an association between two variables [50]. It evaluates how the relationship between two variables can be described using a monotonic function, without making any assumptions about the frequency distribution of the variables [50]. The coefficient value (r_s) can be anywhere between -1 and 1, and the closer r_s is to +1 and -1, the two variables have a stronger monotonic relationship. The r_s value is calculated using equation (4) and the strength of the correlation is interpreted based on the absolute value of r_s; r_s = 00-.19 "very weak", r_s =.20-.39 "weak", r_s = .40-.59 "moderate", r_s = .60-.79 "strong", r_s = .80-1.0 "very strong" [51].

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{4}$$

where d_i is the difference in ranks for variables and n is the number of cases. The calculated r_s values are presented in Table 1.
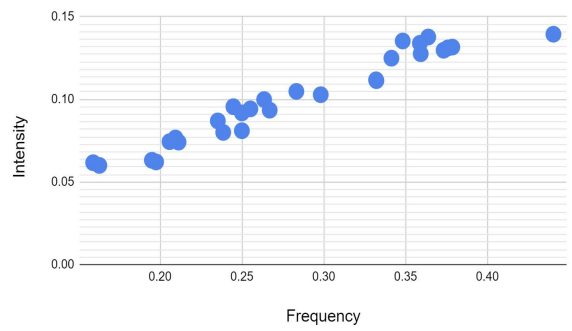
Figure 5: Linear Regression of Frequency and Intensity of Positive Sentiments

Table 1: Coefficient Values of Positive Sentiments and Performance

	Coefficient Value (r_s)	
Frequency and performance	.61	Strong positive correlation
Intensity and performance	.64	Strong positive correlation

The coefficient values of both intensity and frequency of positive sentiments show that both have a strong positive correlation with performance.
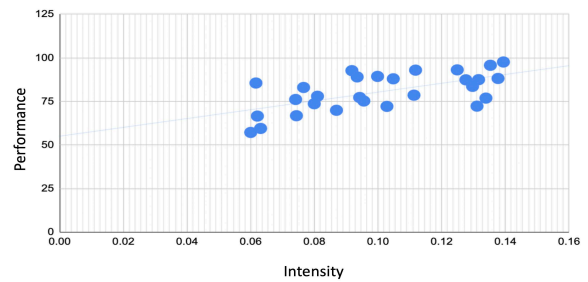
Figure 6: Regression plot of the Intensity of Positive Sentiment vs Performance

The linear regression plot of intensity and frequency of positive sentiments vs performance are presented in Figures 6 and 7. In these plots, the horizontal axis shows the intensity and frequency of positive sentiments, and the vertical axis shows the performance score.

To test the null hypothesis (H_0), we applied the chi-square test and measured the two-tailed p-value with the confidence level of 0.05. The Null hypothesis states: There is no correlation between students' positive sentiment in low-stake teams and their individual performance in the course.
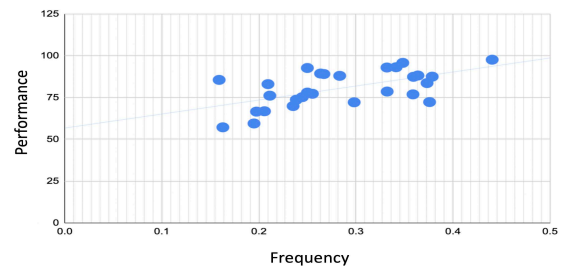
Figure 7: The Regression Plot of Frequency of Positive Sentiment vs Performance

The calculated p-value for the intensity of positive compound value is 0.002 and the p-value for the frequency of compound value is 0.001. The p-values for both the intensity and frequency of positive sentiments are less than the confidence level (0.05), therefore the null hypothesis is rejected which confirms that there is a statistically significant correlation between intensity and frequency of positive sentiment and students' performance. The calculated two-tailed p-values are presented in Table 2.

Table 2: Two-tailed p-values of Positive Sentiment and Performance

	Tow-tailed p-value	
Frequency and performance	.001	Statistically significant correlation
Intensity and performance	.002	Statistically significant correlation

The next step in our analysis is classifying students' sentiments into multiple classes beyond positive, negative, and neutral to investigate which classes can be used as predictive metrics for performance as the target value. The result of multiclass sentiment analysis is presented in the following section.

4.2. Multi-class sentiment analysis

The LIWC text analysis tool was applied to extract sentiment class from the corpora. The selected 63 dimensions of LIWC are in psychological, affective, and cognitive domains. The output features of each domain scale differently, in other words, the range of features in one domain might be greater than the features in another one. For example, the word count (WC) features are greater than the affective features such as negative emotion (negemo). To standardize the distribution of the data we applied the StandardScaler() algorithm from the SKLEARN library. This algorithm standardizes the features by removing the mean and scaling to the unit variance as shown in equation (5) [52]:

$$z = (x-u)/s \qquad (5)$$

where u is the mean and s is the standard deviation of the training sample.

The output features of LIWC have very high dimensionality, which makes it challenging to interpret the features and identify the main ones for developing predictive models. In order to reduce the dimension of the feature space to the most critical ones and preserving as much 'variability' (i.e., statistical information) as possible, we used the Principal Component Analysis (PCA) method. To determine the minimum number of principal components, we calculated the cumulative proportion of variances that the components explain. The acceptance level of cumulative variance can vary from 80% to 95% depending on the application. We identified 14 principal components with 92% variance which is an acceptable number to represent data. Figure 8 shows the Scree plot of the LIWC features.

The heatmap visualization of the 14 principal components and the 63 LIWC features is presented in Figure 9. The horizontal axis shows the total of 63 original features, the vertical axis denotes the identified 14 principal components and the color bar on the right side represents the correlation between the original features

and the principal components. The lighter color indicates a more positive correlation exists between the features and the principal component and the darker colors show a negative correlation.
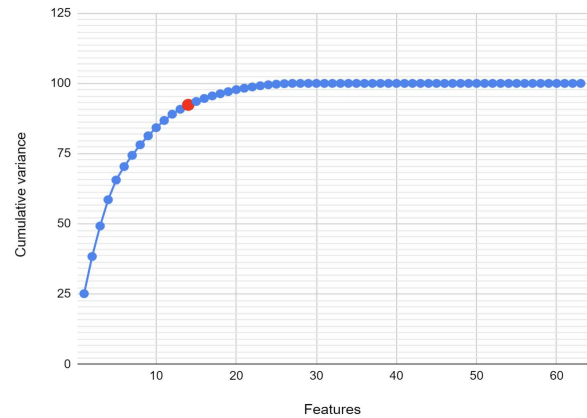
Figure 8: Scree Plot of LIWC Dimensions
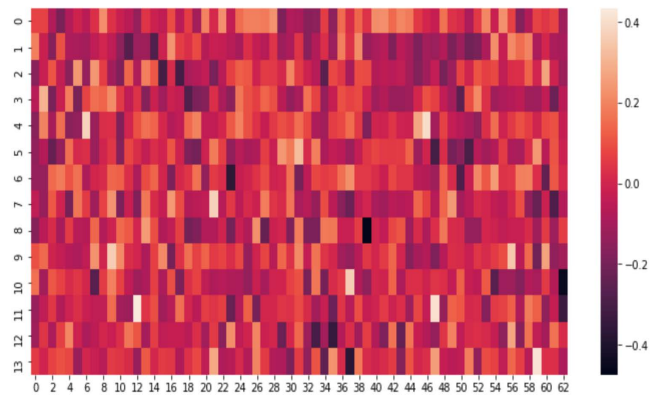
Figure 9: The Heatmap Visualization of Principal Components vs. Features

Table 3: Mapping of Principal Components (PCs) to the Original LIWC Features

PCs	LIWC Feature	PCS	LIWC Feature
PC_1	Anxiety	PC_8	Negative emotion
PC_2	Negations (e.g., no, not, never)	PC_9	Perceptual process (hear)
PC_3	Common verbs	PC_10	Personal pronounce (she/he)
PC_4	Analytical thinking	PC_11	Exclamation marks
PC_5	Drives (including affiliation, achievement, power, reward focus, risk focus)	PC_12	Auxiliary verbs
PC_6	Insight (e.g., think, know, consider)	PC_13	Certainty
PC_7	Anger	PC_14	Non-fluencies (e.g., Er, hm, umm)

To interpret each principal component in terms of the original variables, we measured the magnitude of the coefficients for the original variables such that larger absolute values of coefficients confirm corresponding variables have more importance in calculating the component. The maximum absolute values of the coefficients are mapped to the original LIWC features in Table 3.

The table 3 lists the most principal features that determine the target value which is students' performance.

The identified features can be applied to train the predictive algorithms to predict student's performance based on their psychological, affective, and cognitive behaviors. Fitting the sample data into predictive models is not the scope of this study and will be considered in future work. In the next section, the result of thematic analysis (aspect analysis) on students' positive sentiments is presented.

4.3. Aspect-based sentiment analysis

After identifying the correlation between the positive sentiments and the performance we explore the theme of the positive sentiments to identify whether students expressed a more positive emotion toward the course content or any other topic or experience. In order to identify the themes, we applied word frequencies algorithms to parse the positive sentiment vectors and extract unigrams and bigrams tokens. The frequency of unigrams and bigrams determines the dominant theme in students' positive sentiments.
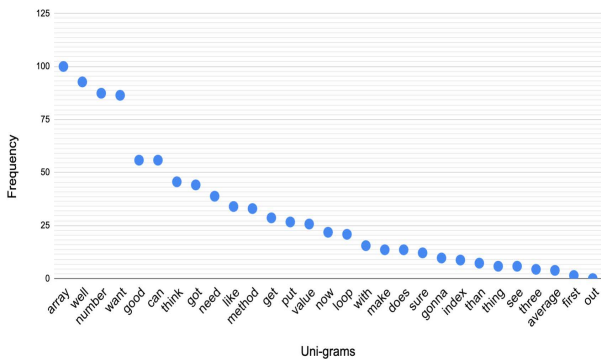
Figure 10: Frequency of Unigrams in Positive Sentiment Vectors (Comp>0)

Figures 10 and 11 plot the normalized frequency of the top 30 unigrams and bigrams of vectors with cv > 0. Based on these plots the dominant theme of word tokens that most frequently were uttered in students' speech is specifically about the course content (i.e., Array, random number).
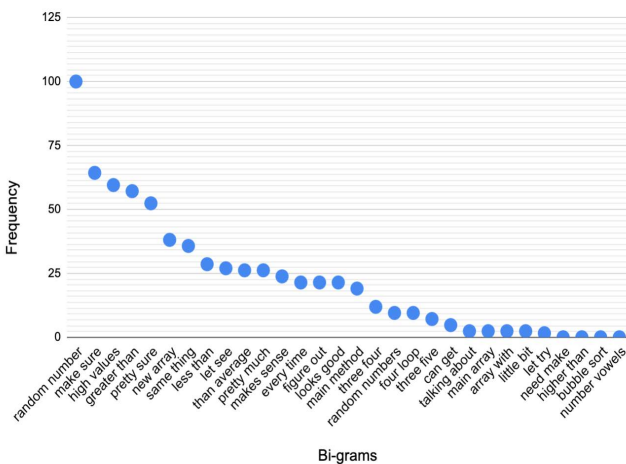
Figure 11: Frequency of Bi-Grams in Positive Sentiment Vectors (Comp>0)

The proportion of course-specific unigrams and bi-grams tokens in the vectors with cv>0 is presented in Figures 12 and 13.
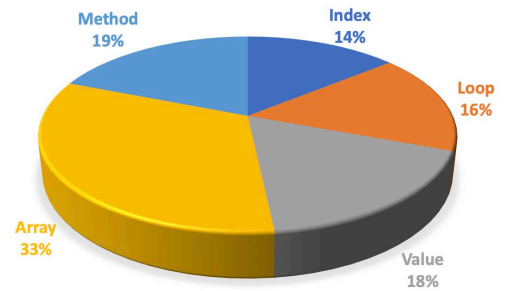
Figure 12: Course-Specific Unigrams in the Vectors with cv>0
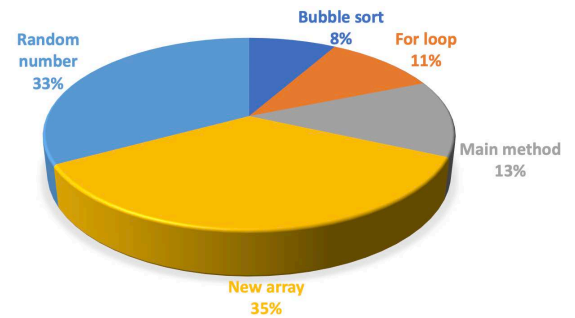
Figure 13: Course-specific Bi-Grams in the Vectors with cv>0

Based on the frequency analysis we conclude that students' positive sentiments were mainly centered on course-related topics. This finding also confirms that the course design was effective in engaging the students in the teamwork activity since they were timed activities that allowed the well-prepared students to finish them within 35-40 minutes and they did not have extra time to discuss the topics outside the course content.

5. Discussion

The k-means clustering of the output sentiments showed that most vectors fall into the neutral class which is the cluster where cv=0 (figure 3.b). We further analyzed students' speech to identify the correlation between the neutral sentiments (cv = 0) as well as the subjectivity level in their speech and their performance. We further conducted a thematic analysis on the negative sentiments' corpora (vectors with (cv<0) to identify the themes in which students expressed negative emotions.
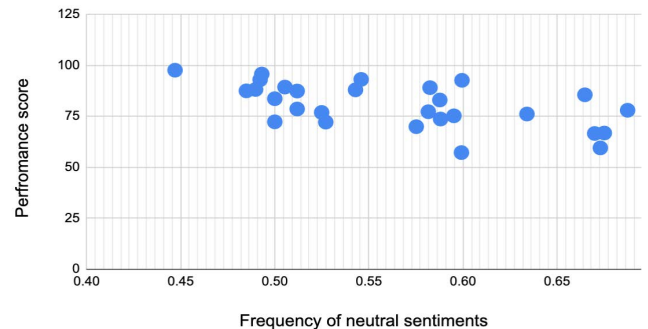
Figure 14: Frequency of Neural Sentiments vs Performance Score

To identify the correlation between neutral sentiments and performance we applied the VADER tool to extract neutral sentiments (cv =0). The regression plot of the frequency of neutral sentiments vs performance is presented in Figure 14. In this figure, the horizontal axis is the frequency of the neutral

sentiments, and the vertical axis is the performance. We observe a homogeneous pattern in which by decreasing the performance the frequency of neutral sentiments increases.

The Spearman's rank correlation coefficient was conducted and the calculated r_s value is -0.61. It confirms there is a strong negative correlation between these two variables of neutral sentiments and performance.

Next, we analyzed the speech to identify the correlation between the level of subjectivity and performance. We applied TextBlob which is a rule-based sentiment analysis tool to extract the subjectivity metric [53]. The output subjectivity score is a float number between 0.0 and 1.0 where 0.0 is very objective and 1.0 is very subjective [54]. We applied Equation (6) to calculate the subjectivity score of each participant.

$$\text{Subjectivity Score} = \sum_{x=1}^{n} Sx \quad (6)$$

where, Sx is the subjectivity score of each vector ($0<=Sx<=1$), and n is the total number of vectors in each dataset.

The subjectivity level vs. performance score is plotted in Figure 15, where the vertical axis is the performance score, and the horizontal axis is the subjectivity score.

Based on Spearman's rank correlation coefficient the calculated r_s value is -.02, which shows a very weak negative correlation exists between the level of subjectivity and students' performance.
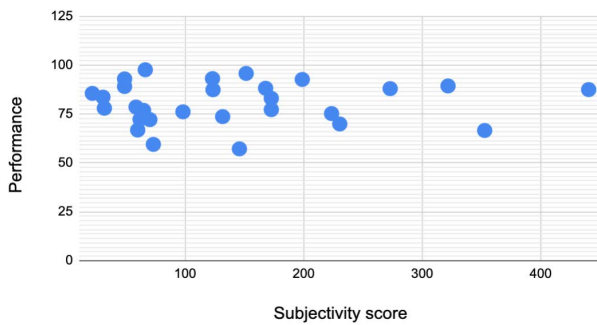
Figure 15: The Regression Plot of Subjectivity vs Performance Score
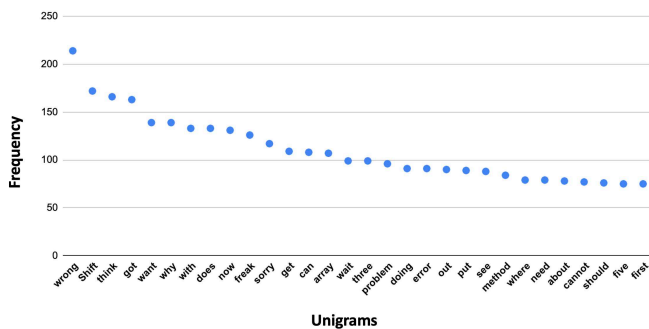
Figure 16: Frequency of Negative Sentiment Unigrams

Next, we conducted a thematic analysis on negative sentiment vectors (cv < 0) by calculating the frequency of unigrams. The result shows that students expressed more negative sentiments

when they made mistakes and faced a problem in solving the problems in the class activities. For the purpose of this study, we have replaced the swear words with "shift" and "freak".

The scatter plot of the frequency of the unigrams in the negative sentiment vectors (cv<0) (Figure 16) shows the most frequent token in negative sentiments is the word "wrong".

And finally, it is worth mentioning that we did not find a statistically significant correlation between vectors with cv <0 and the performance score in this study.

6. Conclusion

Teamwork and collaboration are critical aspects of cognition and learning. In low-stake teams, students don't have assigned roles and tasks since they are more focused on formative learning tasks than final artifact production. This might lead to a lack of contribution from some team members which as a result may lead to poor peer learning and performance. Assessing student's contribution and performance in low stake teams is thus challenging for instructors. We posit that attitude constructs and individuals' emotions can be assessed as an indicator of performance.

In the existing literature, emotion is mainly operationalized in the form of surveys and self-reports [13] which have their own challenges and may not necessarily provide reliable results. A major drawback of surveys is the lack of commitment from participants to provide responses.

Text analysis methods are another way to operationalize sentiments and have been applied to students' textual conversations on discussion forums or blogs [41]. The drawback of these methods is that they are conducted on asynchronous conversations and don't provide sentiment analysis in a real-time context. Speech is an alternative to capture emotional states. Recording and speech analysis have their own challenges, such a being demanding in terms of time, effort, and resources. Due to the environmental noise level, this method of data collection is not often used in educational settings.

In this article, we proposed a novel approach to operationalize students' sentiments by recording their real-time speech in low-stake teams during class. Finding the correlation between different emotions and students' performance can help in developing models to predict students' performance at earlier stages of the semester to provide timely feedback to them and apply pedagogical interventions. The main goal of this research is to identify the correlation between students' positive emotions as they work in teams during class with their performance.

The novelty of our research is sentiment analysis based on speech. This method has minimum distraction for students during the class activity and does not require additional time to answer the surveys. We implemented our method in a CS1 class during one semester with 28 participants. The result of data analysis showed the students who had higher scores in frequency and intensity of positive sentiments performed better in the course. On the other hand, the students who had more frequency in their neutral sentiments had lower performance scores in the course. We did not identify a statistically significant correlation between negative sentiments and students' performance. The result of

thematic analysis on the positive sentiments confirms that the theme of the students' positive emotions were the topics related to the course content.

This study was conducted on 28 samples thus the results of the data analysis may not generalize to other courses or other populations of students. However, the results provide a basis for future research to analyze larger samples of data and for developing predictive models.

The result of this research can benefit both students and instructors. Instructors can identify at-risk students in earlier stages of the semester and help them develop their social skills and learning experience.

6.1. Future work

In the next phase of this research, we aim to collect more samples to study the correlation of positive emotion with performance. In this research, the principal components from multiple features (classes of emotion) were identified that determine the performance. In future work, we will fit these principal features to train machine-learning algorithms to predict performance.

Another future direction is to develop algorithms for automatic and real-time transcriptions of the speech to be able to provide real-time analysis as students work in teams. The development of such a system that provides emotional feedback to both instructors and students as they work in teams helps students to develop emotional awareness and adjust their behavior and interaction with peers. This system can also enable the instructor to observe the climate of different teams and cue them to provide timely feedback to students.

References

[1] N. Dehbozorgi, M. Lou Maher, M. Dorodchi, "Sentiment analysis on conversations in collaborative active learning as an early predictor of performance," in 2020 IEEE Frontiers in Education Conference (FIE), IEEE: 1–9, 2020. DOI: 10.1109/FIE44824.2020.9274119

[2] G. Salomon, D.N. Perkins, "Chapter 1: Individual and social aspects of learning," Review of Research in Education, 23(1), 1–24, 1998. https://doi.org/10.3102/0091732X023001001

[3] C. Latulipe, N.B. Long, C.E. Seminario, "Structuring flipped classes with lightweight teams and gamification," in Proceedings of the 46th ACM Technical Symposium on Computer Science Education, 392–397, 2015. https://doi.org/10.1145/2676723.2677240

[4] N. Dehbozorgi, S. MacNeil, "Semi-automated Analysis of Reflections as a Continuous Course," in 2019 IEEE Frontiers in Education Conference (FIE), IEEE: 1–5, 2019. DOI: 10.1109/FIE43999.2019.9028636

[5] M. Prince, "Does active learning work? A review of the research," Journal of Engineering Education, 93(3), 223–231, 2004. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

[6] M.K. Smith, W.B. Wood, W.K. Adams, C. Wieman, J.K. Knight, N. Guild, T.T. Su, "Why peer discussion improves student performance on in-class concept questions," Science, 323(5910), 122–124, 2009. https://doi.org/10.1126/science.1165919

[7] B.J. Millis, P.G. Cottell Jr, Cooperative Learning for Higher Education Faculty. Series on Higher Education., ERIC, 1997.

[8] P.D. Feden, R.M. Vogel, Methods of teaching: Applying cognitive science to promote student learning, McGraw-Hill Humanities, Social Sciences & World Languages, 2003.

[9] N.H. Narayanan, C. Hundhausen, D. Hendrix, M. Crosby, "Transforming the CS classroom with studio-based learning," in Proceedings of the 43rd ACM technical symposium on Computer Science Education, 165–166, 2012. https://doi.org/10.1145/2157136.2157188

[10] N. Dehbozorgi, "Active learning design patterns for CS education," in ICER 2017 - Proceedings of the 2017 ACM Conference on International Computing Education Research, 2017, doi:10.1145/3105726.3105741.

[11] N. Dehbozorgi, S. MacNeil, M. Lou Maher, M. Dorodchi, "A comparison of lecture-based and active learning design patterns in CS education," in 2018 IEEE Frontiers in Education Conference (FIE), IEEE: 1–8, 2018. https://doi.org/10.1109/FIE.2018.8659339

[12] L. Liu, J. Hao, A.A. von Davier, P. Kyllonen, J.-D. Zapata-Rivera, A tough nut to crack: Measuring collaborative problem solving, IGI Global: 344–359, 2016. https://doi.org/10.4018/978-1-4666-9441-5.ch013

[13] E. Salas, D.L. Reyes, A.L. Woods, The assessment of team performance: Observations and needs, Springer: 21–36, 2017. https://doi.org/10.1007/978-3-319-33261-1_2

[14] D.R. Ilgen, J.R. Hollenbeck, M. Johnson, D. Jundt, "Teams in organizations: From input-process-output models to IMOI models," Annu. Rev. Psychol., 56,517–543,2005.https://doi.org/10.1146/annurev.psych.56.091103.070250

[15] S.W.J. Kozlowski, B.S. Bell, "Work groups and teams in organizations.," Handbook of psychology: Industrial and organizational psychology, London, England: Wiley, 12, 333–375, 2013. https://doi.org/10.1002/9781118133880.hop212017

[16] S.W.J. Kozlowski, D.R. Ilgen, "Enhancing the effectiveness of work groups and teams," Psychological Science in the Public Interest, 7(3), 77–124, 2006. https://doi.org/10.1111/j.1529-1006.2006.00030.x

[17] J. Mathieu, M.T. Maynard, T. Rapp, L. Gilson, "Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future," Journal of Management, 34(3), 410–476, 2008. https://doi.org/10.1177/0149206308316061

[18] L.A. DeChurch, J.R. Mesmer-Magnus, "The cognitive underpinnings of effective teamwork: a meta-analysis.," Journal of Applied Psychology, 95(1), 32, 2010. https://doi.org/10.1037/a0017328

[19] S. Anders, J. Heinzle, N. Weiskopf, T. Ethofer, J.-D. Haynes, "Flow of affective information between communicating brains," Neuroimage, 54(1), 439–446, 2011. https://doi.org/10.1016/j.neuroimage.2010.07.004

[20] M.B. Schippers, A. Roebroeck, R. Renken, L. Nanetti, C. Keysers, "Mapping the information flow from one brain to another during gestural communication," Proceedings of the National Academy of Sciences, 107(20), 9388–9393, 2010. https://doi.org/10.1073/pnas.1001791107

[21] K. Shockley, M.-V. Santana, C.A. Fowler, "Mutual interpersonal postural constraints are involved in cooperative conversation.," Journal of Experimental Psychology: Human Perception and Performance, 29(2), 326, 2003. https://doi.org/10.1037/0096-1523.29.2.326

[22] R. Stevens, T. Galloway, J. Lamb, R. Steed, C. Lamb, Linking team neurodynamic organizations with observational ratings of team performance, Springer: 315–330, 2017. https://doi.org/10.1007/978-3-319-33261-1_20

[23] E. Hudlicka, "To feel or not to feel: The role of affect in human–computer interaction," International Journal of Human-Computer Studies, 59(1–2), 1–32, 2003. https://doi.org/10.1016/S1071-5819(03)00047-8

[24] T.L. Dickinson, R.M. McIntyre, A conceptual framework for teamwork measurement, Psychology Press: 31–56, 1997.

[25] D.E. Wilkins, P.B. Lawhead, "Evaluating individuals in team projects," ACM SIGCSE Bulletin, 32(1), 172–175, 2000. https://doi.org/10.1145/331795.331849

[26] L.K. Michaelsen, M. Sweet, "The essential elements of team-based learning," New Directions for Teaching and Learning, 2008(116), 7–27, 2008. https://doi.org/10.1002/tl.330

[27] R. Stevens, T. Galloway, J. Lamb, R. Steed, C. Lamb, Linking team neurodynamic organizations with observational ratings of team performance, Springer: 315–330, 2017. https://doi.org/10.1007/978-3-319-33261-1_20

[28] E. Salas, C.S. Burke, J.E. Fowlkes, "Measuring team performance in the wild: Challenges and tips", Psychology Press: 263–290, 2014.

[29] J.E. Driskell, E. Salas, S. Hughes, "Collective orientation and team performance: Development of an individual differences measure," Human Factors, 52(2), 316–328, 2010. https://doi.org/10.1177/0018720809359522

[30] P.A. Schutz, R. Pekrun, G.D. Phye, "Emotion in education", Elsevier, 2007.

[31] D.B. McLeod, "Research on affect in mathematics education: A reconceptualization," Handbook of Research on Mathematics Teaching and Learning, 1, 575–596, 1992.

[32] A.C. Araújo-Simões, S.M. Guedes-Gondim, "Performance and affects in group problem-solving," Revista de Psicología Del Trabajo y de Las Organizaciones, 32(1), 47–54, 2016. https://doi.org/10.1016/j.rpto.2015.09.001

[33] M. Munezero, C.S. Montero, M. Mozgovoy, E. Sutinen, "Exploiting sentiment analysis to track emotions in students' learning diaries," in Proceedings of the 13th Koli Calling International Conference on Computing Education Research, 145–152, 2013.

https://doi.org/10.1145/2526968.2526984

[34] P. Maras, P. Kutnick, "Emotional and behavioural difficulties in schools: Consideration of relationships between theory and practice," Social Psychology of Education, **3**(3), 135–153, 1999. https://doi.org/10.1023/A:1009618701899

[35] M. Arguedas, T. Daradoumis, F. Xhafa, "Towards an emotion labeling model to detect emotions in educational discourse," in 2014 Eighth International Conference on Complex, Intelligent and Software Intensive Systems, IEEE: 72–78, 2014. https://doi.org/10.1109/CISIS.2014.11

[36] D.T. Mizokawa, N. Hansen-Krening, "The ABCs of attitudes toward reading: Inquiring about the reader's response," Journal of Adolescent & Adult Literacy, **44**(1), 72–79, 2000.

[37] S. Anders, J. Heinzle, N. Weiskopf, T. Ethofer, J.-D. Haynes, "Flow of affective information between communicating brains," Neuroimage, **54**(1), 439–446, 2011. https://doi.org/10.1016/j.neuroimage.2010.07.004

[38] M.B. Schippers, A. Roebroeck, R. Renken, L. Nanetti, C. Keysers, "Mapping the information flow from one brain to another during gestural communication," Proceedings of the National Academy of Sciences, **107**(20), 9388–9393, 2010.

[39] K. Shockley, M.-V. Santana, C.A. Fowler, "Mutual interpersonal postural constraints are involved in cooperative conversation.," Journal of Experimental Psychology: Human Perception and Performance, **29**(2), 326, 2003. https://doi.org/10.1037/0096-1523.29.2.326

[40] R. Pekrun, T. Goetz, A.C. Frenzel, P. Barchfeld, R.P. Perry, "Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ)," Contemporary Educational Psychology, **36**(1), 36–48, 2011. https://doi.org/10.1016/j.cedpsych.2010.10.002

[41] H. Tarmazdi, R. Vivian, C. Szabo, K. Falkner, N. Falkner, "Using learning analytics to visualise computer science teamwork," in Proceedings of the 2015 ACM Conference on Innovation and technology in computer science education, 165–170, 2015. https://doi.org/10.1145/2729094.2742613

[42] T. Ahmed, A. Bosu, A. Iqbal, S. Rahimi, "SentiCR: a customized sentiment analysis tool for code review interactions," in 2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE), IEEE: 106–111, 2017. https://doi.org/10.1109/ASE.2017.8115623

[43] K. Hood, P.K. Kuiper, "Improving student surveys with natural language processing," in 2018 Second IEEE International Conference on Robotic Computing (IRC), IEEE: 383–386, 2018. https://doi.org/10.1109/IRC.2018.00079

[44] C. Hutto, E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in Proceedings of the International AAAI Conference on Web and Social Media, 2014.

[45] A. Amin, I. Hossain, A. Akther, K.M. Alam, "Bengali vader: A sentiment analysis approach using modified vader," in 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE: 1–6, 2019. https://doi.org/10.1109/ECACE.2019.8679144

[46] V. Bonta, N.K.N. Janardhan, "A comprehensive study on lexicon based approaches for sentiment analysis," Asian Journal of Computer Science and Technology, **8**(S2), 1–6, 2019. https://doi.org/10.51983/ajcst-2019.8.S2.2037

[47] P. Bholowalia, A. Kumar, "EBK-means: A clustering technique based on elbow method and k-means in WSN," International Journal of Computer Applications, **105**(9), 2014.

[48] Y.R. Tausczik, J.W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," Journal of Language and Social Psychology, **29**(1), 24–54, 2010. https://doi.org/10.1177/0261927X09351676

[49] A.S. Hess, J.R. Hess, "Principal component analysis," Transfusion, **58**(7), 1580–1582, 2018. https://doi.org/10.1111/trf.14639

[50] J. Hauke, T. Kossowski, "Comparison of values of Pearson's and Spearman's correlation coefficient on the same sets of data," 2011. https://doi.org/10.2478/v10117-011-0021-1

[51] M.M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," Malawi Medical Journal, **24**(3), 69–71, 2012.

[52] https://scikit-learn.org/

[53] A. Hasan, S. Moin, A. Karim, S. Shamshirband, "Machine learning-based sentiment analysis for twitter accounts," Mathematical and Computational Applications, **23**(1), 11, 2018. https://doi.org/10.3390/mca23010011

[54] S. Loria, "textblob Documentation," Release 0.15, 2, 269, 2018.

ASTES

A Monthly Rainfall Forecasting from Sea Surface Temperature Spatial Pattern

Prattana Deeprasertkul[*], Royol Chitradon

Hydro - Informatics Institute (Public Organization), Bangkok, 10900, Thailand

A R T I C L E I N F O	A B S T R A C T
	The ocean surface temperatures or sea surface temperatures have a significant influence on local and global weather. The change in sea surface temperatures will lead to the change in rainfall patterns. In this paper, the long-term rainfall forecasting is developed for planning and decision making in water resource management. The similarity of sea surface temperature images pattern that was applied to analyze and develop the monthly rainfall forecasting model will be proposed. In this work, the convolutional neural network and autoencoder techniques are applied to retrieve the similar sea surface temperature images in database store. The accuracy values of the monthly rainfall forecasting model which is the long-term forecasting were evaluated as well. The average value of the model accuracies was around 82.514%.

1. Introduction

The rainfall forecast modeling systems are available to monitor and warn flood and drought events in Thailand. There are two fundamental approaches which are dynamical and empirical methods for rainfall forecasting. The dynamical prediction is generated by physical model based on numerical analysis that represents the evolution in seasonal forecasting by using observed Sea Surface Temperature anomalies as a boundary condition. The empirical approach relies on the observed and measured data. It is used to analyse the atmospheric and oceanic variables in the real world [1].

Tropical rainfall is influenced strongly by nearby sea temperatures in the real world [1]. Recently, there are some climate predictions used empirical relations between Sea Surface Temperature or SST and atmospheric variables [2]. The empirical approaches is the most widely used to predict the climate using linear regression, machine learning, artificial neural network, and the methods of data-driven handling [3, 4].

In this paper, the monthly rainfall forecasting model is implemented by using the convolutional autoencoder technique for efficient unsupervised encoding of image features. We used 7 years (2014-2020) datasets of the global climate data such as SST and the Global Satellite Mapping of Precipitation project or GSMaP datasets that both are the satellite images. This model has predicted the monthly rainfall amount in image format. This is the

improvement of the data-driven analytics which is a project of Hydro – Informatics Institute (Public Organization) [5].

The monthly rainfall forecasting amount with 6 months ahead were predicted. The experimental results show the accuracy that was calculate between the prediction rainfall amount and the actual rainfall amount.

The remaining of this paper is organized as follows: Section 2 describes the background knowledge related to this work. Section 3 explains the analysis of SST matching. In section 4, the experimental results of monthly rainfall forecasting model are presented. Finally, this work is concluded in section 5.

2. Background

The datasets and image processing technologies using in this monthly rainfall forecasting model are described in this section.

2.1. Data

The set of data using in this work consist of two datasets that were SST as displayed in Figure 1 and the satellite precipitation or GSMaP as displayed in Figure 2. Firstly, SST is the ocean temperature that closes to the ocean's surface. The sea surface is between 1 millimetre and 20 meters below the sea surface [6]. The weather satellite sensors are available to measure and define sea surface temperatures. Secondly, the GSMaP project [6,7,8,9] is a high resolution global precipitation map production applying satellite data. This project is supported by Core Research for Evolutional Science and Technology (CREST) of the Japan

[*]Corresponding Author: Prattana Deeprasertkul, prattana@hii.or.th

Science and Technology Agency (JST) during 2002-2007. The GSMaP project activities are promoted by the JAXA Precipitation Measuring Mission (PMM) Science Team since 2007.
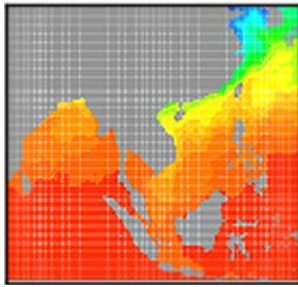
Figure 1: The example of SST data.

Figure 2: The example of GSMaP data.

2.2. Type and Format of Data

Because SST and GSMaP are the satellite images in the digital formats, to display both on Map and to have a good performance, both are converted from images (JPEG) format to text (ESRI ASCII raster) format. The ESRI ASCII raster file format has the basic structure as illustrated in Table 1 [10]. The header fields at the top of the file displayed in "Parameter" column in Table 1 [10]. The example of ESRI ASCII raster file is shown in Figure 3.

Table 1: The description of the header fields of ESRI ASCII raster file.

Parameter	Description	Requirements
NCOLS	Number of cells in a column.	Integer
NROWS	Number of cells in a row.	Integer
XLLCENTER	Coordinate X on the origin point (at lower left corner of the cell).	Match with Y coordinate type.
YLLCENTER	Coordinate Y on the origin point (at lower left corner of the cell).	Match with X coordinate type.
CELLSIZE	Cell size.	Decimal
NODATA_VALUE	The value means NoData in the raster file.	Optional

```
ncols 480
nrows 450
xllcorner 378923
yllcorner 4072345
cellsize 30
nodata_value -32768
43 2 45 7 3 56 2 5 23 65 34 6 32 54 57 34 2 2 54 6
35 45 65 34 2 6 78 4 2 6 89 3 2 7 45 23 5 8 4 1 62 ...
```

Figure 3: The example of ESRI ASCII raster file.

2.3. Convolutional Autoencoder

A Convolutional Autoencoder or ConvAE shown in Figure 4 is a type of Convolutional Neural Network (CNN) designed for unsupervised deep learning. CNN is widely used in the field of pattern recognition. CNN architecture is constructed from various layers in the neural network. The convoluted layers are used for automatic extraction of an image feature hierarchy. The size of the input layer is the same as output layer. An autoencoder consists of image encoder and image decoder. The training of the autoencoder encodes image features into compressed images. The encoded image features are then clustered using k-means clustering. a small sub volume set is characterized by the decoded cluster center [11]. This technique is applied to reduce the storage usage, speeding up the computation time, improving performance by removing redundant variables [12].
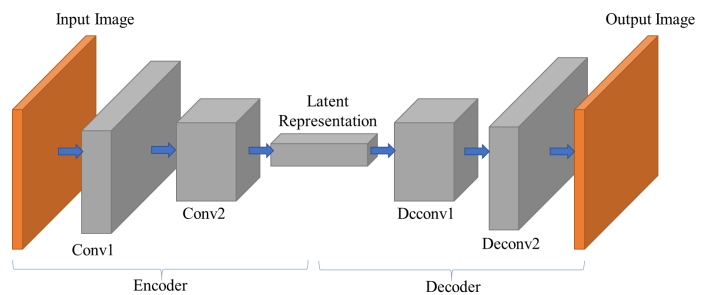
Figure 4: Convolutional Autoencoder.

- An encoder as shown in Figure 4 is an algorithm used to transfer one format to another format. In this work, the encoder is applied to the original image. Then, the original image is represented in smaller feature vectors image or it is compressed [13].
- The decoder as shown in Figure 4 is an algorithm used for image decompression. The important features of the original image are not lost after the decoding process is applied in the encoded image. The compressed image is returned to its original image [13].

2.4. Image Retrieval

To retrieve the similar images stored in the images database and to get the top-k most similar database images, K-Nearest Neighbours or KNN method with cosine similarity as the distance metric. KNN is applied to verify the nearest neighbours and to get k nearest points as the similar images from a query image.

The Content Based Image Retrieval method was applied in this work as the process flow shown in Figure 5. In training process, the encoded representation of all images from training database is calculated using ConvAE and saved in a file. In testing process, the encoded representation of query image from testing database is calculated using ConvAE and saved in a file. The distance between the encoded representations of query image to encoded representation of all the images in training database is calculated to find nearest neighbors. Finally, KNN retrieves the

top-k images in descending order of distance to the query image [14].
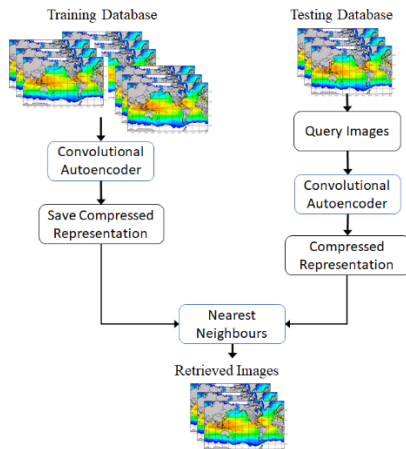
Figure 5: The Content Based Image Retrieval process flow.

2.5. Image Matching

Image matching is method to evaluate the maximum similarity values between two digital images. The method is divided into two categories: 1) gray-based matching 2) feature-based matching. The gray-based matching is a matching algorithm of the grayscale feature applied in the image pixels. The feature-based matching is a matching algorithm of the feature vector points [15]. Image similarity is the similarity measurement between two images with the metrics. Therefore, some similarity criteria rules are used to verify the corresponding points of two images. The similarity quality is evaluated by the distance. Recently, there are many metrics of the distance measurement such as Euclidean distance, humming distance, cosine distance, and etc. the Hamming distance is a metric for measuring the distance between two binary strings.

2.6. Image Hashing

Image hashing has various methods which can be the mapping of the similar images with similar hashing codes. The well-known image hashing methods are the followings [16]:

- Average image hashing is the simplest hashing method. The images are converted to the greyscales which are transformed to binary codes with the average values.
- Perceptual image hashing extracts the real image features to hash values with discrete cosine transformation. This image hashing is the popular hashing method in signal processing.
- Difference image hashing uses the relative gradient technique to calculate the difference between adjacent pixels.

Hamming distance is the metric to compare and calculate the distance between two hashing strings. From comparing between the equal length of two hashing strings, hamming distance computes the number of bits which are different [17].

3. Sea Surface Temperature Matching

This section has described the convolutional autoencoder layers when the process was implemented and the image matching process for monthly rainfall forecast application.

3.1. Convolutional Autoencoder Layers

The convolutional autoencoder aims to predict monthly rainfall from the Sea Surface Temperature images. The rainfall predictor is implemented by adding the following layers to the neural network as described in Table 2 [13]. There is a procedure as the following:

- Step 1: A convoluted layer with shape of (287, 309, 16) is added and collected separately in the output layer 1. It is extracted significant features from the original image before reducing the shape in smaller features.
- Step 2: A maximum layer selected and pooled with the shape of (144, 155, 16), is associated to the first step convoluted layer and it resizes the layer to smaller features.
- Step 3: A dropout layer, with the shape of (144, 155, 8), is regularization technique to avoid overfitting of neural network.
- Step 4: A maximum layer pooled with the shape of (72, 78, 8), is associated to the previous step convoluted layer.
- Step 5: The previous maximum layer is pooled to the shape of (72, 78, 8) before a convoluted layer is applied.
- Step 6: A maximum layer selected and pooled with the shape of (36, 39, 8), is associated to the previous step convoluted layer.
- Step 7: The previous maximum layer is pooled to the shape of (36, 39, 8) before a convoluted layer is applied.
- Step 8: An upsampled layer with the shape of (72, 78, 8), is applied to increase the image resolution to the larger image shape.
- Step 9: The previous convoluted layer is performed by the additional operation on the output layer 2 with the shape of (72, 78, 8).
- Step 10: An upsampled layer with the shape of (144, 156, 8), is applied to increase the image resolution to the larger image shape.
- Step 11: The previous convoluted layer is performed by the additional operation on the output layer 1 with the shape of (142, 154, 16).
- Step 12: An upsampled layer with the shape of (284, 308, 16), is applied to increase the image resolution to the larger image shape.
- Step 13: The final layer of neural networks is applied to the convoluted layer with the shape of (284, 308, 4) to receive the output.

Table 2: The layers of Autoencoder method [13].

Layer (type)	Output Shape
Input Layer	(287, 309, 4)
Conv	(287, 309, 16)
Maximum Pooling	(144, 155, 16)
Conv	(144, 155, 8)

Maximum Pooling	(72, 78, 8)
Conv	(72, 78, 8)
Maximum Pooling	(36, 39, 8)
Conv	(36, 39, 8)
Upsampling	(72, 78, 8)
Conv	(72, 78, 8)
Upsampling	(144, 156, 8)
Conv	(142, 154, 16)
Upsampling	(284, 308, 16)
Conv	(284, 308, 4)

3.2. Image Retrieval Process

This application is the SST pattern data analysis for 6 months rainfall forecasting. In this work, the Content Based Images Retrieval technique to estimate the similarity values was applied in monthly rainfall forecasting of Thailand. Figure 6 illustrates the SST image matching process. The current SST has been detected and matched with the historical SST dataset stored in the database. In this project, MongoDB database application is used to collect the images which are the unstructured data. The program code is developed to retrieve the similar database images with comparing to the current SST images. The top three highest values of image similarities are the results of monthly rainfall forecasting.
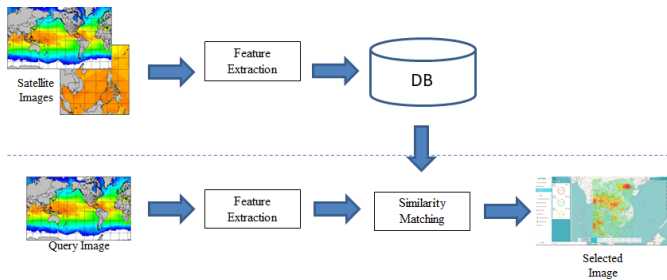
Figure 6: The image matching process flow [5].
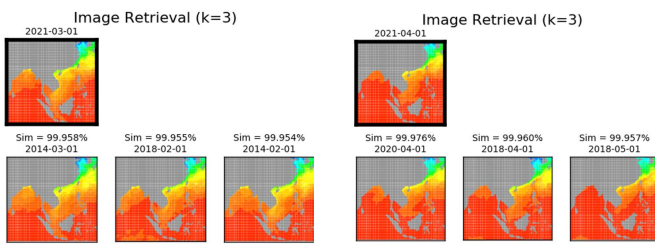
Figure 7: The results of the top three highest SST similarities.

4. Experimental Results

The program has processed and generated the results of the top three highest SST similarities written by Python. These results are shown in Figure 7. The rainfall amounts using in this work are the accumulated rainfall or the accumulated GSMaP rainfall at the same months of top three highest SST matching. These months results are used to define the predicted months of rainfall data and they are shown on the application screen in Figure 8. In addition, the SST and GSMaP images can be simultaneously overlaid on application screen as illustrated in Figure 9.

The accurate measurement of this monthly rainfall forecasting analysis is described the detail of process in this

section as well. The process flow is displayed in Figure 10. For this measurement, the image hashing method was applied to retrieve the GSMaP image similarities. The hashing codes are used to be the index to discover the most similar database images. The image hashing codes are used to compare the images which are easier and faster method than using the whole of original images [18].
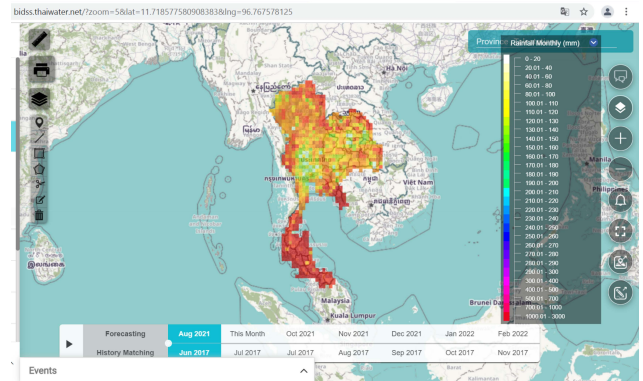
Figure 8: The example of monthly GSMaP data overlaid on map.
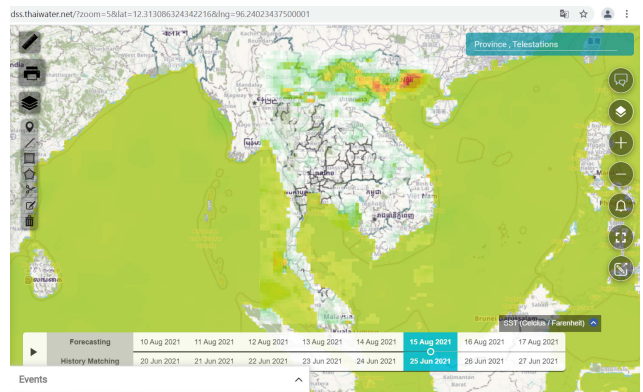
Figure 9: SST and GSMaP data were overlaid simultaneously on the application screen.
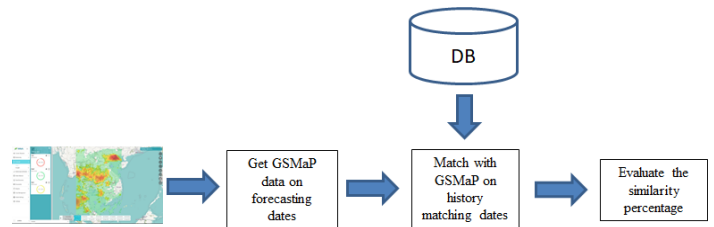
Figure 10: The process flow of the accurate measurement of rainfall forecasting model [5].

The program code of accuracy evaluation of this monthly rainfall forecasting model by using image pattern similarity was written by Python Jupyter as shown in Figure 11. The current monthly GSMaP rainfall datasets were matched with the historical GSMaP datasets collected in MongoDB. Both of the current and the historical GSMaP hashing codes were created and then the similarity values between two images were computed with their hashing method. The example between the nth months

GSMaP rainfall forecasting and the nth months of the historical matching GSMaP as shown in Table 3 and Table 4.


```
In [8]:  d = "20161201"
         s = "20201201"

In [9]:  check_similar_gsm(d, s, imagehash.phash, hashsize=16, isPlot=True, last=7)

Out[9]:  [('20170501', '20210501', 79.12),
          ('20170601', '20210601', 79.78),
          ('20161201', '20201201', 79.88),
          ('20170401', '20210401', 79.91),
          ('20170301', '20210301', 80.06),
          ('20170201', '20210201', 80.9),
          ('20170101', '20210101', 85.94)]
```

Figure 11: The example of Python Jupyter displaying for GSMaP comparison.

Table 3: The example I of the rainfall forecasting date and the historical matching date of GSMaP data.

Forecasting date	History matching date
Dec. 2020	Dec. 2016
Jan. 2021	Jan. 2017
Feb. 2021	Feb. 2017
Mar. 2021	Mar. 2017
Apr. 2021	Apr. 2017
May 2021	May 2017
Jun. 2021	Jun. 2017

Table 4: The example II of the rainfall forecasting date and the historical matching date of GSMaP data.

Forecasting date	History matching date
Nov. 2020	Nov. 2016
Dec. 2020	Dec. 2016
Jan. 2021	Jan. 2017
Feb. 2021	Feb. 2017
Mar. 2021	Mar. 2017
Apr. 2021	Apr. 2017
May 2021	May 2017

The GSMaP rainfall images on the forecasting months were computed the similarity values with the historical GSMaP matching months. The example image of GSMaP similarity (%) case is shown in Figure 12. In addition, the example of the similarity values of monthly GSMaP rainfall forecasting and their historical matching were calculated as shown in Table 5. Table 5 is the example 1 which shows the GSMaP data matching between forecasting months, 2020/12 – 2021/06, and history matching months, 2016/12 – 2017/06. The similarity values of example 1 are 81.21%, 93.18%, 84.14%, 83.75%, 82.3%, 79.53%, and

80.52%, respectively and the average of similarity values is 83.52%. For the example 2 in Table 6, the forecasting months are 2020/11 – 2021/05 that match with the history matching months, 2016/11 – 2017/05 and the similarity values are 81.46%, 81.21%, 93.18%, 84.14%, 83.75%, 82.3%, and 79.53%, respectively. The average of similarity values in the example 2 is 83.65%.
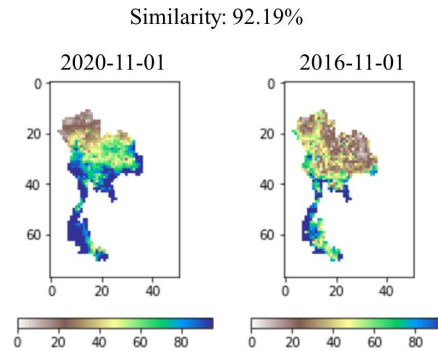
Figure 12: The example images of GSMaP similarity (%)

Table 5: The example I of the average of monthly GSMaP forecasting and the historical matching similarity

Forecasting months	History matching months	Similarity (%)
2020/12	2016/12	81.21
2021/01	2017/01	93.18
2021/02	2017/02	84.14
2021/03	2017/03	83.75
2021/04	2017/04	82.3
2021/05	2017/05	79.53
2021/06	2017/06	80.52
	Avg.	83.52

Table 6: The example II of the average of monthly GSMaP forecasting and the historical matching similarity

Forecasting months	History matching months	Similarity (%)
2020/11	2016/11	81.46
2020/12	2016/12	81.21
2021/01	2017/01	93.18
2021/02	2017/02	84.14
2021/03	2017/03	83.75
2021/04	2017/04	82.3
2021/05	2017/05	79.53
	Avg.	83.65
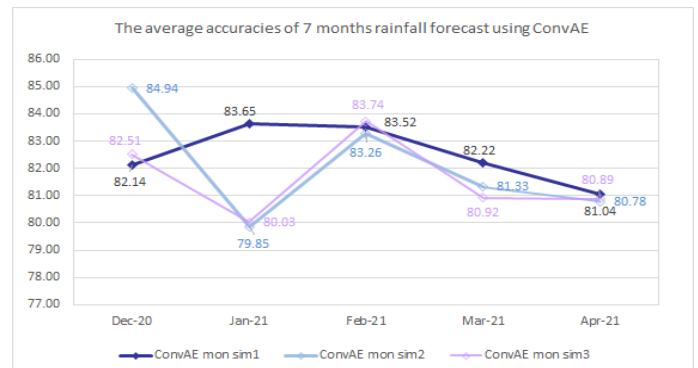
Figure 13: The graph of the monthly rainfall forecast similarity since December 2020 to April 2021

Figure 13 is the comparison between the accurate values of top three highest similarities or top three rainfall forecasting. The graph shows the accurate values of 5 months GSMaP rainfall forecast since December 2020, January 2021, February 2021, March 2021, to April 2021. The accurate values of the highest similarity (dark blue line) are 82.14%, 83.65%, 83.52%, 82.22%, and 81.04%, respectively. The average of the first accuracy is 82.51%. The average of the second and the third accurate values are 82.03% and 81.62%, respectively. Therefore, it can be found that the highest GSMaP rainfall forecast is better than others.

5. Conclusion

The monthly rainfall forecasting was the SST image patterns analysis by computing the average similarity between the recent monthly SST image datasets and historical monthly SST image datasets and applying convolutional autoencoder method to retrieve the top three highest of SST similarities. Moreover, the accurate value of this monthly rainfall forecasting was evaluated by applying the GSMaP image matching method as well. Then the result of this monthly rainfall forecasting was 82.514% which was an average value in matching GSMaP testing datasets.

However, this monthly rainfall forecasting model will be improved the accuracy, in the near future, by considering more atmospheric variables related to rainfall in Thailand. In addition, the weekly rainfall forecasting will be added in the application.

References

[1] M. Kannan et al., "Rainfall Forecasting Using Data Mining Technique," International Journal of Engineering and Technology, **2**(6):397-401, 2010.

[2] University of Reading, "Tropical Rainfall and Sea Surface Temperature Link Could Improves Forecast," https://www.reading.ac.uk/news-and-events/releases/PR849878.aspx, 26 October 2020.

[3] M. Haghbin et al., "Applications of soft computing models for predicting sea surface temperature: a comprehensive review and assessment," Progress in Earth Planetary Science, **8:**4, 2021, doi:10.1186/s40645-020-00400-9.

[4] O. Ieremeiev, V. Lukin, K. Okarma, and K. Egiazarian, "Full-Reference Quality Metric Based on Neural Network to Assess the Visual Quality of Remote Sensing Images," Remote Sensing, **12**(15):2349, 2020, doi:10.3390/rs12152349.

[5] P. Deeprasertkul, "An Assessment of Rainfall Forecast using Image Similarity Processing," Proceedings of the 2020 3rd International Conference on Image and Graphics Processing, 141-145, 2020, doi:10.1145/3383812.3383821.

[6] K. Okamoto et al., "The Global Satellite Mapping of Precipitation (GSMaP) project," 25th IGARSS Proceedings, 3414-3416, 2005, doi: 10.1109/IGARSS.2005.1526575.

[7] K. Aonashi, J. Awaka, M. Hirose, T. Kozu, T. Kubota, G. Liu, S. Shige, S. Kida, S. Seto, N. Takahashi, and Y. N. Takayabu, "GSMaP passive, microwave precipitation retrieval algorithm: Algorithm description and validation," J. Meteor. Soc. Japan, 87A, 119-136, 2009.

[8] T. Kubota, S. Shige, H. Hashizume, K. Aonashi, N. Takahashi, S. Seto, M. Hirose, Y. N. Takayabu, K. Nakagawa, K. Iwanami, T. Ushio, M. Kachi, and K. Okamoto, "Global Precipitation Map using Satelliteborne Microwave Radiometers by the GSMaP Project : Production and Validation," IEEE MicroRad, 290-295, 2006, doi: 10.1109/MICRAD.2006.1677106.

[9] T. Ushio, T. Kubota, S. Shige, K. Okamoto, K. Aonashi, T. Inoue, N. Takahashi, T. Iguchi, M. Kachi, R. Oki, T. Morimoto, and Z. Kawasaki, "A Kalman filter approach to the Global Satellite Mapping of Precipitation (GSMaP) from combined passive microwave and infrared radiometric data," J. Meteor. Soc. Japan, 87A, 137-151, 2009, doi:10.2151/jmsj.87A.137.

[10] ESRI Inc., "ESRI ASCII raster format," http://resources.esri.com/help/9.3/arcgisengine/java/GPToolRef/spatial_analyst_tools/esri_ascii_raster_format.htm.

[11] X. Zeng, M. Ricardo Leung, T. Zeev-Ben-Mordehai, and M. Xu, "A convolutional autoencoder approach for mining features in cellular electron cryo-tomograms and weakly supervised coarse segmentation," Journal of Structural Biology, **202**(2):150-160, 2018, doi:10.1016/j.jsb.2017.12.015.

[12] Y. Zhang, "A Better Autoencoder for Image: Convolutional Autoencoder." In ICONIP17-DCEC, 2018.

[13] P. Kumar and N. Goel, "Image Resolution Enhancement Using Convolutional Autoencoders." 7th Electronic Conference on Sensors and Applications Proceedings, 8259, 2020, doi:10.3390/ecsa-7-08259.

[14] P. Deore, V. S. Patil, and V. B. Patil, "Content Based Image Retrieval using Deep Convolutional Variational Autoencoder," Proteus Journal, **11** (10), 2020, ISSN/eISSN: 0889-6348.

[15] G. Jianhua, Y. Fan, T. Hai, W. Jing-xue, and L. Zhiheng, "Image matching using structural similarity and geometric constraint approaches on remote sensing images," Journal of Applied Remote Sensing, **10**(4), 2016, doi:10.1117/1.JRS.10.045007.

[16] T. Zhenjun, W. Shuozhong, Z. Xinpeng, and W. Weimin, "Structural Feature-Based Image Hashing and Similarity Metric for Tampering Detection," Fundamenta Informticae, **106**(1):75-91, 2011, doi:10.3233/FI-2011-377.

[17] N. Raut, "What is Hamming Distance?," TutorialPoint, 2018.

[18] Study.com, "Discuss the advantages of hashing," 2019.

A Summary of Canonical Multivariate Permutation Entropies on Multivariate Fractional Brownian Motion

Marisa Mohr[*,1,2], Ralf Möller[1,3]

[1]*University of Lübeck, Institute of Information Systems, Lübeck, 23562, Germany*

[2]*inovex GmbH, Data Management & Analytics, Hamburg, 22763, Germany*

[3]*German Research Center for Artificial Intelligence, Stochastic Relational AI in Healthcare, Lübeck, 23562, Germany*

A R T I C L E   I N F O

A B S T R A C T

Real-world applications modelled by time-dependent dynamical systems with specific properties such as long-range dependence or self-similarity are usually described by fractional Brownian motion. The investigation of the qualitative behaviour of its realisations is an important topic. For this purpose, efficient mappings from realisations of the dynamical system, i.e., time series, to a set of scalar-valued representations that capture certain properties are considered. Permutation entropy is a well-known measure to quantify the complexity of univariate time series in a scalar-valued representation, for example, to derive estimates for self-similarity or as features or representations in learning tasks. However, since many real-world problems involve multivariate time series, permutation entropy needs to be extended to the multivariate case. This work summarises the behaviour of pooled permutation entropy (PPE), multivariate multi-scale permutation entropy (MMSPE), and multivariate weighted permutation entropy (MWPE) on multivariate fractional Brownian motion, and this work fills the gaps in existing research. In addition, we provide a new study of multivariate ordinal pattern permutation entropy (MOPPE) on multivariate fractional Brownian motion. We conclude with a detailed experimental evaluation and comparison between all multivariate extensions, for example, demonstrating identical behaviour of PPE and MMSPE or uncovering different aspects such as amplitude and cross-correlations by using MWPE and MOPPE, respectively.

1   Introduction

This paper is an extension of two works originally presented in the *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2020 (APSIPA ASC)* [1] and in the *International Florida Artificial Intelligence Society Conference Proceedings (FLAIRS-34)* [2]. Both papers study the qualitative behaviour of multivariate fractional Brownian motion by analysing the distribution of up and down encodings in terms of ordinal patterns and permutation entropy.

Time series data is part of many real-world applications, e.g., weather records, patient health evolution metrics, (industrial) sensor data, stock prices, website clicks, server metrics, network data, political or sociological indicators, video or voice messages – just to name a few. Time series can be understood as a series of (real-valued) random variables of a time-dependent system that changes over time, also called a dynamical system. For many real-world applications, modelling time-dependent dynamical systems require

specific properties such as *long-range dependence* or *self-similarity*. For example, the long memory is fundamental in internet traffic [3] or for financial data [4]. If financial return series exhibit a long-range dependence, this indicates that observed returns over time are not independent. When returns are not independent, past returns may help predict future returns [5]. This work focuses on a special class of dynamical systems or stochastic processes that describe long-range dependence and self-similarity properties, called fractional Brownian motion (fBm) and denoted as $B_H(t)$, where the so-called Hurst parameter H describes the self-similarity.

While classical time series analysis examines the values of the time series themselves, symbolic time series analysis considers a non-parametric mapping or encoding into a sequence of symbols. This is particularly promising in the case of dynamical systems such as fBm, as the overall dynamics of the generating system is uncovered [6]. To measure the qualitative behaviour of a generating system, efficient measures of the sequence of symbols to a set of scalar-valued representations, also called features, that capture

*Corresponding Author: Marisa Mohr, mohr@ifis.uni-luebeck.de

specific properties are considered. Entropy, a measure that has its origins in information theory, is promising through an encoding that preserves the information content of an object [7]. PE is a robust, scalar-valued measure for determining the degree of complexity of dynamical systems or observed time series by analysing the distribution of *symbolic sequences of ordinal patterns* [8]. While the associated entropy is low for a deterministic time series, it approaches its maximum value in case of uncorrelated randomness or high complexity. To distinguish between randomness and complexity, an additional measure is necessary. As a complement to permutation entropy (PE), multi-scale permutation entropy (MSPE) captures the complexity of time series on different time scales [9]. On higher time scales, random noise tends to cancel out, resulting in a low entropy value, where complex signals retain a high entropy value. In this manner, it is possible to gain a deeper insight into the randomness and complexity of a system. Another limitation of PE is the inability to distinguish between different patterns of a given motif in amplitudes. In [10], the author introduce weighted permutation entropy (WPE), which takes into account patterns that differ in amplitudes by assigning weights to each extracted pattern.

The behaviour of PE on fBm in the univariate case is well understood [8, 11, 12]. Not only the PE but also the distributions of ordinal patterns of certain lengths yield interesting properties. The distributions of ordinal patterns of lengths (also called order) two and three yield parameter functions that can be used for descriptive purposes such as autocorrelation. The distribution of order four has irrational values and depends on the noise distribution. The resulting closed formulas for computation are used in this work. The distributions of ordinal patterns of lengths greater than four do not yield closed formulas. Thus, a relationship between autocorrelation or spectrum on the one hand and ordinal patterns, on the other hand, does no longer exist [8]. In [13], the author investigate the behaviour of MSPE on fBm, denoted as $B_H(t)$, and show that MSPE of the fractional Gaussian noise, the increment process of fBm, i.e., $X(t) = B_H(t + 1) - B_H(t)$, is time-scale invariant. In [2] we investigate the behaviour of WPE on fBm and show that the distribution of ordinal patterns of order $d = 2$ and thus also WPE are independent of the weighting.

Moreover, in many fields of applications, multivariate measurements are performed. Examples of multivariate fractional Brownian motion (mfBm) can be found in economic time series [14], or functional Magnetic Resonance Imaging of several brain regions [15]. This work summarises the behaviour of several multivariate extensions of PE on mfBm. The first original conference paper [1] studies the behaviour of two canonical multivariate extensions of PE, known as pooled permutation entropy (PPE) and multivariate multiscale permutation entropy (MMSPE) of orders $d = 2$ and $d = 3$, on mfBm. The second original conference paper [2] investigates the behaviour of multivariate weighted permutation entropy (MWPE) of orders $d = 2$ and $d = 3$ on mfBm. We extend original works [1] and [2] by

- a **comprehensive review** of all existing extensions to multivariate permutation entropy including code for computing on Github[1] and Python Package Index (PyPI)[2],

- a **complement** in existing theoretical and experimental investigations of PPE, MMSPE and MWPE on mfBm in a variation of its Hurst parameter H, i.e., we fill the gaps by investigating the behaviour of additional orders $d = 4$ and $d = 5$ as well as other parameters such as in particular different time delays, time scales and number of variables of mfBms,

- a **new study** on the behaviour of multivariate ordinal pattern permutation entropy (MOPPE) on mfBm, and

- a **comparison** of all considered multivariate extensions.

Section 3	Review on Multivariate Permutation Entropy
Section 4 Section 5 Section 6	**Complements on** • PPE on mfBm [1] • MMSPE on mfBm [1] • MWPE on mfBm [2]
Section 7	**New study of** • MOPPE on mfBm
Section 8	Comparison
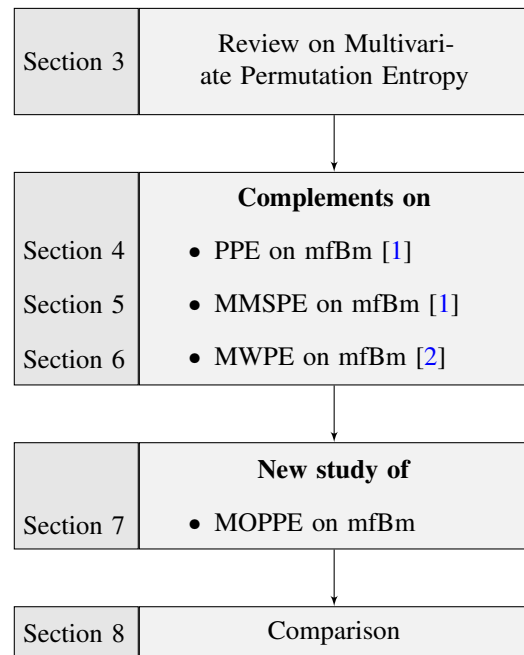
Figure 1: Structure of this research.

This paper is structured as follows and as visualised in Figure 1. After presenting background on (m)fBm and univariate PE in Section 2, needed for the paper to be self-consistent, we provide in Section 3 a review of multivariate extensions of PE and categorise them according to their approach. Since this paper focuses on canonical extensions, we present all those extensions in detail and provide algorithms for their computations. In Section 4, Section 5, and Section 6 we investigate the behaviour of PPE, MMSPE and MWPE, respectively, on mfBm in variation of the Hurst parameter H, and we complement the original work with previously mentioned investigations. In Section 7 we present a new study on the behaviour of MOPPE on mfBm in variation to the Hurst parameter H. In Section 8 we compare our findings and show that PPE and MMSPE behave identically for all investigated orders, delays, scales, and the number of variables. In addition, MWPE and MOPPE of orders $d > 2$ uncover different aspects of mfBm, i.e., amplitudes and cross-correlations, respectively. In Section 9 we conclude our work with mentioning future work.

2 Background

We shortly formalise fBm and subsequently key concepts of ordinal patterns and PE. We then summarise known results from the application of PE on fBm in the univariate case.

2.1 Multivariate Fractional Brownian Motion

A stochastic process, or more generally a mathematical object that is similar to itself at all scales, is called a fractal. When you zoom in on a fractal, it resembles or looks exactly like the original shape. The mathematical property is called self-similarity and is expressed by the so-called Hurst parameter $H \in (0, 1)$ in honour of Harold Edwin Hurst [16]. The work described here focuses on a class of special stochastic processes, namely *fractional Brownian motion* (fBm), involving the properties of self-similarity and long-range dependency.

Definition 2.1 (Fractional Brownian Motion) A stochastic process is called *fractional Brownian motion* (fBm) with Hurst parameter $H \in (0, 1)$ and denoted as $B_H(t)$ if it is

1. Gaussian,

2. self-similar, i.e., there exists $H \in (0, 1)$ such that for any $a > 0$ it is

$$B_H(at) \sim a^H B_H(t), \qquad (1)$$

   where $\sim$ denotes the equality of probability distributions, and it has

3. stationary increments, i.e., $B_H(t) - B_H(s) \sim B_H(t - s)$.

In case $H = 1/2$, fBm corresponds to the ordinary Brownian motion. In case $H > 1/2$, the process has a persistence property and positively correlated increments, i.e., an upward jump is more likely followed by another upward jump and vice versa, and the process exhibits long-range dependence. For $H \to 1$, the process becomes smoother, less irregular and more trendy. In case $H < 1/2$, the process has negatively correlated increments and an anti-persistence property. Figure 2 shows three different realisations of fBms of length $T = 5000$ with different Hurst parameters $H \in \{0.25, 0.5, 0.75\}$ illustrating the properties mentioned above.
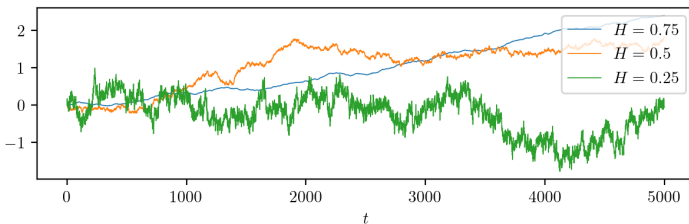
Figure 2: Three different realisations of fBm of length $T = 5000$ with different H.

Stationarity refers to the fact that the distribution of the process doesn't change in time, which has important consequences. In particular, the $B_H(t)$ are identically distributed, i.e., the expectation values and variances of components do not depend on time t. Furthermore, the distribution $B_H(t) - B_H(s)$ depends only on $t - s$, so the correlations of the components also depend only on $t - s$.

Multivariate generalisations of fBms are introduced in [17, 18] and require an extension of self-similarity to multivariate processes first. While the definition given in [19] concerns a far-reaching generalisation of fBm, where self-similarity becomes operator self-similarity for the multivariate case, in [18] the authors restrict themselves to joint self-similarity, where joint self-similarity can be seen as a special case of operator self-similarity where the operator is diagonal. For the sake of simplicity, we focus on the definition introduced in [18].

Definition 2.2 (Multivariate Self-Similar Process) A multivariate process $((X^i(t))_{i=1}^m)_{t \in \mathbb{R}} = (X^1(t), \dots, X^m(t))_{t \in \mathbb{R}}$ with variable-dimension $m \in \mathbb{N}$ is called *self-similar*, if there exists a vector $H = (H_1, \dots, H_m)$ with $H_i \in (0, 1)$ for $i = 1, \dots, m$ such that for any $a > 0$ it holds that

$$(X^1(at), \dots, X^m(at))_{t \in \mathbb{R}} \sim (a^{H_1} X^1(t), \dots, a^{H_m} X^m(t))_{t \in \mathbb{R}},$$

where $\sim$ denotes the equality of finite-dimensional distributions.

With Definition 2.2, mfBm is defined as follows.

Definition 2.3 (Multivariate Fractional Brownian Motion [20]) An m-multivariate process $((X^i(t))_{i=1}^m)_{t \in \mathbb{R}}$ is called *multivariate fractional Brownian motion (mfBm)* with self-similarity or Hurst parameter $H = (H_1, \dots, H_m)$ with $H_i \in (0, 1)$ for $i = 1, \dots, m$, and is denoted as $\mathbf{B}_H^m(t)$, if it is

1. Gaussian distributed,

2. self-similar with Hurst parameter H, and it has

3. stationary increments, i.e., $\mathbf{B}_H^m(t) - \mathbf{B}_H^m(s) \sim \mathbf{B}_H^m(t - s)$.

Multivariate self-similarity imposes some constraints on the covariance structure of mfBm [18]. The covariance structure is characterised by three parameters $\sigma_i > 0$, $\rho_{ij} \in (-1, 1)$ and $\eta_{ij} \in \mathbb{R}$ for $i, j = 1, \dots, m$, which allow two components to be more or less correlated and the process to be reversible in time or not. Parameter $\sigma_i > 0$ is the standard deviation of the i-th variable at time $t = 1$. Parameter $\rho_{ij} = \rho_{ji}$ is the correlation coefficient between the variables i and j at time $t = 1$. Parameters $\eta_{ij} = -\eta_{ji}$ are antisymmetric and linked with the time-reversibility of mfBm. Multivariate fractional Brownian motion can be characterised by its covariances and cross-covariances of its variables as follows.

Lemma 2.1 (Covariance Function of mfBm [21]) *The mfBm $\mathbf{B}_H^m(t)$ is marginally an fBm, such that the covariance function of the i-th variable $B_{H_i}^i$ of mBfm is as in the univariate case, i.e.,*

$$Cov(B_{H_i}^i(s), B_{H_i}^i(t)) = \frac{\sigma_i^2}{2}(|s|^{2H_i} + |t|^{2H_i} - |t - s|^{2H_i}). \qquad (2)$$

where $\sigma_i^2 = Var(B_{H_i}^i(1))$. The cross-covariances of mfBm for all $(i, j) \in \{1, \dots, m\}^2$ and $i \neq j$ are given by

$$Cov(B_{H_i}^i(s), B_{H_j}^j(t)) = \frac{\sigma_i \sigma_j}{2}(w_{ij}(-s) + w_{ij}(t) - w_{ij}(t - s)) \qquad (3)$$

where the function $w_{ij}(h)$ is defined by

$$w_{ij}(h) = \begin{cases} (\rho_{ij} - \eta_{ij} \operatorname{sign}(h))|h|^{H_i + H_j} & \text{if } H_i + H_j \neq 1, \\ \rho_{ij}|h| + \eta_{ij} h \log |h| & \text{if } H_i + H_j = 1. \end{cases} \qquad (4)$$

Moreover, the setting of $\rho_{ij} = 1$ and $\eta_{ij} = 0$ in (3) and (4) matches with the definition in the univariate case in (2). For $m = 1$, Definition 2.3 matches with the univariate Definition 2.1 of fBm. Figure 3 shows three realisations of mfBm of length $T = 2000$, $\rho_{i,j} = 0.3$, $\eta_{i,j} = 0.1/(1 - H_i - H_j)$ with different Hurst parameters.

(a) mfBm with $m = 3$ and $H_i = 0.25$ for all $i = 1, \ldots, 3$

(b) mfBm with $m = 5$ and $H_i = 0.75$ for all $i = 1, \ldots, 5$

(c) mfBm with $m = 5$ and different $H_i \in [0.35, 0.75]$
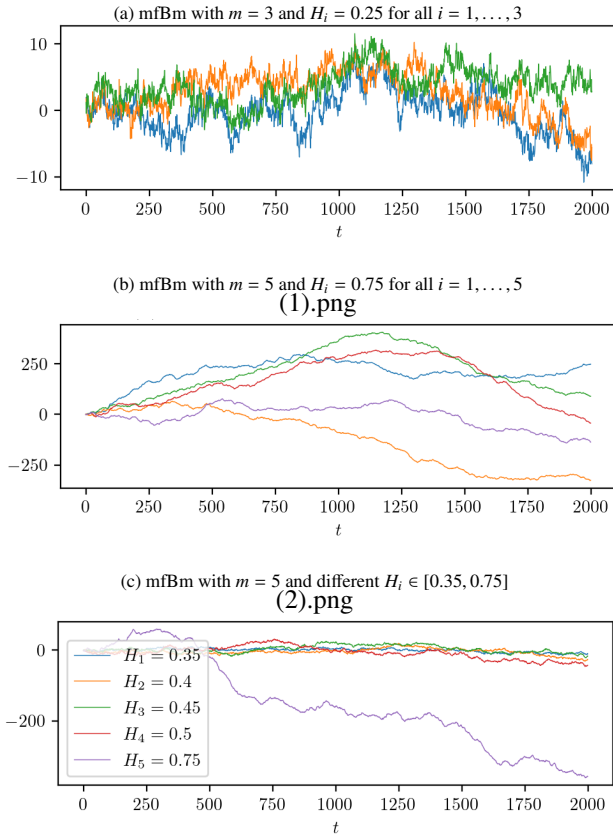
Figure 3: Three different realisations of mfBm with different Hurst parameters.

2.2 Ordinal Pattern Representations

To investigate the qualitative behaviour of mfBm, we use ordinal pattern symbols. Before we get into the details of multivariate ordinal patterns, we introduce univariate concepts.

2.2.1 Ordinal Pattern Symbolisation

In the mathematical theory of symbolic dynamics, a dynamical system is modelled by a discrete space consisting of infinite (sequences of) abstract symbols. These sequences of symbols are the subject of advanced analysis such as entropies. As far as current research is concerned, there are two general approaches to encode the sequence of real-valued measurements into a sequence of symbols as visualised in Figure 4.

On the one hand, classical symbolisation approaches partition the data range according to specified mapping rules in order to then encode a numerical time series into a sequence of discrete symbols from a predefined alphabet Σ. An example of the partitioning of a data range can be found in Figure 4a, while the assignment of the symbols (encoding) to the time series is visualised in Figure 4b. A corresponding and well-known algorithm for determining data range

partitions is Symbolic Aggregate ApproXimation (SAX) presented in [22]. On the other hand, ordinal pattern symbolisation is another approach that, independent of the data range of the time series, encodes the total order between two or more neighbours ($x < y$ or $x > y$) into symbols. This ordinal pattern approach is based on the idea of Bandt and Pompe [23] and visualised in Figures 4c and 4d. Since we use the ordinal symbolisation scheme in the remainder of this work, we present its formalism and advantages in detail in the following.

Ordinal patterns describe the total order between two or more neighbours encoded by permutations.

Definition 2.4 (Univariate Ordinal Pattern)
A vector $(x_1, \ldots, x_d) \in \mathbb{R}^d$ has *ordinal pattern* $(r_1, \ldots, r_d) \in \mathbb{N}^d$ of order $d \in \mathbb{N}$ if $x_{r_1} \geq \cdots \geq x_{r_d}$ and $r_{l-1} > r_l$ in the case $x_{r_{l-1}} = x_{r_l}$.

Note that equality of two values within a pattern is not allowed. In this case, without loss of generality, the newer value is replaced with a smaller value. Figure 4c shows all possible ordinal patterns of order $d = 3$ of a vector (x_1, x_2, x_3). To symbolise a time series $(x_1, x_2, ..., x_T) \in \mathbb{R}^T$ each point in time $t \in \{d, ..., T\}$ is assigned its ordinal pattern of order d. The order d is chosen to be much smaller than the total length T of the time series to look at smaller ordinal pattern windows within the series and their distributions of "up and down" movements. To assess the overarching trend, delayed behaviour is of interest. The time delay $\tau \in \mathbb{N}$ is the delay between successive points in the symbol sequences. Different delays show different details of the structure of a time series. Figure 4d visualises ordinal pattern determination of order $d = 3$ and time delay $\tau = 1$ of three different time points in a time series marked in blue, orange and magenta. Note that ordinal patterns are determined at any arbitrary point in time.

The ordinal approach has notable advantages in its practical application. First of all, the method is conceptually simple as it reflects man's natural thought of up and down movements and is, therefore, open to interpretation. Second, prior knowledge of the data range or the type of time series is not necessary. The concept can be applied to any time series as long as the range of values is ordered, e.g. $x_t \in \mathbb{R}$. Third, the ordinal approach supports robust and fast implementations [24, 25]. Fourth, it allows for an easier estimation of a good symbolisation scheme compared to the classical symbolisation approaches [26, 27].

2.2.2 Permutation Entropy

Not the ordinal patterns themselves, but their distributions in a *univariate time series* $(x_t)_{t=1}^T$ are of interest. Thus, each pattern is identified with exactly one of the *ordinal pattern symbols* $j = 1, 2, ..., d!$. To assess the (dis)order of the identified symbols in the system or in the time series, we use a measure of the dispersion from the field of statistics, namely entropy. The greater the disorder, the higher the entropy. The statistical interpretation of entropy corresponds to Shannon's information entropy used in information theory or computer science. Applying the well-known formula for (Shannon) entropy

$$I(Z) = -\sum_{z \in Z} p_z \ln p_z, \tag{5}$$

$a = (-\infty, 10]$
$b = (10, 20]$
$c = (20, 30]$
$d = (40, 50]$
$e = (50, 60]$
$f = (60, \infty]$
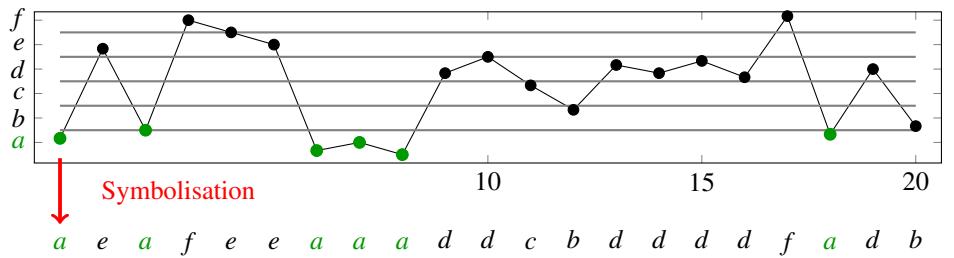
(a) All symbols from an alphabet Σ of size $|\Sigma| = 6$.

(b) Partitioning of the data range and classical symbolisation determination in a univariate time series.

(0 1 2) (2 1 0) (1 0 2)

(2 0 1) (0 2 1) (1 2 0)
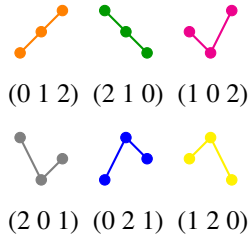
(c) All possible ordinal patterns of order $d = 3$.
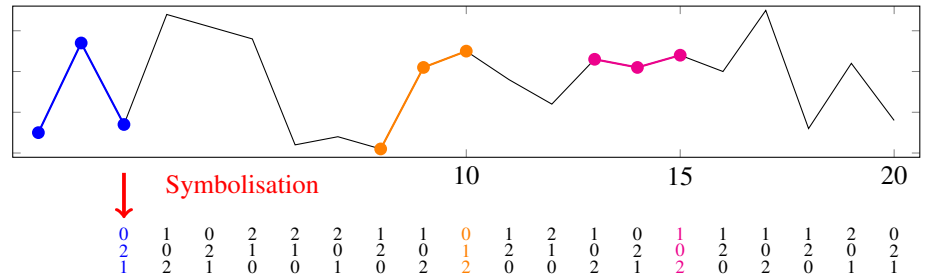
(d) Ordinal pattern determination of order $d = 3$ and time delay $\tau = 1$ at any time point $t \in [d\tau - \tau + 1, T]$.

Figure 4: Two approaches for symbolising a univariate time series: (a) and (b) classical symbolisation and (c) and (d) ordinal pattern symbolisation. Best viewed in colour.

where Z is an alphabet of symbols $Z = \{z_1, z_2, \ldots, z_n\}$ to ordinal pattern symbols leads directly to the definition of permutation entropy.

Definition 2.5 (Permutation Entropy [23])
Permutation entropy (PE) of order $d \in \mathbb{N}$ and delay $\tau \in \mathbb{N}$ of a univariate time series $\mathbf{x} = (x_t)_{t=1}^T$, $T \in \mathbb{N}$ is defined by

$$\text{PE}_{d,\tau}(\mathbf{x}) = - \sum_{j=1}^{d!} p_j^{\tau,d} \ln p_j^{\tau,d}, \qquad (6)$$

where $p_j^{\tau,d}$ is the frequency of occurrence of ordinal pattern j in the time series.

In time series with maximally random ordinal pattern symbols, the ordinal patterns are equally distributed so that PE is $\ln(d!)$. For a time series with a regular pattern, e.g., in the case of strict monotony, PE is equal to zero [7].

2.2.3 Multi-Scale-Permutation Entropy

For example, biological and physiological time series, as well as time series from other fields, often contain complex correlations between both temporal and spatial levels (scales) that need to be uncovered. PE achieves a maximum entropy value both on completely random time series and on series with complex correlations based on a determinism. To uncover multi-scale complex correlations, multi-scale entropy (MSE) proposed in [28] provides a systematic procedure to assign small complexity values to both fully predictable and fully random uncorrelated time series. In contrast, correlated processes across different scales have a high complexity value. In [9], the author extend the concept of MSE to univariate ordinal patterns.

For the consideration of different scales of a time series and an associated definition of MSPE, a coarse-grained procedure is used:

From the original time series, several consecutive time data points are averaged within a non-overlapping time window of scaling length ϵ, also called *scaling factor*. Each element of the coarse-grained time series $\mathbf{y} = (y_l^{(\epsilon)})_{l=1}^{T/\epsilon}$ is calculated as:

$$y_l^{(\epsilon)} = \frac{1}{\epsilon} \sum_{t=(l-1)\epsilon+1}^{l\epsilon} x_t \qquad (7)$$

for $1 \leq l \leq \frac{T}{\epsilon}$.

Definition 2.6 (Multi-Scale Permutation Entropy [9])
Multi-scale permutation entropy (MSPE) of order $d \in \mathbb{N}$ and delay $\tau \in \mathbb{N}$ of a univariate time series $\mathbf{x} = (x_t)_{t=1}^T$, $T \in \mathbb{N}$ is defined as PE of its coarse-grained time series $\mathbf{y} = (y_l^{(\epsilon)})_{l=1}^{T/\epsilon}$, that is

$$\text{MSPE}_{d,\tau,\epsilon}(\mathbf{x}) = \text{PE}_{d,\tau}(\mathbf{y}). \qquad (8)$$

2.2.4 Weighted Permutation Entropy

Another shortcoming in the above Definition 2.5 of PE is that when the ordinal patterns are extracted, no information other than the order structure is preserved. In particular, information about the amplitude in a time series is lost. However, ordinal patterns with large differences in amplitude should contribute in different ways to the computation of PE. The weighted permutation entropy introduced in [10] allows for the weighting of ordinal patterns by exploiting amplitude information resulting from small fluctuations in the time series due to the effect of noise to be weighted less than ordinal patterns with large amplitudes.

Definition 2.7 (Weighted Permutation Entropy [10])
Weighted permutation entropy (WPE) of order $d \in \mathbb{N}$ and delay

$\tau \in \mathbb{N}$ of a univariate time series $\mathbf{x} = (x_t)_{t=1}^T$, $T \in \mathbb{N}$ is defined as

$$\text{WPE}_{d,\tau}(\mathbf{x}) = - \sum_{j}^{d!} p_{w_j}^{\tau,d} \ln p_{w_j}^{\tau,d} \qquad (9)$$

with

$$p_{w_j}^{\tau,d} = \frac{\sum_{t \leq T} w_t \cdot [(x_{t-(d-1)\tau}, ..., x_{t-\tau}, x_t) \text{ has pattern } j]}{\sum_{t \leq T} w_t \cdot [(x_{t-(d-1)\tau}, ..., x_{t-\tau}, x_t)]} \qquad (10)$$

where $[x] = 1$ if x is true, and 0 otherwise, and

$$w_t = \frac{1}{d} \sum_{k=1}^{d} (x_{t-(k-1)\tau} - \overline{x}_t^{d,\tau})^2 \qquad (11)$$

is the empirical variance of the sub-sequence and $\overline{x}_t^{d,\tau}$ denotes its arithmetic mean, i.e., $\overline{x}_t^{d,\tau} = \frac{1}{d} \sum_{k=1}^{d} x_{t-(k-1)\tau}$.

2.3 On the Behaviour of Permutation Entropy on Fractional Brownian Motion

This work investigates the behaviour of PE on fBm in a multivariate setting. To this purpose, we recapitulate univariate results from [8]. They investigate the distribution of ordinal patterns in fBm of different orders and, if possible, provide closed formulas for computation of pattern distributions as follows:

Order $d = 2$. The ordinal patterns are equally distributed, more specifically

$$p_{01}^\tau = p_{10}^\tau = 1/2, \qquad (12)$$

such that $\text{PE}_{2,\tau}(B_H(t)) = -\ln(1/2)$ for all τ.

Order $d = 3$. The distribution of ordinal patterns is given by

$$p_{012}^\tau = \frac{1}{\pi} \arcsin 2^{H-1} =: u \qquad (13)$$

for all τ. Furthermore, for a Gaussian process with stationary increments (as fBm), it holds that

$$p_j^\tau = \begin{cases} u & \text{if } j = (012), (210), \\ (1-2u)/4 & \text{otherwise,} \end{cases} \qquad (14)$$

for all τ. In particular, $\text{PE}_{3,\tau}(B_H(t))$ is independent of the delay τ but monotonically dependent on the Hurst parameter H.

Order $d = 4$. The distribution of ordinal patterns can also be expressed, albeit in a more complex formula

$$p_j^\tau = \frac{1}{8} + \frac{1}{4\pi} \cdot v_j \qquad (15)$$

for all delays τ, where

$$v_j = \begin{cases} \arcsin \alpha_1 + 2 \arcsin \alpha_2 & \text{if } j = (0123), (3210), \\ \arcsin \alpha_4 - 2 \arcsin \alpha_5 & \text{if } j = (3120), (0213), \\ 2 \arcsin \alpha_6 + \arcsin \alpha_1 & \text{if } j = (1032), (2301), \\ \arcsin \alpha_7 - \arcsin \alpha_1 - \arcsin \alpha_5 & \text{if } j = (0132), (1023), \\ & \qquad (2310), (3201), \\ \arcsin \alpha_7 - \arcsin \alpha_4 - \arcsin \alpha_5 & \text{if } j = (0312), (3021), \\ & \qquad (2130), (1203), \\ \arcsin \alpha_3 + \arcsin \alpha_8 - \arcsin \alpha_5 & \text{if } j = (2013), (0231), \\ & \qquad (3102), (1320), \\ \arcsin \alpha_6 - \arcsin \alpha_8 + \arcsin \alpha_2 & \text{if } j = (0321), (3012), \\ & \qquad (1230), (2103), \end{cases} \qquad (16)$$

with

$$\alpha_1 = \frac{1 + 3^{2H} - 2^{2H+1}}{2}, \quad \alpha_2 = 2^{2H-1} - 1, \quad \alpha_3 = \frac{1 - 3^{2H} - 2^{2H}}{2 \cdot 6^H},$$

$$\alpha_4 = \frac{3^{2H} - 1}{2^{2H+1}}, \qquad \alpha_5 = 2^{2H-1}, \qquad \alpha_6 = \frac{-1 - 3^{2H} + 2^{2H}}{2 \cdot 3^H},$$

$$\alpha_7 = \frac{3^{2H} - 2^{2H} - 1}{2^{2H+1}}, \qquad \text{and} \qquad \alpha_8 = \frac{2^{2H-1}}{3^H}.$$

In particular, $\text{PE}_{4,\tau}(B_H(t))$ is independent of the delay τ but monotonically dependent on the Hurst parameter H.

Order $d \geq 5$. There are no closed formulas.

The formulas for the distribution of univariate order patterns of different orders lead to the following behaviour of PE and its extensions MSPE and WPE on fBm.

Theorem 2.2 *For order $d = 2$, it holds*

$$\begin{aligned} \text{PE}(B_H(t)) &= -\ln(1/2) & (17) \\ &= \text{MSPE}(B_H(t)) & (18) \\ &= \text{WPE}(B_H(t)). & (19) \end{aligned}$$

for all delays $\tau \in \mathbb{N}$ and all scales ϵ.

Proof. Equation (17) follows directly from (12). Equations (18) or (19) are shown in [1] and [2], respectively. The independence of MSPE from the scale ϵ follows from the invariance of the coarse-grained procedure of fBm. $\quad \square$

Theorem 2.3 *For orders $d = 3$ and $d = 4$, $\text{PE}(B_H(t))$ and $\text{MSPE}(B_H(t))$ are independent of all delays $\tau \in \mathbb{N}$ but monotonically dependent on the Hurst parameter H. Moreover, $\text{MSPE}(B_H(t))$ is independent of all scales ϵ, and thus $\text{PE}(B_H(t))$ and $\text{MSPE}(B_H(t))$ are identical.*

Proof. The independence of MSPE from the scale ϵ follows from the invariance of the coarse grain procedure of fBm and is shown in [1]. Thus, independence of the delay τ and the dependence on the Hurst parameter H follows directly from (13). $\quad \square$

In [2] we observe that for WPE of order $d = 3$, certain ordinal patterns, namely the strictly ascending ordinal pattern (012) and the strictly descending pattern (210), have higher weights and thus have more impact on the computation of WPE than the other four ordinal patterns. As a result, WPE decreases faster than PE for increasing Hurst parameter H. We elaborate on this experimental discovery in Section 6.

3 A Review of Multivariate Permutation Entropy

This section provides a comprehensive overview of existing extensions of permutation entropy to the multivariate case. We categorise existing multivariate extensions according to their procedure before presenting them in detail.

3.1 Different Approaches for Multivariate Permutation Entropy

A *multivariate time series* $((x_t^i)_{i=1}^m)_{t=1}^T$ such as a certain realisation of mfBm has more than one time dependent variable. Each variable x^i for $i \in 1, ..., m$ not only depends on the respective past values in time but also has some dependence on other variables in phase space. Considering two time points $(x_t^i)_{i=1}^m$ and $(x_{t+1}^i)_{i=1}^m$ with m variables, simply put two vectors, it is not possible to establish a total order between them. A total order is only possible if $x_t^i > x_{t+1}^i$ or $x_t^i < x_{t+1}^i$ for all $i \in 1, ..., m$. Therefore, there is no trivial generalisation of the PE algorithm to the multivariate case.

Nevertheless, numerous studies deal with the multivariate extensions of PE. We classify proposed extensions into four procedures. The multivariate time series is

a) projected into univariate ordinal space by determining univariate ordinal pattern *between neighbouring values in time space* (row marked in red in Figure 5a) in each single variable i, and then pooled over all m variables for multidimensionality, or

b) projected into univariate ordinal space by determining univariate ordinal pattern *between values of all m variables* (column marked in red in Figure 5b) for each single time point t, and then pooled over all T time points for multidimensionality, or

c) projected into multivariate ordinal space by determining the multivariate ordinal pattern *between d vectors of variable dimension m* (box marked in blue in Figure 5c for $d = 2$), or

d) projected onto a single-dimensional reduction first (box marked in green in Figure 5d), and then transformed into ordinal space by determining univariate ordinal pattern.

The first procedure a) is a canonical extension of the univariate definition and is presented in [29] as pooled permutation entropy (PPE). PPE is probably the most widely used multivariate extension. While PPE measures the complexity of each variable in time space, procedure b) analyses the complexity of the variables in phase space. A variant of procedure b) is introduced in [30] as Multivariate Permutation Entropy (MvPE). The third procedure c) first requires an extension of univariate ordinal pattern to multidimensionality. Some theoretical basis for this is set in [31, 32]. A detailed discussion and challenge on real-world data can be found in [33]. The idea of the fourth procedure d) is to first reduce the number of variables m to a single dimension, i.e., $m = 1$, by applying a distance measure or dimension reduction algorithm. Consequently, Definition 2.5 can then be used for PE computation directly. Rayan, in [34], the author propose several distance measures to reduce the dimensionality, that is Euclidian distance with

reference point $(x_{t=0}^i)_{i=1}^m$, Manhattan distance with reference point $(x_{t=0}^i)_{i=1}^m$, and Euclidian distance with reference point $\mathbf{0}$. To consider correlations between variables in space, in [33], the author propose to reduce the dimension using principal component analysis.

(a) Univariate ordinal pattern in time.

(b) Univariate ordinal pattern in phase space.

(c) Multivariate ordinal pattern.

(d) Dimensionality reduction.
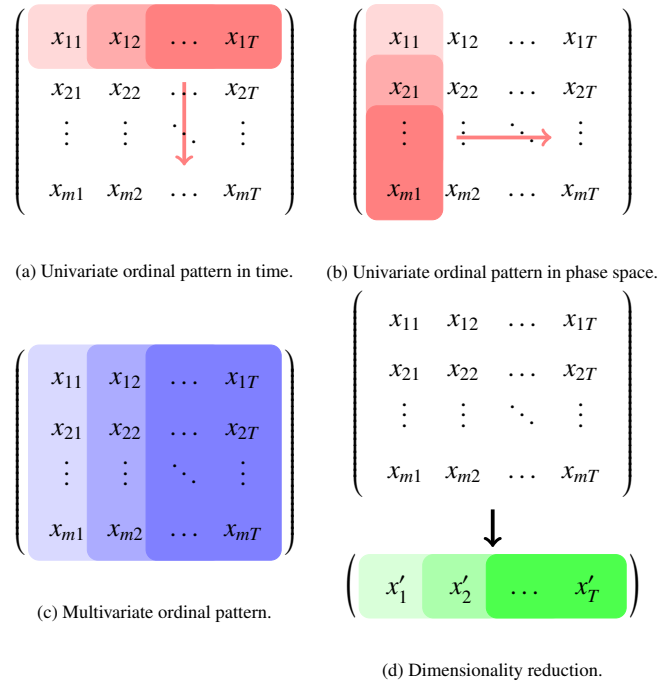
Figure 5: Four procedures of ordinal pattern determination in a multivariate time series.

Since the determination of ordinal patterns in both procedures b) and d) requires transformations of the structure of mfBms, we restrict ourselves in the following to extensions a) and c) that are some kind of canonical extensions.

3.2 Pooled Permutation Entropy and Adaptions

The procedure in Figure 5a is a canonical extension of PE to the multivariate case, i.e., univariate ordinal patterns are determined for each variable of the multivariate time series before each pattern of all variables is pooled. This idea is introduced in [29] and referred to as *pooled permutation entropy* (PPE). In addition, we introduce two adaptations based on PPE that take into account different aspects of a time series, i.e., different scales or amplitudes, called *multivariate multi-scale permutation entropy* (MMSPE) or *multivariate weighted permutation entropy* (MWPE), respectively.

3.2.1 Pooled Permutation Entropy

The idea of PPE is to use marginal frequencies of $d!$ ordinal patterns regarding all m variables as input for entropy computation. For the determination of marginal frequencies, an auxiliary matrix has to be established first as follows.

1. For each variable $i = 1, .., m$ and for each ordinal pattern $j = 1, ..., d!$, count all time steps $s \in [\tau(d-1) + 1, T]$, for which the variable-time pair (i, s) has the ordinal pattern j.

2. Divide the counts by $m \cdot \delta$, where $\delta := T - \tau(d-1)$ is the total count of ordinal patterns each variable has.

3. Store the results, i.e., frequencies $p_{ij}^{\tau,d}$ in a so-called *pooling matrix* $P \in (0,1)^{m \times d!}$, which reflects the distribution of ordinal patterns in the multivariate time series across its m variables.

It holds $\sum_{i=1}^{m} \sum_{j=1}^{d!} p_{ij}^{\tau,d} = 1$. For computational reasons, marginal frequencies $p_{\cdot j}^{\tau,d} = \sum_{i=1}^{m} p_{ij}^{\tau,d}$ must not vanish for $j = 1, ..., d!$. If they vanish, set the value close to zero.

Definition 3.1 (Pooled Permutation Entropy [29])
Pooled permutation entropy (PPE) of a multivariate time series $\mathbf{X} = ((x_t^i)_{i=1}^{m})_{t=1}^{T}$ is defined as PE of the marginal frequencies $p_{\cdot j}^{\tau,d} = \sum_{i=1}^{m} p_{ij}^{\tau,d}$ for $j = 1, ..., d!$ describing the distribution of the ordinal pattern and can be calculated as

$$\text{PPE}_{d,\tau}(\mathbf{X}) = - \sum_{j}^{d!} p_{\cdot j}^{\tau,d} \ln p_{\cdot j}^{\tau,d}. \tag{20}$$

In the literature, PPE is often referred to as Multivariate Permutation Entropy. To avoid confusion with other multivariate versions, we use the original naming. A major advantage of PPE is that within the procedure also univariate PE can be derived for each variable by calculating the PE based on the frequencies p_{ij} of all ordinal patterns $j = 1, \ldots, d!$ for corresponding $i = 1, \ldots, m$, i.e. row by row on the pooling matrix. Algorithm 1 provides pseudocode for computing PPE.

Algorithm 1: Computation of PPE

Input: Multivariate Time Series $X^{m \times T}$, Order d, Delay τ
Function pooling(X,d,τ):
  $P^{m \times d!} \leftarrow$ pooling matrix initialised with zeros
  for *every time series variable* $i = 1, \ldots, m$ *in X* **do**
    for *each ordinal pattern* $j = 1, \ldots, d!$ **do**
      $c \leftarrow$ number of time steps $s \in [\tau(d-1)+1, T]$ with pattern j
      $P_{ij} \leftarrow$ divide c by the number of total time steps $\delta \cdot m$ in the multivariate time series X
  return P
Function marginalisation(P):
  $p^{1 \times d!} \leftarrow$ vector of marginalisation initialised with zeros
  for *every column* $j = 1, \ldots, d!$ *in $P^{m \times d!}$* **do**
    $p_j \leftarrow$ sum up p_{ij}
  return p
return PE(p)

For example, PPE is successfully used in the analysis of electroencephalography (EEG) signals, as cross-channel regularities between spatially distant variables, i.e., on different hemispheres or in different areas, can be extracted by long-range spatial nonlinear correlations [29]. Furthermore, in [24], the author successfully used PPE to study the 19-channel scalp EEG reflecting changes in brain dynamics of a boy with lesions predominantly in the left temporal lobe caused by connatal toxoplasmosis. In [35], the author use PPE to characterise sleep EEG signals from more than

80 hours of nocturnal sleep recordings and classify sleep stages. They find that each sleep stage is characterised by statistically different PPE values and that the observed pattern of PPE is consistent with the physiological properties of the EEG in each sleep stage.

3.2.2 Multivariate Multi-Scale Permutation Entropy

In addition to PPE and by analogy with MSPE, Morabito et al. [9] provide a canonical definition of multivariate multi-scale permutation entropy. MSPE and MMSPE are calculated on different time scales by processing the coarse-grained time series as a function of the scale factor ϵ. Per variable i in a multivariate time series, several consecutive time data points are averaged within a non-overlapping time window of the scaling length ϵ. Each element of the coarse-grained time series $\mathbf{Y} = ((y_{i,l}^{(\epsilon)})_{l=1}^{T/\epsilon})_{i=1}^{m}$ is calculated as:

$$y_{i,l}^{(\epsilon)} = \frac{1}{\epsilon} \sum_{t=(l-1)\epsilon+1}^{l\epsilon} x_{i,t} \tag{21}$$

for all $i = 1, ..., m$ and $1 \le l \le \frac{T}{\epsilon}$.

Definition 3.2 (Multivariate Multi-Scale PE [9])
Multivariate multi-scale permutation entropy (MMSPE) of order $d \in \mathbb{N}$ and delay $\tau \in \mathbb{N}$ of a multivariate time series $\mathbf{X}$ is defined as PPE of its coarse-grained time series $\mathbf{Y}$, that is

$$\text{MMSPE}_{d,\tau,\epsilon}(\mathbf{X}) = \text{PPE}_{d,\tau}(\mathbf{Y}). \tag{22}$$

The simultaneous utilisation of a multi-scale approach and the consideration of multiple variables of the time series facilitates assessing the complexity of the underlying dynamical system. Algorithm 2 provides pseudocode for computing MMSPE.

Algorithm 2: Computation of MMSPE

Input: Multivariate Time Series $X^{m \times T}$, Order d, Delay τ, Scale Factor ϵ
$Y^{m \times (T/\epsilon)} \leftarrow$ coarse-grained multivariate time series initialized with zeros
for *every time series variable* $i = 1, \ldots, m$ *in X* **do**
  $y_i \leftarrow$ calculate coarse-grained univariate time series
                                    // see Eq.(21)
return PPE(Y, d, τ)

3.2.3 Multivariate Weighted Permutation Entropy

In addition to PPE and by analogy with WPE, in [2] we provide a canonical definition of multivariate weighted permutation entropy. Again, for the determination of MWPE, an auxiliary matrix has to be established first:

1. For each variable $i = 1, .., m$ and for each ordinal pattern $j = 1, ..., d!$, select all time steps $s \in [\tau(d-1)+1, T]$ for which the variable-time pair (i, s) has the ordinal pattern j.

2. Add up the weights w_t, i.e., $\mathbf{w_{ij}} = \sum_{t=d\tau-\tau+1}^{T} w_t$ for all selected ordinal pattern vectors j and for each variable $i = 1, ..., m$. Note that the total count of weights w_t^i for each variable i is $\delta := T - (d\tau - \tau)$.

3. Divide the weighted sum $\mathbf{w_{ij}}$ by the total sum of all $m \cdot \delta$ weights to obtain the weighted frequencies for every ordinal pattern j.

4. Store the results, i.e., weighted frequencies $p_{w_{ij}}^{\tau,d}$ in a so-called *weighted pooling matrix* $P_w^{\tau,d} \in \mathbb{R}^{m \times d!}$, which reflects the weighted distribution of ordinal patterns in the multivariate time series across its m variables.

Based on the weighted pooling matrix $P_w^{\tau,d}$, multivariate weighted permutation entropy can be calculated as follows.

Definition 3.3 (Multivariate Weighted PE [2])
Multivariate weighted permutation entropy (MWPE) of a multivariate time series $\mathbf{X} = ((x_t^i)_{i=1}^m)_{t=1}^T$ is defined as PE of the marginal weighted frequencies $p_{w\cdot j}^{\tau,d} = \sum_{i=1}^m p_{w_{ij}}^{\tau,d}$ for $j = 1, ..., d!$ describing the distribution of the weighted ordinal pattern and can be calculated as

$$\mathrm{MWPE}_{d,\tau}(\mathbf{X}) = -\sum_j^{d!} p_{w\cdot j}^{\tau,d} \ln p_{w\cdot j}^{\tau,d}. \qquad (23)$$

Algorithm 3 provides pseudocode for computing MWPE.

Algorithm 3: Computation of MWPE

Input: Multivariate Time Series $X^{m \times T}$, Order d, Delay τ
Function `weightedPooling(X,d,τ)`:
  $P^{m \times d!} \leftarrow$ weighted pooling matrix initialised with zeros
  for *every time series variable $i = 1, \ldots, m$ in X* **do**
    for *each ordinal pattern $j = 1, \ldots, d!$* **do**
      $w_{ij} \leftarrow$ add up weights w_t // see Eq.(11)
      $p_{ij} \leftarrow$ divide w_{ij} by the total sum of all $m \cdot \delta$
        weights
  return P
$p \leftarrow$ `marginalisation(P)`
return PE(p)

3.3 Multivariate Ordinal Pattern Permutation Entropy

A *multivariate* time series $((x_t^i)_{i=1}^m)_{t=1}^T$ and its corresponding data matrix $X \in \mathbb{R}^{m \times T}$ has more than one time-dependent variable. Each variable x^i for $i \in 1, ..., m$ not only depends on the respective past values in time but also has some dependence on other variables in phase space. Since all procedures from the previous section consider univariate ordinal patterns per variable i separately, interdependence is not considered. To involve interdependence, an intuitive idea introduced in [33] is to store the univariate ordinal patterns of all variables at a time point t together into one common multivariate ordinal pattern.

Definition 3.4 (Multivariate Ordinal Pattern) A data matrix $(x_1, ..., x_d) \in \mathbb{R}^{m \times d}$ has *multivariate ordinal pattern*

$$\begin{pmatrix} r_{11} & \cdots & r_{1d} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{md} \end{pmatrix} \in \mathbb{N}^{m \times d} \qquad (24)$$

of order $d \in \mathbb{N}$ if $x_{r_{i1}} \geq ... \geq x_{r_{id}}$ for all $i = 1, ..., m$ and $r_{il-1} > r_{il}$ in the case $x_{r_{il-1}} = x_{r_{il}}$.

Figure 6 shows all possible multivariate ordinal patterns of order $d = 3$ and number of variables $m = 2$.
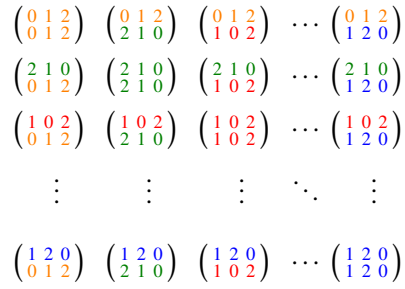
Figure 6: All possible multivariate ordinal patterns of order $d = 3$ with $m = 2$ variables. Best viewed in colour.

With the natural extension of univariate ordinal pattern, which leads to Definition 3.4, it is possible to apply the PE algorithm from Definition 2.5 to multivariate time series in its original form.

Definition 3.5 (Multivariate Ordinal Pattern PE [33])
Multivariate ordinal pattern permutation entropy (MOPPE) of order $d \in \mathbb{N}$ and delay $\tau \in \mathbb{N}$ of a multivariate time series $\mathbf{X} = ((x_t^i)_{i=1}^m)_{t=1}^T$ is defined by

$$\mathrm{MOPPE}_{d,\tau}(\mathbf{X}) = -\sum_{j=1}^{d!} p_j^{\tau,d} \ln p_j^{\tau,d}, \qquad (25)$$

where $p_j^{\tau,d}$ is the frequency of multivariate ordinal pattern j in the multivariate time series.

Algorithms 1, 2, 3 and 4 for computing PPE, MMSPE, MWPE and MOPPE, respectively, can also be found on Github[1] and PyPI[2].

Algorithm 4: Computation of MOPPE

Input: Multivariate Time Series $X^{m \times T}$, Order d, Delay τ
$p^{1 \times d!} \leftarrow$ frequencies initialised with zeros
for *every time step t in X* **do**
  for *each ordinal pattern $j = 1, \ldots, d!$* **do**
    $c \leftarrow$ count all time points t with multivariate ordinal
      pattern j // see Def. 3.4
    $p_j \leftarrow$ divide c by the number of total time steps
      $T - \tau(d-1)$ in the multivariate time series X
return PE(p)

The number of the possible multivariate ordinal pattern increases exponentially with the number of variables m, i.e., $(d!)^m$. Therefore, if d and m are too large, depending on the application, each pattern occurs only rarely or some not at all, resulting in a uniform distribution of ordinal patterns. This has the consequence that subsequent learning procedures may fail. Nevertheless, for small order d and sufficiently large length T of the time series, the use of multivariate ordinal patterns can lead to higher accuracy in learning tasks, e.g., classification [33], because they incorporate interdependence.

4 PPE Applied to mfBm

In this section, we investigate the behaviour of PPE of different orders and delays on mfBm in the variation of its Hurst parameter from both theoretical and experimental points of view. For the sake of completeness, we recapitulate results from previous work and fill in gaps with further results.

4.1 Theoretical Analysis

Orders 2 **and** 3. In previous work, [1] we show that for order $d = 2$ the PPE of mfBm is constant for all delays τ as well as any number of variables m, in particular, it holds that

$$\text{PPE}_{2,\tau}(\mathbf{B}_H^m(t)) = -\ln(1/2) \tag{26}$$

for all τ, m. Furthermore, we show that for order $d = 3$, the PPE of mfBm stays also independent of the delay τ, but depends monotonically on the number of variables m as well as the Hurst parameter H. In the case $H_i = H_j$ for all variables i, j, PPE is also independent of the number of variables m. Details and corresponding formulas for entropy calculation can be found in the original paper [1].

Orders greater than 3. In this work, we show that a behaviour analogous to that of PPE with order $d = 3$ can be transferred to orders $d > 3$.

Theorem 4.1 *PPE*$(\mathbf{B}_H^m(t))$ *of order $d = 4$ is independent of all delays τ but monotonically dependent on the number of variables m and the Hurst parameter H of mfBm. In the case $H_i = H_j$ for all i, j, it is independent on the number of variables m.*

Proof. Let $d = 4$. Since mfBm $\mathbf{B}_H^m(t)$ is marginally an fBm $B_H(t)$, we conclude from Definition 2.3 and (15), that

$$p_{ij}^{\tau,4} = \frac{1}{m}\left(\frac{1}{8} + \frac{1}{4\pi} \cdot v_j\right) \tag{27}$$

for all $i = 1, ..., m$ and v_j as in (16).

The marginal frequencies for $j = 1, ..., d!$ are given by

$$p_{.j}^{\tau,4} = \sum_{i=1}^{m} \frac{1}{m}\left(\frac{1}{8} + \frac{1}{4\pi} \cdot v_j\right), \tag{28}$$

where v_j is as in (16), while

$$\alpha_1 = \frac{1 + 3^{2H_i} - 2^{2H_i+1}}{2}, \quad \alpha_2 = 2^{2H_i-1} - 1, \quad \alpha_3 = \frac{1 - 3^{2H_i} - 2^{2H_i}}{2 \cdot 6^{H_i}},$$

$$\alpha_4 = \frac{3^{2H_i} - 1}{2^{2H_i+1}}, \qquad \alpha_5 = 2^{2H_i-1}, \qquad \alpha_6 = \frac{-1 - 3^{2H_i} + 2^{2H_i}}{2 \cdot 3^{H_i}},$$

$$\alpha_7 = \frac{3^{2H_i} - 2^{2H_i} - 1}{2^{2H_i+1}}, \qquad \text{and} \qquad \alpha_8 = \frac{2^{2H_i-1}}{3^{H_i}}.$$

Considering the distribution of all possible realisations of fBm, we see from the use of (28) in Definition 3.1 that the number of variables m in the case of $H_i = H_j$ for all i, j cancels out. Thus, PPE$(\mathbf{B}_H^m(t))$ is independent of the number of variables m. In addition, PPE$(\mathbf{B}_H^m(t))$ stays independent of all delays τ. Monotonic

dependence on the Hurst parameter H is preserved since the transformations carried out, i.e. additions, do not change the monotonicity. $\square$

For order $d > 4$, no closed formulas for the distribution of ordinal patterns in fBm exist (see Section 2.3). Nevertheless, analogous behaviour can also be observed in this case, and we we evaluate that in an experimental setting in the next section.

4.2 Experimental Evaluation

In this experimental evaluation, we investigate the behaviour of PPE on mfBm in the variation of the Hurst parameter H, the number of variables m and the delay τ. All experimental computations are based on simulations of mfBms using an algorithm described in [20] with parameters $H_i = H_j$ for all i, j, $\rho_{i,j} = 0.0$ and $\eta_{i,j} = 0.1/(1 - H_i - H_j)$ for Lemma 2.1. The lengths $T = 10,000$ of mfBms are assumed to be large. The results of computations are visualised in Figure 7.

In Figure 7a we show the independence of PPE of all orders $d = 2$ (green), $d = 3$ (orange), $d = 4$ (blue), and $d = 5$ (red) on mfBm from number of variables m. For this, the delay is fixed to $\tau = 1$. All green lines, i.e., for $d = 2$, are constant, namely $-\ln(1/2)$, for each m, confirming the independence of numbers of variables m in (26). All other lines, i.e., for $d > 2$, are also the same for any m, but monotonically dependent on the Hurst parameter H, i.e., in particular confirming Theorem 4.1. In Figure 7b we show the independence of PPE on mfBm from delay τ. For this, the number of variables is fixed to $m = 3$. Again, all green lines, i.e., for $d = 2$, are constant $-\ln(1/2)$ for each τ, confirming the independence of delay τ in (26). All other lines, i.e., for $d > 2$, are also the same for any τ, but monotonically dependent on the Hurst parameter H, in particular confirming Theorem 4.1.

The deviations of the values of PPE in Figure 7 with increasing Hurst parameter H thus result from the length restriction in the simulation, i.e., the distributions as given in the theoretical part are only to be expected if $T \to \infty$ converges [13]. Thus, for a small length T, the estimates of the probabilities for the ordinal patterns differ from the true values of a hypothetical mfBm of infinite length. All in all, the experiment underpins our theoretical results. The results are not surprising since PPE pools over the univariate patterns so that mfBm can be understood by the computational logic as a kind of extended univariate fBm. This kind of extension ensures that the deviations for greater Hurst parameter H and larger number of variables become smaller and can also be observed in Figure 7a.

5 MMSPE Applied to mfBm

In this section, we investigate the behaviour of MMSPE of different orders, delays, and scales on mfBm in the variation of its Hurst parameter from both theoretical and experimental points of view. For the sake of completeness, we recapitulate results from previous work and fill in remaining results.

5.1 Theoretical Analysis

To investigate the behaviour of MMSPE on mfBm, coarse-grained fractional Brownian motion has to be considered. By using (7)
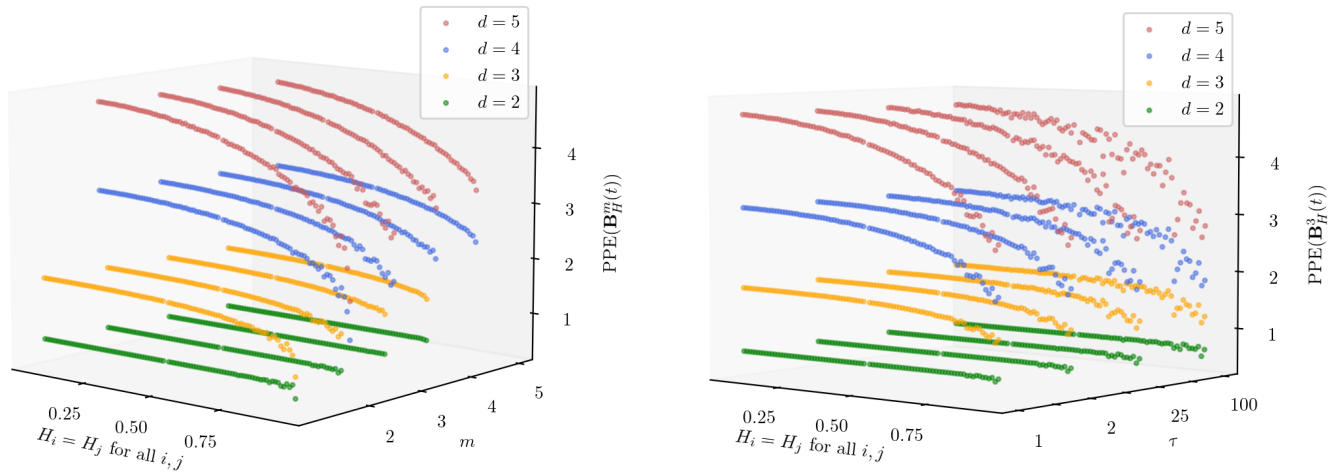
(a) PPE of fixed delay $\tau = 1$ on mfBm for different numbers of variables $m = 3$.

(b) PPE of different delays τ on mfBm for fixed variable dimension $m = 3$.

Figure 7: Experimental computations of PPE of orders $d = 2$ (green), $d = 3$ (orange), $d = 4$ (blue), and $d = 5$ (red) on mfBm in variation of its Hurst parameters H.

or (21) on fBm, coarse-grained fractional Brownian motion (cfBm) is defined as

$$B_{H_i}^{i,(\epsilon)}(l) = \frac{1}{\epsilon} \sum_{j=1}^{\epsilon} B_{H_i}^{i}((l-1)\epsilon + j) \qquad (29)$$

for $l = 1, ..., T/\epsilon$ and all $i = 1, ..., m$.

In [1] we show that the structure of the covariance function is the same as the original fBm for all variables $i = 1, ..., m$, but with additional information of the scale factor ϵ. Therefore the covariance function of the original fBm or mfBm is invariant to the coarse-grained procedure. For this reason, the results from the previous Section 4 are directly transferable to MMSPE.

Orders 2 and 3. For order $d = 2$ the MMSPE on mfBm is constant for all delays τ, number of variables m, as well as any scales ϵ. In particular holds

$$\text{MMSPE}_{2,\tau,\epsilon}(\mathbf{B}_H^m(t)) = -\ln(1/2) \qquad (30)$$

for all τ, m, ϵ. Furthermore, for order $d = 3$, the MMSPE on mfBm is also independent of delays τ and scales ϵ but depends monotonically on the number of variables m as well as the Hurst parameter H of mfBm. In the case $H_i = H_j$ for all variables i, j, MMSPE is also independent of the number of variables m. Details and corresponding formulas for entropy calculation can be found in the original work [1].

Orders greater than 3. With previous insights, the transfer of results from MMSPE of order $d = 3$ to MMSPE of order $d = 4$ is straightforward. For the sake of completeness, we formulate the following theorem.

Theorem 5.1 MMSPE($\mathbf{B}_H^m(t)$) *of order $d = 4$ is independent of all delays τ and scales ϵ, but monotonically dependent on the number of variables m and Hurst parameter H of mfBm. In the case $H_i = H_j$ for all i, j, it is independent on the number of variables m.*

Proof. For mfBm, each variable $i = 1, ..., m$ is marginally an fBm. Since the covariance function of fBm is invariant to the coarse-grained procedure in (29), the independence of scales ϵ follows [1]. Then Theorem 4.1 is directly applicable, which implies independence from the delay τ, the number of variables m and monotonic dependence on the Hurst parameter H. □

For order $d > 4$, no closed formulas for the distribution of ordinal patterns in fBm exist (see Section 2.3). Nevertheless, analogous behaviour can also be observed in this case. We evaluate this behaviour in an experimental setting in the next section.

Summarising the results of this subsection, the inclusion of different scales in mfBm has no influence. This is particularly related to or reflects the fractal property because it resembles the original shape when a fractal is zoomed in or scaled. Thus, MMSPE and PPE show identical behaviour, which we investigate and discuss in detail in Section 8.

5.2 Experimental Evaluation

In this experimental evaluation, we investigate the behaviour of MMSPE on mfBm in the variation of the Hurst parameter H, delay τ, number of variables m, as well as scale ϵ. The experimental computations are based on the same simulations of mfBms as in Section 4.2.

In Figure 8a we show the independence of MMSPE of all orders $d = 2$ (green), $d = 3$ (orange), $d = 4$ (blue), and $d = 5$ (red) on mfBm from number of variables m. For this, the delay is fixed to $\tau = 1$, and the scale is fixed to $\epsilon = 2$. All green lines, i.e., for $d = 2$, are constant, namely $-\ln(1/2)$, for each m, confirming the independence of numbers of variables m in (30). All other lines, i.e., for $d > 2$, are also the same for any m, but monotonically dependent on the Hurst parameter H, in particular confirming Theorem 5.1. In Figure 8b we show the independence of MMSPE on mfBm from delay τ. For this, the number of variables is fixed to $m = 3$, and the scale is fixed to $\epsilon = 2$. Again, all green lines, i.e., for $d = 2$, are constant with value $-\ln(1/2)$ for each τ, confirming the independence of delay τ in (30). All other lines, i.e., for $d > 2$, are also the same for any τ, but monotonically dependent on the Hurst parameter H,

in particular confirming Theorem 5.1.
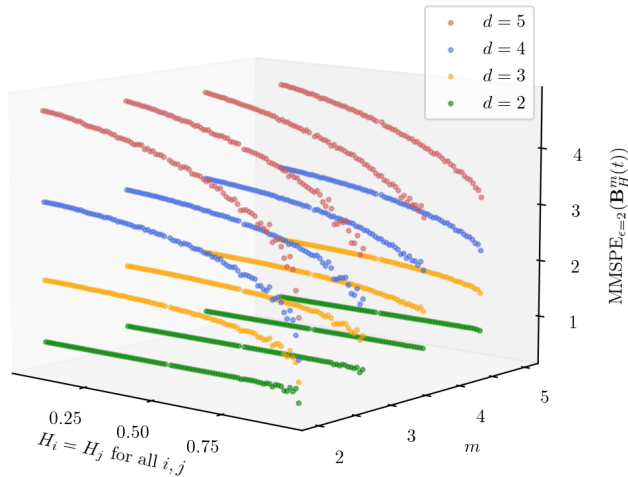
(a) MMSPE of fixed delay $\tau = 1$ and fixed scale $\epsilon = 2$ on mfBm for different numbers of variables $m = 3$.
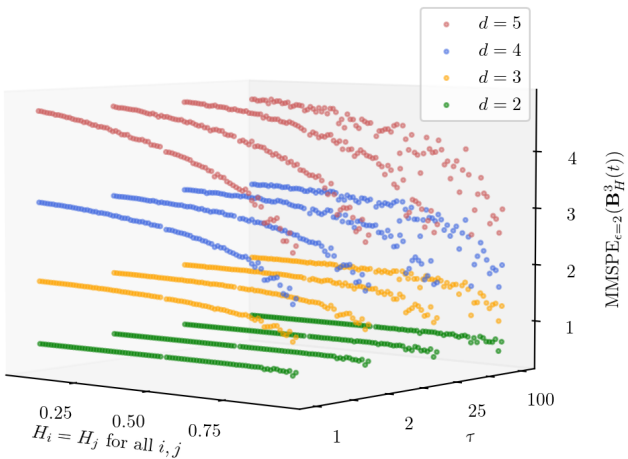
(b) MMSPE of different delays τ and fixed scale $\epsilon = 2$ on mfBm for fixed variable dimension $m = 3$.
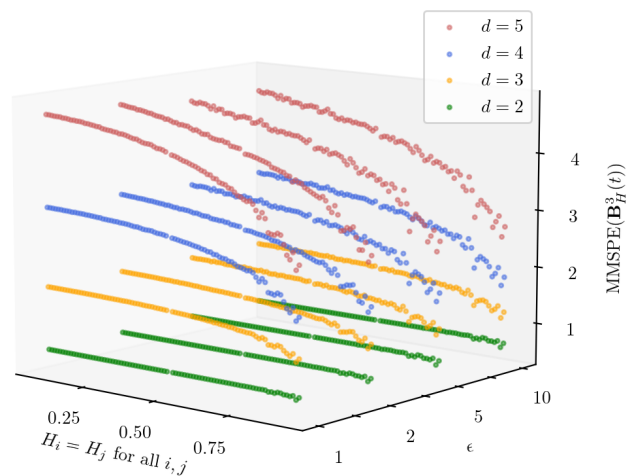
(c) MMSPE of fixed delay $\tau = 1$ and different scales ϵ on mfBm for fixed variable dimension $m = 3$.

Figure 8: Experimental computations of MMSPE of orders $d = 2$ (green), $d = 3$ (orange), $d = 4$ (blue), and $d = 5$ (red) on mfBm in variation of its Hurst parameters H.

In Figure 8c we show the independence of MMSPE on mfBm

from scale ϵ. For this, the number of variables is fixed to $m = 3$ and the delay to $\tau = 1$. Again, all green lines, i.e., for $d = 2$, are constant with value $-\ln(1/2)$ for each ϵ, confirming the independence of scale ϵ in (30). All other lines, i.e., for $d > 2$, are also the same for any scale ϵ, but monotonically dependent on the Hurst parameter H, in particular confirming Theorem 5.1. Again, the reason for the deviations of the values of MMSPE in Figure 8 with increasing Hurst parameter H is the same as in the previous Section 4, namely from length restriction $T < \infty$ of mfBm.

All in all, the experiment underpins our theoretical results from Since MMSPE is based on PPE and scaling has no influence on mfBm, the behaviour of MMSPE on mfBm in Figure 8 is to be expected the same as that of PPE in Figure 7. A direct comparison is discussed in Section 8.

6 MWPE Applied to mfBm

In this section, we investigate the behaviour of MWPE of different orders and delays on mfBm in the variation of its Hurst parameter from both theoretical and experimental points of view. For the sake of completeness, we recap results from past work and add some further results to fill in some of the remaining research gaps.

6.1 Theoretical Analysis

Order 2. In previous work, [2] we show that for order $d = 2$, PPE of mfBm is constant for all delays τ as well as numbers of variables m.

Theorem 6.1 *MWPE of order $d = 2$ on mfBm $B_H(t)$ is given by*

$$\text{MWPE}_{2,\tau}(\mathbf{B}_H^m(t)) = -\ln(1/2) \qquad (31)$$

for all τ, m.

Proof. From [2]: WPE or MWPE differs from PE or PPE, respectively, in that the ordinal patterns are weighted depending on their position t according to (11). For a weight w_t of order $d = 2$, i.e., of two time steps $x_{t-\tau}$ and x_t, we have

$$w_t = \frac{1}{2} \sum_{k=1}^{2} (x_{t-(k-1)\tau} - \bar{x}_t^{2,\tau})^2 \qquad (32)$$

$$= \frac{1}{2}(x_t - x_{t-\tau})^2. \qquad (33)$$

Since $x_t \sim B_H(t)$, we conclude from Definition 2.3, property (3), i.e., fBm has stationary increments, that

$$\frac{1}{2}(B_H(t) - B_H(t - \tau))^2 \sim \frac{1}{2}(B_H(\tau))^2 \qquad (34)$$

with $\text{Var}(B_H(\tau)) = \sigma^2 \tau^{2H}$ and $\sigma^2 = \text{Var}(B_H(1)) = s^{2H}$ as in (2). Consequently, the weights w_t are independently distributed from t, i.e.,

$$w_t \sim \mathcal{N}(0, \frac{1}{2}(s\tau)^{2H}). \qquad (35)$$

Considering the distribution of all possible realisations of fBm, we see from the use of the weights w_t from (11) in the computation of WPE that the weights cancel out for a constant delay $\tau \in \mathbb{N}$. With (12) it follows that $\text{WPE}(B_H(t)) = -\ln(1/2)$.

Calculating the frequencies of ordinal patterns for the weighted pooling matrix $P_w^{\tau,d} \in \mathbb{R}^{m \times d!}$ results in an equal distribution with

$$P_{w_{i,(1,0)}}^{\tau,2} = P_{w_{i,(0,1)}}^{\tau,2} = \frac{1}{2m} \tag{36}$$

for all $i = 1, .., m$. By calculating the marginal relative frequencies, the number of variables m is reduced, i.e.,

$$P_{w_{\cdot j}}^{\tau,2} = \frac{1}{2} \tag{37}$$

for $j = \{(0,1),(1,0)\}$. Hence the claim follows. $\qquad\square$

Orders greater than 2. For orders $d = 3$ and $d = 4$, the following theorem can be derived by inspecting corresponding weights of ordinal patterns in combination with results from PPE.

Theorem 6.2 MWPE($\mathbf{B}_H^m(t)$) *of orders $d = 3$ and $d = 4$ are independent of all delays τ but monotonically dependent on the number of variables m and Hurst parameter H of mfBm. In the case $H_i = H_j$ for all i, j, it is independent on the number of variables m.*

Proof. The weights for ordinal patterns of order $d = 3$ are given by

$$w_t = \frac{2}{9} \cdot (x_t^2 - x_t(x_{t-\tau} - x_{t-2\tau}) + x_{t-\tau}^2 - x_{t-\tau}x_{t-2\tau} + x_{t-2\tau}^2). \tag{38}$$

In contrast to $d = 2$, in particular in contrast to (33), due to the lack of symmetry, the weights depend on the time t, and therefore do not cancel and thus influence the value of MWPE. Since the distribution of the ordinal pattern is independent of the weights w_t, the independence of the delay τ is retained. In addition, the multiplication of weights w_t and counts do not affect the monotony that follows from (13) and (14) of PPE. The independence of MWPE from the number of variables m for $H_i = H_j$ for all variables i, j follows with the same line of reasoning as from the end of the proof of Theorem 4.1 when calculating the marginal frequencies.

The weights for ordinal patterns of order $d = 4$ are given by

$$\begin{aligned} w_t \;=\; \frac{1}{16} \cdot \; & (3x_t^2 - 2x_t(x_{t-\tau} + x_{t-2\tau} + x_{t-3\tau}) \\ & + 3x_{t-\tau}^2 - 2x_{t-\tau}(x_{t-2\tau} + x_{t-3\tau}) \\ & + 3x_{t-2\tau}^2 - 2x_{t-2\tau}x_{t-3\tau} + 3x_{t-3\tau}^2). \end{aligned} \tag{39}$$

Using the same reasoning as in the case of $d = 3$ and with (15), the independence of the delay τ, the number of variables m and the monotonic dependence of the Hurst parameter H for MWPE are preserved. $\qquad\square$

For order $d > 4$, no closed formulas for the distribution of ordinal patterns in fBm exist (see Section 2.3). Nevertheless, analogous behaviour can also be observed in this case, which we evaluate in an experimental setting in the next section.

For increasing Hurst parameter $H \to 1$, m(fBm) is more positively correlated, i.e., after an upward jump, a further upward jump is more likely to follow and vice versa. Consequently, it is more likely that the strictly ascending ordinal pattern $(0, 1, 2)$ or the strictly descending ordinal pattern $(2, 1, 0)$ occurs than the other four ordinal patterns $(0, 2, 1)$, $(1, 0, 2)$, $(2, 0, 1)$ and $(1, 2, 0)$. Since for $H \to 1$ and $t \to \infty$ m(fBm) becomes trendier, the weights of $(0, 1, 2)$ and $(2, 1, 0)$ respectively must be larger on average. Thus, by weighting the ordinal patterns of orders $d > 2$, the values of MWPE are expected to decrease faster than the values of PPE.

6.2 Experimental Evaluation

In this experimental evaluation, we investigate the behaviour of MWPE on mfBm in the variation of Hurst parameter H, delay τ, and the number of variables m. The experimental computations are based on the same simulations of mfBms as in Section 4.2.
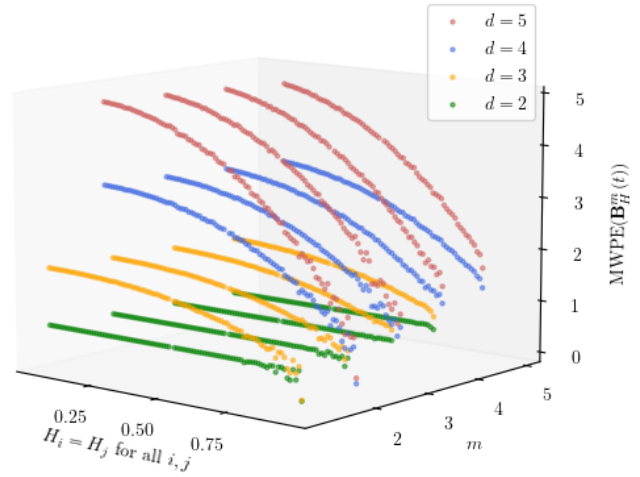
(a) MWPE of fixed delay $\tau = 1$ on mfBm for different numbers of variables $m = 3$.
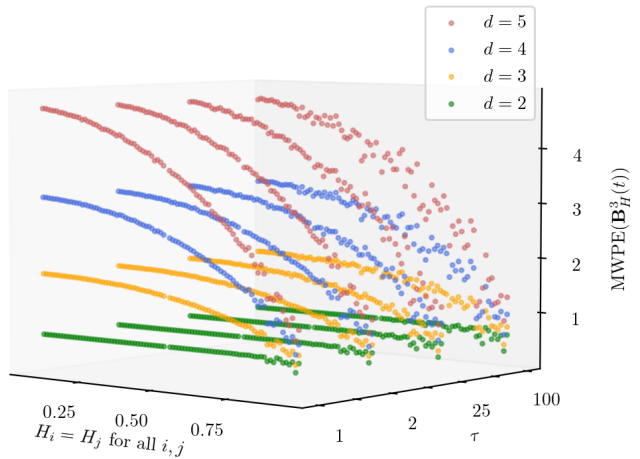
(b) MWPE of different delays τ on mfBm for fixed variable dimension $m = 3$.

Figure 9: Experimental computations of MWPE of orders $d = 2$ (green), $d = 3$ (orange), $d = 4$ (blue), and $d = 5$ (red) on mfBm in variation of its Hurst parameters H.

In Figure 9a we show the independence of MWPE of all orders $d = 2$ (green), $d = 3$ (orange), $d = 4$ (blue), and $d = 5$ (red) on mfBm from number of variables m. For this experiment, the delay is fixed to $\tau = 1$. All green lines, i.e., for $d = 2$, are constant, namely $-\ln(1/2)$, for each m, confirming the independence of numbers of variables m in (31). All other lines, i.e., for $d > 2$, are the same for any number of variables m, but monotonically dependent on the Hurst parameter H. In Figure 9b we show the independence of MWPE on mfBm from delay τ. For this experiment, the number of variables is fixed to $m = 3$. Again, all green lines, i.e., for $d = 2$, are constant with value $-\ln(1/2)$ for each delay τ, confirming the independence of delay τ in (31). All other lines, i.e., for $d > 2$, are

also the same for any delay τ, but monotonically dependent on the Hurst parameter H. The reason for the deviations of the values of MWPE in Figure 9 with increasing Hurst parameter H is the same as in the previous Section 4, i.e., the deviations are caused by length constraint $T < \infty$ of mfBm.

All in all, the experiment underpins our theoretical results statet in Theorem 6.1 and Theorem 6.2. In particular, the behaviour of MWPE in Figure 9 is similar to that of PPE in Figure 7 in terms of the independence of the parameters τ and m, as well as the constant or monotonic progression at orders $d = 2$ or $d > 2$, respectively.

In the theoretical analysis, we discussed the influence of the weighting of ordinal patterns of different orders on the values of MWPE. For $d = 2$, we show that weighting has no influence (Theorem 6.1). For orders $d > 2$, we assume that certain ordinal patterns (strictly ascending and strictly descending) are weighted more than others. Thus the values of MWPE fall faster than the values of PPE as the Hurst parameter H increases. Exemplarily, Table 1 shows the frequencies of (weighted) ordinal patterns (p_{w_j}) p_j of order $d = 3$ on a simulated mfBm of length $T = 10000$ with number of variables $m = 3$, correlation coefficient $\rho = 0.0$, and Hurst parameter $H = (0.8, 0.8, 0.8)$. While for PPE it holds that $w_t = 1$ for every pattern j, we see that the weights w_t of ordinal patterns $(0, 1, 2)$ and $(2, 1, 0)$ on average are larger than those of the other four ordinal patterns. Consequently, the frequencies of these ordinal patterns are also higher than in the unweighted case of PPE. For more details on the difference between MWPE and PPE, see Section 8.

j	counts	w_t	p_{w_j}	p_j
$(0, 1, 2)$	10539	0.76	0.48	0.35
$(0, 2, 1)$	2470	0.19	0.03	0.08
$(1, 0, 2)$	2482	0.19	0.03	0.08
$(1, 2, 0)$	2450	0.19	0.03	0.08
$(2, 0, 1)$	2462	0.19	0.03	0.08
$(2, 1, 0)$	9588	0.72	0.40	0.33

Table 1: Frequencies of (weighted) ordinal patterns (p_{w_j}) p_j of order $d = 3$.

7 MOPPE Applied to mfBm

In this section, for the first time, the behaviour of MOPPE of different orders and delays on mfBm is investigated in the variation of its Hurst parameter, both from a theoretical and experimental point of view.

7.1 Theoretical Analysis

Since multivariate ordinal patterns are defined as univariate ordinal patterns combined in a matrix, joint probabilities are considered to investigate the behaviour of MOPPE on mfBm. Let p_j^τ be a probability for a univariate ordinal pattern j of arbitrary order d in fBm given in Section 2.3. Since any univariate ordinal pattern of any order d is independent of the delay τ, we write p_j in the following.

Order 2. In contrast to the previous sections, the behaviour of MOPPE on mfBm is obviously dependent on the number of variables m but remains independent of the delays τ.

Lemma 7.1 *Let $B_{H_k}^k(s)$ and $B_{H_l}^l(t)$ for every $k, l = 1, \ldots, m$ conditionally independent, then it holds that*

$$\text{MOPPE}_{2,\tau}(\mathbf{B}_H^m(t)) = -\ln\left(\frac{1}{2^m}\right) \tag{40}$$

for all τ.

Proof. The independence from the delay τ follows directly from Definition 3.4 and the distribution of the univariate ordinal patterns of order $d = 2$ in (12). Let $j \in \{(0, 1), (1, 0)\}$ be a univariate ordinal patterns of order $d = 2$. Let $X_k = B_{H_k}^k(t)$ and $X_l = B_{H_l}^l(t)$ for every $k, l = 1, \ldots, m$ conditionally independent, then the joint probability function satisfies

$$P(X_1 = j, \ldots, X_m = j) = P(X_1 = j) \cdot \cdots \cdot P(X_m = j) \tag{41}$$
$$= p_j^m \tag{42}$$

With (12) it is $p_j = 1/2$ for every j, so that the joint distribution of every m-fold combination of ordinal patterns $j \in \{(0, 1), (1, 0)\}$ stays p_j^m. For the number of variables m, there exist 2^m multivariate ordinal patterns as combinations from univariate ordinal patterns so that

$$\text{MOPPE}_{2,\tau}(\mathbf{B}_H^m(t)) = -\sum_{j=1}^{2^m} \frac{1}{2}^m \cdot \ln\left(\frac{1}{2}^m\right). \tag{43}$$

$\square$

Orders greater than 2. For orders $d = 3$ and $d = 4$, the following theorem can be derived by considering joint probabilities of univariate ordinal pattern distributions introduced in Section 2.3.

Lemma 7.2 *MOPPE$(\mathbf{B}_H^m(t))$ of orders $d = 3$ and $d = 4$ are independent of all delays τ, but dependent on number of variables m, and monotonically dependent on the Hurst parameter H.*

Proof. The independence from the delay τ follows directly from Definition 3.4 and the distribution of the univariate ordinal patterns or orders $d = 3$ or $d = 4$ in (13) or (15), respectively. As introduced in Section 2.3, the distribution of ordinal patterns of order $d = 3$ is twofold (see (14)). Therefore, let $A = \{(0, 1, 2), (2, 1, 0)\}$ and $B = \{(0, 2, 1), (1, 2, 0), (1, 0, 2), (2, 0, 1)\}$. Since multivariate ordinal patterns are combinations of univariate ordinal patterns, in the following, we consider three cases where the ordinal patterns of all m variables are either from A, from B or from A and B. Again, let $B_{H_k}^k(s)$ and $B_{H_l}^l(t)$ be conditionally independent for each $k, l = 1, \ldots, m$.

1. Let the univariate pattern a_i of the i-th variable be $a_i \in A$ for all $i = 1, \ldots, m$. Using (13) and (41) the joint distributions of all 2^m combinations of $a_1, \ldots, a_m \in A$ are given by

$$\prod_{i=1}^m \frac{1}{\pi^m} \arcsin 2^{H_i - 1} =: \prod_{i=1}^m \frac{1}{\pi^m} u_i. \tag{44}$$

2. Let the univariate pattern b_i of the i-th variable be $b_i \in B$ for all $i = 1, \ldots, m$. Using (14) and (41) the joint distributions of all 4^m combinations of $b_1, \ldots, b_m \in B$ are given by

$$\prod_{i=1}^m \frac{1}{4}(1 - \frac{1}{\pi^m} u_i). \tag{45}$$
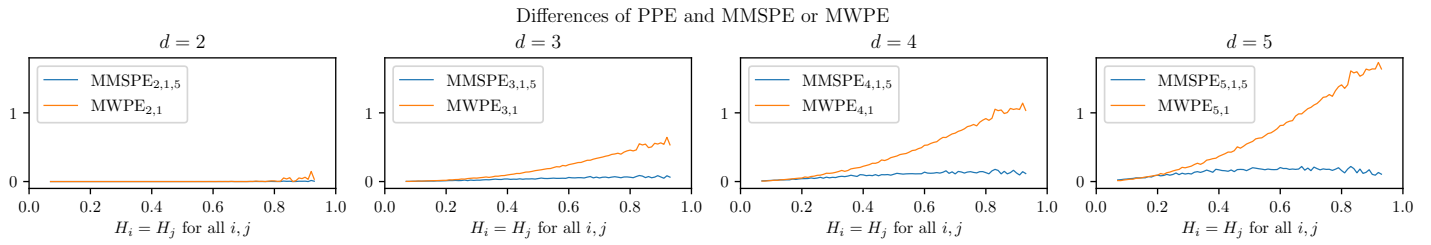
Figure 10: Differences of PPE and MMSPE or MWPE of different orders (from left to right) and fixed delay $\tau = 1$ on simulations of mfBm with fixed number of variables $m = 3$ in variation of Hurst parameters $H_i = H_j$ for all i, j.

3. Let the univariate pattern c_i of the i-th variable be, where $c_1, \ldots, c_k \in A$ and $c_{k+1}, \ldots, c_m \in B$. Then using (13), (14), and (41) the joint distributions of all $6^m - 4^m - 2^m$ remaining combinations $c_1, \ldots, c_m$ are given by

$$\prod_{i=1}^{k} \frac{1}{\pi^k} u_i \cdot \prod_{i=k+1}^{m} \frac{1}{4}(1 - \frac{1}{\pi^{m-k}} u_i). \tag{46}$$

In particular, the joint distributions or the distributions of the multivariate ordinal patterns remain monotonically dependent on the Hurst parameter H after the monotonic transformations.

For order $d = 4$, formulas for joint distributions corresponding to multivariate ordinal patterns are derivable with (15) and (41) in the same way as for order $d = 3$. □

For orders $d > 4$, no closed formulas for the distribution of ordinal patterns in fBm exist. Nevertheless, analogous behaviour of MOPPE of orders $d > 4$ can also be observed in this case, which we evaluate in an experimental setting in the next section.

7.2 Experimental Evaluation

In this experimental evaluation, we investigate the behaviour of MOPPE on mfBm in the variation of the Hurst parameter H and delay τ. Since the number of multivariate ordinal patterns depends on the number of variables m and thus also on MOPPE, we refrain from an experimental evaluation of different m, i.e., we assume $m = 2$ small for computational reasons. The experimental computations are based on the same simulations of mfBms as in Section 4.2 with the difference that in this experiment, we include more than one correlation parameter as MOPPE takes into account the interdependence of variables.

In Figure 11a we show the independence of MOPPE on mfBm from delay τ. For this, the number of variables is fixed to $m = 2$. Again, all green lines, i.e., for $d = 2$, are constant $-\ln(1/4)$ for each τ, confirming the independence of delay τ in (40). All other lines, i.e., for $d > 2$, are also the same for any τ, but monotonically dependent on the Hurst parameter H, i.e., for increasing Hurst parameter H the value of MOPPE decreases. The reason for the deviations of the values of MOPPE in Figure 11 with increasing Hurst parameter H is the same as in the previous Section 4. All in all, the experiment underpins our theoretical results from Theorem 7.1 and Theorem 7.2.
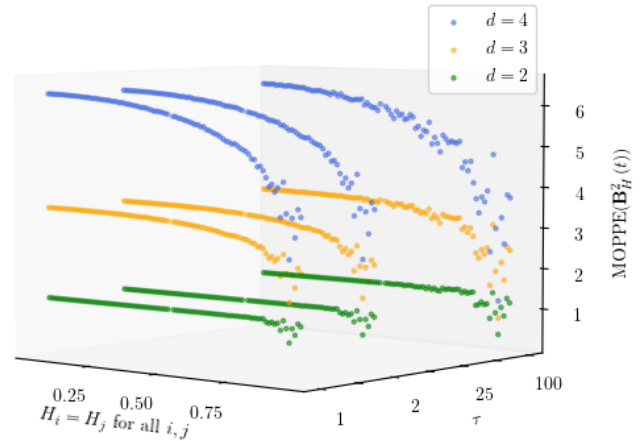
(a) MOPPE of different delays τ on mfBm with fixed number of variables $m = 2$.
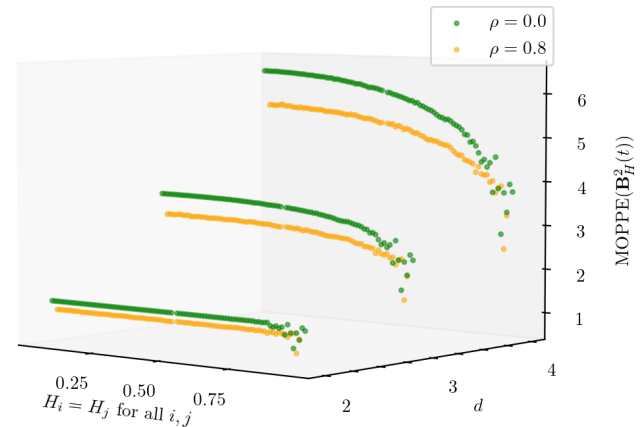
(b) MOPPE of fixed delay $\tau = 1$ on mfBm with fixed number of variables $m = 2$.

Figure 11: Experimental computations of MOPPE for different delays τ on mfBm with different correlations $\rho = \{0.0, 0.8\}$ in variation of its Hurst parameters H.

Although we have focused on independent variables in the theoretical part for simplicity as well as due to the lack of results in the literature, we examine the dependence on the correlations of variables in space in Figure 11b, since the definition of multivariate ordinal patterns involves the dependence on several variables. A high cross-covariance or correlation of the variables causes the behaviour of the individual variables (for different Hurst parameters) to converge. Another influences the behaviour of one variable, and so is the occurrence of the higher-order ordinal patterns that are dependent on the Hurst parameter. For the investigation of correlations using MOPPE, the delay or the number of variables are fixed to $\tau = 1$ or $m = 2$, respectively. In Figure 11b, the yellow

lines, corresponding to a correlation coefficient of 0.8, are constantly below the green lines corresponding to a correlation coefficient of 0.0. This reflects the fact that due to a higher correlation of the variables, there is a greater fit of the individual univariate ordinal pattern order in the multivariate combination, resulting in a smaller MOPPE value. Advantages are discussed in Section 8.

Figure 12: Differences of MOPPE values on mfBm with different correlations $\rho_{ij} = \{0.0, 0.8\}$ in variation of its Hurst parameter H.

8 Comparison

Up to this point, we have examined the behaviour of numerous multivariate representations, namely pooled permutation entropy (PPE), multivariate multi-scale permutation entropy (MMSPE), multivariate weighted permutation entropy (MWPE), and multivariate ordinal pattern permutation entropy (MOPPE). In this section, we summarise both the differences and the similarities between the various representations. This leads to different possible applications as well as to different recommendations.

PPE can be understood as a canonical extension of univariate ordinal patterns to a multivariate variant that pools the univariate ordinal patterns over the variables. MMSPE and MWPE belong to the same family as PPE. They are all based on a pooled matrix while addressing different aspects of the time series and ordinal patterns, namely scaling and amplitudes. One advantage for these approaches is that, in contrast to MOPPE, they also allow univariate analyses on the individual variables within the algorithm. Figure 10 compares PPE with its extensions MMSPE and MWPE with different orders $d = 1, \ldots, 5$ from left to right. Since, as evaluated in the individual sections, all measures are independent of the delay and the number of variables, we set $\tau = 1$ and $m = 3$. Similarly, MMSPE is independent of the scale parameter ϵ, so we set the scale factor $\epsilon = 5$. The left sub-figure shows the equality of PPE, MMSPE and MWPE for order $d = 2$, which again underpin (26), (30) and (31). Thus, scaling or weighting has no effect. The three right sub-figures in Figure 10 show the equality of PPE and MMSPE (blue lines) for orders $d = 3, 4, 5$. Similarly, they show a difference between PPE and MWPE as the orange lines increase in variation of increasing H, indicating a faster decrease of MWPE.

MOPPE can be understood as a canonical extension of univariate ordinal patterns to multivariate ordinal patterns by conceiving univariate patterns as multidimensional patterns in a matrix. The advantage of this approach lies in consideration of the interdependencies of several variables at a single time point. A disadvantage is the exponentially increasing number of possible ordinal patterns, which increase the complexity of the computation and result in a uniform distribution and thus maximum entropy for $T \ll \infty$. Furthermore, MOPPE considers the multivariate time series as a whole, i.e., not each variable individually, which means that an analysis of the individual variables using MOPPE within the algorithm is not possible, in contrast to PPE.
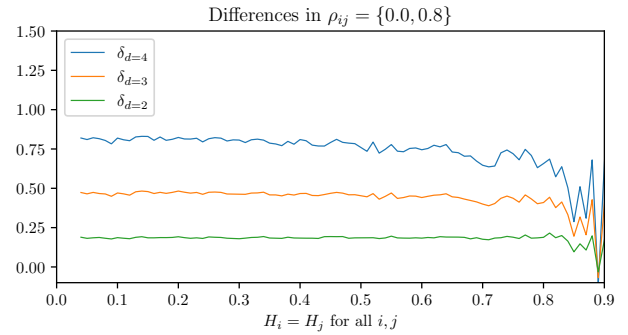
Figure 12 visualises the differences of MOPPE values of different orders $d = 2$ (green), $d = 3$ (orange), and $d = 4$ (blue) and fixed delay τ on mfBm with fixed variable dimension $m = 2$ and different correlations $\rho_{ij} = \{0.0, 0.8\}$ in variation of its Hurst parameter $H = (H_1, H_2)$, where $H_1 = H_2$. As the visualisation shows, the differences for any $d \in \{2, 3, 4\}$ are almost constant for all Hurst parameters H. Increasing differences for increasing order d result from increasing maximum entropy values per order d.

The differences in the multivariate extensions of PE are important because they may be relevant to the use of different applications. In the field of inverse problems, the aim is to estimate the parameters, e.g., the Hurst parameter, of a generating dynamic system from observed realisations [36]. As all multivariate variants for orders $d > 2$ involve a direct relationship between the distribution of univariate ordinal patterns and the Hurst parameter H, PPE, MWPE and MOPPE are suitable candidates to solve the inverse problem regarding H. Given a realisation of mfBm, for example, the calculation of entropies that decrease as H increases can provide conclusions about the value of H. In addition to estimating H, MOPPE has the potential to uncover additional correlations by taking interdependencies into account, and it does so from one source. However, MOPPE is only practical for a small number of variables m. In machine learning, a high accuracy of the prediction depends not only on the model used but also on the representations of the data. Therefore, extracting good or expressive representations, also called features, from the data is essential. Especially in classification tasks, it is necessary to extract features that separate well [37]. In this case, MWPE may promise advantages over PPE because MWPE favours certain ordinal patterns (strictly rising and strictly falling) and thus decreases faster compared to PPE.

9 Conclusion and Future Work

Fractional Brownian motion (fBm) is usually used for modelling real-world applications with specific properties such as long-range dependence or self-similarity. Since many real-world applications also contain multivariate interdependencies between several variables, we focus on multivariate fractional Brownian motion (mfBm). This paper provides a comprehensive study of several multivariate measures that study the qualitative behaviour of multivariate fractional Brownian motion (mfBm), namely pooled permutation entropy (PPE), multivariate multi-scale permutation entropy

(MMSPE), multivariate weighted permutation entropy (MWPE), and multivariate ordinal pattern permutation entropy (MOPPE). All measures are understood as multivariate extensions of univariate permutation entropy (PE), which is based on the distribution of ordinal patterns, i.e., on the ups and downs of different orders $d \in \mathbb{N}_{>1}$. The main contributions of this paper are the extensions of the theoretical and experimental investigations of PPE, MMSPE and MWPE on mfBm in the original papers [1, 2] to additional orders $d = 4$ and $d = 5$. Furthermore, this paper provides a new study on the behaviour of MOPPE on mfBm, and for the first time, a comparison of all multivariate extensions considered.

Our theoretical and experimental analyses show the following results on the behaviour of PPE, MMSPE, MWPE and MOPPE on mfBm. For order $d = 2$, for all four cases the multivariate permutation entropy is constant $-\ln(1/2)$ regardless of number of variables, Hurst parameters, delays, scales, or weights (or $-\ln(1/2^m)$ in case of MOPPE). Because all measures are equal, in particular constant and independent of all parameters, no characteristics can be determined via these measures with order $d = 2$. Therefore, usage is not reasonable in applications. Since scaling does not change the structure of mfBm, MMSPE of any scale ϵ is equal to PPE and analysis with MMSPE of higher orders $d > 2$ does not provide any additional insight than PPE.

However, for orders $d > 2$, the use of PPE, MWPE and MOPPE provide interesting insights and possible applications. The distribution of ordinal patterns, and thus also PPE, MWPE and MOPPE, are directly related to the Hurst parameter H. For example, considering the estimation of H is an inverse problem, PPE, MWPE and MOPPE can be used since they all depend monotonically on H, i.e., the entropy decreases as H increases. In contrast to PPE and MOPPE, MWPE is substantially influenced by strictly ascending and descending ordinal patterns. For this reason, MWPE decreases faster than PPE as H increases, providing more expressive representations that may promise better discriminability. Since multivariate ordinal patterns involve the dependence of several variables on mfBm, MOPPE is promising for estimating the Hurst parameter H and correlations from one hand, which the other measures cannot. To prove the exact theoretical relationship, conditional probabilities of univariate ordinal patterns must first be derived.

In the literature, there are additional approaches to multivariate permutation entropy determination, specifically Multivariate Permutation Entropy (MvPE), introduced in [30], or multivariate permutation entropy based on principal component analysis (MPE-PCA), introduced in [33]. Since their application requires a prior transformation of mfBm and thus is not a canonical extension of the univariate definition, its study is not part of this work. Nevertheless, both have the potential to uncover (additional) structures in mfBm but are left for future work. In addition, the theoretical results of the measures investigated in this work motivate various potential applications, e.g., for estimating the Hurst parameter or as a feature in a learning task in the field of machine learning. Similarly, the application of these measures is of great interest not only in the context of mfBm but also for any real-world data set. Since this work focuses on the theoretic contexts, we leave concrete applications for future work.

References

[1] M. Mohr, N. Finke, R. Möller, "On the Behaviour of Permutation Entropy on Fractional Brownian Motion in a Multivariate Setting," in Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2020 (APSIPA-ASC), 2020.

[2] M. Mohr, F. Wilhelm, R. Möller, "On the Behaviour of Weighted Permutation Entropy on Fractional Brownian Motion in the Univariate and Multivariate Setting," The International FLAIRS Conference Proceedings, **34**, 2021, doi: 10.32473/flairs.v34i1.128464.

[3] T. Konstantopoulos, S.-J. Lin, "Fractional Brownian Approximations of Queueing Networks," in P. Glasserman, K. Sigman, D. D. Yao, editors, Stochastic Networks, 257–273, Springer New York, New York, NY, 1996.

[4] R. J. Elliott, J. van der Hoek, "Fractional Brownian Motion and Financial Modelling," in M. Kohlmann, S. Tang, editors, Mathematical Finance, 140–151, Birkhäuser Basel, 2001.

[5] B. S. Nath, B. Mousumi, "Long Memory in Stock Returns: A Study of Emerging Markets," Iranian Journal of Management Studies, **5**(2), 67–88, 2012.

[6] F. Traversaro, F. O. Redelico, M. R. Risk, A. C. Frery, O. A. Rosso, "Bandt-Pompe symbolization dynamics for time series with tied values: A data-driven approach," Chaos: An Interdisciplinary Journal of Nonlinear Science, **28**(7), 2018, doi:10.1063/1.5022021.

[7] J. Amigó, Permutation Complexity in Dynamical Systems: Ordinal Patterns, Permutation Entropy and All That, Springer Series in Synergetics, Springer-Verlag, Berlin Heidelberg, 2010.

[8] C. Bandt, F. Shiha, "Order Patterns in Time Series," Journal of Time Series Analysis, **28**(5), 646–665, 2007.

[9] F. C. Morabito, D. Labate, F. La Foresta, A. Bramanti, G. Morabito, I. Palamara, "Multivariate Multi-Scale Permutation Entropy for Complexity Analysis of Alzheimer's Disease EEG," Entropy, **14**(7), 1186–1202, 2012, doi: 10.3390/e14071186.

[10] B. Fadlallah, B. Chen, A. Keil, J. Principe, "Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information," Physical review. E, Statistical, nonlinear, and soft matter physics, **87**, 022911, 2013, doi:10.1103/PhysRevE.87.022911.

[11] M. Sinn, K. Keller, "Estimation of Ordinal Pattern Probabilities in Gaussian Processes with Stationary Increments," Comput. Stat. Data Anal., **55**(4), 1781–1790, 2011, doi:10.1016/j.csda.2010.11.009.

[12] L. Zunino, D. G. Pérez, M. T. Martín, M. Garavaglia, A. Plastino, O. A. Rosso, "Permutation entropy of fractional Brownian motion and fractional Gaussian noise," Physics Letters A, **372**(27), 4768–4774, 2008.

[13] A. Dávalos, M. Jabloun, P. Ravier, O. Buttelli, "Theoretical Study of Multiscale Permutation Entropy on Finite-Length Fractional Gaussian Noise," in 2018 26th European Signal Processing Conference (EUSIPCO), 1087–1091, 2018, doi:10.23919/EUSIPCO.2018.8553070, iSSN: 2076-1465.

[14] L. A. Gil-Alana, "A fractional multivariate long memory model for the US and the Canadian real output," Economics Letters, **81**(3), 355–359, 2003, doi:10.1016/S0165-1765(03)00217-9.

[15] S. Achard, D. S. Bassett, A. Meyer-Lindenberg, E. Bullmore, "Fractal connectivity of long-memory networks," Physical Review E, **77**(3), 2008.

[16] B. B. Mandelbrot, J. R. Wallis, "Noah, Joseph, and Operational Hydrology," Water Resources Research, **4**(5), 909–918, 1968, doi:10.1029/WR004i005p00909.

[17] G. Didier, V. Pipiras, "Integral representations and properties of operator fractional Brownian motions," Bernoulli, **17**(1), 1–33, 2011, doi:10.3150/10-BEJ259.

[18] F. Lavancier, A. Philippe, D. Surgailis, "Covariance function of vector self-similar processes," Statistics & Probability Letters, **79**(23), 2415–2421, 2009, doi:10.1016/j.spl.2009.08.015.

[19] G. Didier, V. Pipiras, "Integral representations and properties of operator fractional Brownian motions," Bernoulli, **17**(1), 1–33, 2011.

[20] P.-O. Amblard, J.-F. Coeurjolly, F. Lavancier, A. Philippe, "Basic properties of the Multivariate Fractional Brownian Motion," Séminaires et congrès, **28**, 65–87, 2013.

[21] P.-O. Amblard, J.-F. Coeurjolly, "Identification of the Multivariate Fractional Brownian Motion," IEEE Transactions on Signal Processing, **59**(11), 5152–5168, 2011, doi:10.1109/TSP.2011.2162835.

[22] B. Chiu, E. Keogh, S. Lonardi, "Probabilistic Discovery of Time Series Motifs," in Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '03, 493–498, ACM, 2003.

[23] C. Bandt, B. Pompe, "Permutation Entropy: A Natural Complexity Measure for Time Series," Physical Review Letters, **88**(17), 174102, 2002, publisher: American Physical Society.

[24] K. Keller, T. Mangold, I. Stolz, J. Werner, "Permutation Entropy: New Ideas and Challenges," Entropy, **19**(3), 134, 2017, doi:10.3390/e19030134.

[25] A. B. Piek, I. Stolz, K. Keller, "Algorithmics, Possibilities and Limits of Ordinal Pattern Based Entropies," Entropy, **21**(6), 547, 2019, doi:10.3390/e21060547.

[26] K. Keller, S. Maksymenko, I. Stolz, "Entropy determination based on the ordinal structure of a dynamical system," Discrete and Continuous Dynamical Systems - Series B, **20**(10), 3507–3524, 2015, doi:10.3934/dcdsb.2015.20.3507.

[27] I. Stolz, K. Keller, "A General Symbolic Approach to Kolmogorov-Sinai Entropy," Entropy, **19**(12), 675, 2017, doi:10.3390/e19120675.

[28] M. Costa, A. L. Goldberger, C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," Physical Review Letters, **89**(6), 068102, 2002.

[29] K. Keller, H. Lauffer, "Symbolic Analysis of High-Dimensional Time Series," International Journal of Bifurcation and Chaos, **13**(09), 2657–2668, 2003, doi:10.1142/S0218127403008168.

[30] S. He, K. Sun, H. Wang, "Multivariate Permutation Entropy and Its Application for Complexity Analysis of Chaotic Systems," Physica A: Statistical Mechanics and its Applications, **461**, 812–823, 2016, doi:10.1016/j.physa.2016.06.012.

[31] K. Keller, "Permutations and the Kolmogorov-Sinai Entropy," Discrete & Continuous Dynamical Systems - A, **32**(3), 891, 2012, doi:10.3934/dcds.2012.32.891.

[32] A. Antoniouk, K. Keller, S. Maksymenko, "Kolmogorov-Sinai Entropy via Separation Properties of Order-Generated Sigma-Algebras," Discrete and Continuous Dynamical Systems, **34**(5), 1793–1809, 2013, doi:10.3934/dcds.2014.34.1793.

[33] M. Mohr, F. Wilhelm, M. Hartwig, R. Möller, K. Keller, "New Approaches in Ordinal Pattern Representations for Multivariate Time Series," in Proceedings of the 33rd International Florida Artificial Intelligence Research Society Conference (FLAIRS-33), 124–129, AAAI Press, 2020.

[34] Y. Rayan, Y. Mohammad, S. A. Ali, "Multidimensional Permutation Entropy for Constrained Motif Discovery," in N. T. Nguyen, F. L. Gaol, T.-P. Hong, B. Trawiński, editors, Intelligent Information and Database Systems, Lecture Notes in Computer Science, 231–243, Springer International Publishing, Cham, 2019, doi:10.1007/978-3-030-14799-0_20.

[35] N. Nicolaou, J. Georgiou, "The Use of Permutation Entropy to Characterize Sleep Electroencephalograms," Clinical EEG and Neuroscience, **42**(1), 24–28, 2011, doi:10.1177/155005941104200107.

[36] B. D'Ambrogi-Ola, "Inverse Problem for Fractional Brownian Motion with Discrete Data," 2009.

[37] Y. Bengio, A. Courville, P. Vincent, "Representation Learning: A Review and New Perspectives," IEEE Transactions on Pattern Analysis and Machine Intelligence, **35**(8), 1798–1828, 2013, doi:10.1109/TPAMI.2013.50.

A S T E S

Machine Learning Algorithms for Real Time Blind Audio Source Separation with Natural Language Detection

Arwa Alghamdi[1], Graham Healy[2], Hoda Abdelhafez[*,1,3]

[1]*College of Computer & Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, 12211, Saudi Arabia*

[2]*School of Computing, Dublin City University, Dublin, D09 NA55, Ireland*

[3]*Faculty of Computer & Informatics, Suez Canal University, Ismailia, 41611, Egypt*

A R T I C L E   I N F O

A B S T R A C T

The Conv-TasNet and Demucs algorithms, can differentiate between two mixed signals, such as music and speech, the mixing operation proceed without any support information. The network of convolutional time-domain audio separations is used in Conv-TasNet algorithm, while there is a new waveform-to-waveform model in Demucs algorithm. The Demucs algorithm utilizes a procedure like the audio generation model and sizable decoder capacity. The algorithms are not pretrained; so, the process of separation is blindly without any function of three Natural Languages (NL) detection. This research study evaluated the quality and execution time of the separation output signals. It focused on studying the effectiveness of NL in Both algorithms based on four sound signal experiments: (music & male), (music &female), (music & conversation), and finally (music & child). In addition, this research studies three NL, which are English, Arabic and Chinese. The results are evaluated based on R square and mir_eval libraries, mean absolute Error (MAE) scores and root mean square error (RMSE). Conv-TasNet has the highest Signal-to-distortion-Ratio (SDR) score is 9.21 of music at (music & female) experiment, and the highest SDR value of child signal is 8.14. The SDR score of music at (music & female) experiment is 7.8 during the Demucs algorithm, whereas child output signal has the highest SDR score 8.15. However, the average execution time of English experiment of Conv-TasNet is seven times faster than Demucs. For accuracy measure, RMSE indicates absolute values, and MAE handles the errors between observations and prediction signals. Both algorithms show high accuracy and excellent results in the separation process.

1. Introduction

This paper is an extension of work originally presented in conference Novel Intelligent and Leading Emerging Sciences Conference (NILES) [1]. As everybody knows that a human is intelligent and has an aural source separation that helps to identify singular sounds within a mixed sound. The type of mixed sound could be a speech or music, which can be detected via the human aural system; because the brain nerves are trained since birth to differentiate between human speech and music. Artificial intelligence, machine learning algorithms, and other Technological advancements have contributed to the revolution of machines to simulate the human intelligence.

Machine learning algorithms contributed to the development in multiple fields such as signal processing. An approach called Blind Source Separation (BSS) can separate unknown set of source signals and the mixing methodology. Furthermore, BSS recovers all individual sources simultaneously as output from the mixture, which helps in performing a real-time technique. Thus, BSS can process sounds in a real-time using many algorithms such as independent component analysis (ICA), Principal Component Analysis (PCA), Degenerate Unmixing Estimation Technique (DUET), Short Time Fourier Transform (STFT) [2], [3]. Some studies applied this methodology to improve other algorithms such as the Deep Neural Network (DNN) algorithm [4], the Independent Deeply Learned Matrix Analysis (IDLMA)

algorithm [5], the Minimum-Volume Beta-Divergence (NMF) algorithm [6], and others.

The real-time approach can be applied simultaneously when using BSS. This enhances separation by using a synchronization procedure during processing. One of the most significant evaluation criteria for comparison is real-time processing because it prefers the algorithm that has a minimum execution time during the separation process. The execution time value computes the actual time through the separation process.

This paper compares the efficiency of Conv-TasNet and Demucs algorithms which use the BSS approach to separate a mixture of two sounds (human speech and music). The biggest challenges are training and testing the algorithms to distinguish between original sounds with high-quality output signals in real-time, with a small execution time value. Both algorithms implement a synchronization method, which substantially reduces the execution time and applies a multi-channel input audio source as well.

2. Related Work

In [7], the authors applied time-domain methodology with the BSS approach. They presented a real-time approach using a Deep Learning System (DLS), which is a robust speech processing during multi-talker environments. Moreover, the authors used signals on nonnegative encoder outputs through the encoder-decoder framework that implemented a model called Time-domain Audio separation Network (TasNet) [8]. The Wall Street Journal (WSJ0-mix) dataset was used to study speech separation problems using two speakers. The dataset was separated into testing and training sets randomly produced mixture. The evaluation appeared that the DLS was 0.23 milliseconds which is six times faster than the state-of-art based system [9]. Nevertheless, the DLS system is not applicable in this study, because DLS supports a short latency like telecommunication applications and listening devices. Furthermore, the experiment is limited only on speech for robust speech processing environment, which represents different object in this study.

Another study implemented Degenerate Unmixing Estimation Technique (DUET) with the Levenberg-Marquardt algorithm for improving the synchronization and signal ratio [10]. This study aimed to build a new DUET with a real-time application while enhancing the quality of the separation process. They used the Bach10 audio dataset [11], which has ten excerpts of classical music, and the single tracks are obtainable. The random mixing with distinct parameters was used to compute the simulated stereophonic mixtures. Using Signal-to-Interference Ratio (SIR), according to the evaluation, a higher SIR means a higher separation of signals quality where they found that SIR was 0.06. Improving the disjoint orthogonality between prediction and observations sources using a recursive implementation in Levenberg-Marquardt algorithm. This result added value to our research study; in terms of enhancing the quality of separation of the BSS in real-time.

In [12], the authors proposed Time-Frequency (TF) method with BSS by estimating the quantity of TF masks and sources with a current spatial aliasing problem. They used DUET and Observation Vector Clustering (OVC). Also, they developed offline and online approach to deal with spatial aliasing. The online separation algorithm operated on gradient ascent and single frame basis. In addition, it optimized using on gradient ascent search with fixed known of the maximum quantity of sources N. The audio dataset was the (CHAINS corpus) story [13]; the authors chose the six-voice signals of the story. Their experiments were implemented simulating a cocktail party situation. Higher distance-miking is the best for the separation performance. The online algorithm was able to track moving sources by estimating of a fixed known of the maximum quantity of sources. The authors mentioned that in mildly reverberant environments, both offline and online algorithms show a good separation performance. However, they do not mention the processing speed of online or offline algorithms.

BSS approaches can be implemented in real-time on hearing aid systems [14]. It is possible to extract and reproduce the target component from mixed binaural sounds using a pocket-size real-time BSS module. The author compared Single-Input Multiple-Output (SIMO) of ICA module with SIMO-of binary masking on convolutive mixtures of speech. Then, produce binaural sounds from the decomposition of mixed binaural sounds. Their experiment evaluated both objectively and subjectively, using 9 users (8 male and 1 female). The distance between the sound source and user was 1.5 m. Furthermore, the speech sample was limited to 9 s and the sampling frequency was 8 kHz. The performance of the modules was comparatively high because of ICA value estimation. Besides, noise reduction rating NRR proposed as a better method rather than conventional methods. However, the authors did not mention more details about their datasets.

In [15], the authors used DUET algorithm of BSS in their research study. They proposed random number of sources while using only two mixtures. Tracking mixing parameters via an online algorithm was applied in a Maximum Likelihood based gradient search method. They generated a mixture dataset using different noises and voices, and they recorded varying angles in echoic and anechoic rooms. Anechoic and echoic rooms of 1500 tested mixtures, anechoic and echoic were 15 dB and 5 dB respectively, of office mixtures. The evaluation environment was a 750 MHz laptop computer, which run more than five times faster than real-time. They noticed that the separation results improved as the angle difference increased. The dataset of authors is completely different from our paper.

The Support Vector Machine (SVM) algorithm is also the solution of BSS under-determined convolutive problem. A study in [16], the authors applied an audio signals sparsity property as an alternative to the independence or stationery of the sources' expectation. Their test was based on the vocal of male and female. They recorded the audio set in a room using microphones with

0.2 seconds of reverberation time. The rate of sampling was 16KHz, and an interval of recording was 28 seconds. The authors presented visualization of signals, but without including a consistent description. The use of the SVM algorithm proved that the BSS could be enhancing via classification using the sparsity property during the separation procedure.

In [17]. the authors proposed optimization approaches about (DUET) algorithm. They developed and implemented DUET for distant speech recognition engine. The authors used two standard techniques for BSS real time: omni-directional microphones and soundcard at reverberation environments. Also, they had several speakers and noise sources, such as a TV. Their experiment focused on two human speakers in a room. They studied the Maximum Microphone Distance (MMD) and became MMD >1m or 2m, but this was a negative effect on the separation quality. However, there are some factors should be considered in a real-time approach when there are multiple sources of a streaming algorithm using position tracking, correlation, arbitrary statistics, condensed metrics. In addition to peak classier processed in 0.55s of execution time and running audio was 9.7s through a single thread with a five-intel core processor. They applied a huge and complex operations to come up with the result. Those techniques will be useful for creating new algorithms which has BSS in a real-time.

On the other hand, in [18], the authors introduced a BSS in real-time oriented to voice separation. They used low-power technique of a processor and synchronized signal for enhancing a real-time methodology. they focused on a cocktail party problem. The sampling rate was 8KHz and the size of data was 16 bit/sample. The audio sets were music, drum, male voice, and more recorded in an audio studio with a professional equipment environment. The authors realized that the algorithm was unable to increase the separation performance using a real-time approach; because of the high quality of separation procedure which requires five times level up iteration. Thus, their research study would be helpful in medical aspects. On the other hand, this research could be implemented in many signals field, mainly when there is a need of a real-time separation.

Demucs implements through following the methodology of WAVE-U-NET algorithm with some disagreements. WAVE-U-NET focuses on manipulating the feature on different period intervals for resampling feature maps frequently. Furthermore, Demucs convert the mixture audio signal into individual original audio signals. Also, Demucs has a technique which is a transferred convolution, that it requires low cost of computation and low memory [19] [20]. Conv-TasNet is a state-of-the-art approach for the separation procedure of speech and music, which helps waveform domain to predict a mask and to improve the whole spectrogram. In [20], the authors speed up the Conv-TasNet during source separation to enhance the execution time which supports a real-time methodology [21].

This paper is an extension of work [1] to study the effectiveness of the Natural Language on the separation process of Demucs and Conv-TasNet algorithms. The experiments' inputs were English, Arabic and Chinese Audioset. This is for studding the ability of the algorithm to separate speech from music blindly, without any previous known of the language type.

In [22], the authors presented the graph-clustering algorithm of Chinese Whispers. The evaluation of their research focused on Natural Language Processing (NLP) issues such as language separation, acquisition of word sense disambiguation, and syntactic word classes. This concludes to employee many graphs in NLP to optimize the separation.

In [23], the authors introduced continuous Arabic speech segmentation. They aimed to use high segmentation accuracies in Concatenative Speech Synthesis and Continuous Automatic Speech Recognition Systems. A tonal language such as Arabic has multiple allophonic phonemes; that will increase the segments computational cost. There are several factors such as ambiguous boundaries between phonemes and allophonic variations, makes recognition more challenging and affect the speech segmentation accuracy. FFT algorithm uses two main methodology frequency and time domains. The evaluation signals are recorded in a constricted environment. The audio sets were Ayat from Qur'an with a duration time 45s. Authors selected 10 audio speeches of 10 readers recorded with no interfering sounds such as noise. The samples file are 128 samples with mono- channel, bit rate 128kbps, type was .wav, audio format is PCM and audio sample size are 16-bit. In segmenting, they had a high segmentation result of intensity value and continuous signals. In the Arabic vowels, the process of the algorithm was accurate at segmentation. This research cannot be compared with the our study results because the differences in algorithms type and propose. The dataset of the experiment is a Small World graphs (SW-graph) [24]. PCW algorithm can process a huge number of graphs in a standard time, however its nondeterministic produces indecisive results.

3. Research Methodology

In this section, we focus on two algorithms for signal separation blindly on a real-time that were used in this research. These algorithms are the Demucs and Conv-TasNet algorithms to compare. The two algorithms are supported and updated in year 2020 by Alexandre Défossez. Moreover, both algorithms are efficient at computational cost and memory process. They are comparable; because they are based on state-of-the-art approaches [20], However, they have an MIT license that can be used and edited. Both Conv-TasNet and Demucs employ machine-learning algorithms to build the models. Also, discuss the evaluation method via the performance of the algorithm and the input and output signals. This research study was extended [1] to apply the Natural Language (NL) of three different languages English, Arabic and Chinese.

TasNet time-domain audio separation network reproduced with the convolution evaluation and encoder-decoder convolutional architecture to come up with Conv-TasNet. The encoder-decoder architecture is like an ICA algorithm operation. The input-signals of the algorithm are mixed sound. The encoder

builds a convolutional block with a particular kernel amount. In addition, the output-signals are a decoders' parameter produced in a non-negativity constraint form. Thus, the computational cost has a significant effect using long short-term memory (LSTM) network in the separation module process. This produces a limitation while training the original TasNet algorithm; because of determining the shorter length of the segment's waveform. Furthermore, LSTM requires a large quantity of parameters which has an immediate effect on expanding the cost of the operational procedure [21].

Conv-TasNet uses entirely time-domain convolution network for sound separation. It contains three stages of memory process: encoding, manipulating and decoding. At the first stage, the encoder module encodes the input signal into overlapping segments with a length (L). The selection depends upon the shorter L of a segment. The output of encoder becomes an input to the manipulating or separation stage, for estimating the mask of a waveform segments. The last stage is decoder, to transform the masks. It reverses the output of separation stage linearly to formulate the number of blocks. Figure 1 shows mathematically the Conv-TasNet process.

Input
Mixture waveform matrix X

Encoder
Divides (X) into overlapping segments of fixed temporal length (L) to transform them into an N-dimensional (1-D convolution operation), using the nonlinear function: $w = H(xU)$

Estimation
Estimates the C vectors and the masks mi of the speaker's number in the mixture matrix. Then, multiply the encoder output (w) via the element-wise multiplication function: $d_i = w \odot m_i$

Decoder
Reconstructs (X), and the rows in V are the basis functions of the decoder with the length (L): $\hat{x} = wV$

Each source has a waveform where i = 1,2 ….. C. to be reconstructed by encoder: $\hat{s}_i = d_i V$

Output
Separated speakers (estimation from mixture waveform matrix (X))

Figure 1: The Stages of Conv-TasNet Algorithm

Conv-TasNet algorithm is the evolution of TasNet algorithm, with application of entirely time-domain convolution network. As shown in figure 1, with the addition of the Temporal Convolutional Network (TCN) methodology. TCN assists in reducing the complexity of the model and the capacity and also reducing the computational cost. It can be applied with or without

a real-time approach, minimize the execution time, and the precision of a separation process. In view of all these reasons, Conv-TasNet is suitable for our study. However, the algorithm has some limitation, it is incapable to denoise the signal or remove the artifacts of a signal such as reverberation or vibrations. Nevertheless, using a multi-channel could solve this problem. Additional limitation is that the constant temporal length can failure with lengthy tracking. Accordingly, this algorithm needs to develop in many aspects, such as accuracy and execution time.

Basic approach of Conv-TasNet algorithm is (1-D convolution separation model):

$$D - conv(Y, K) = concat(y_j \circledast k_j), j = 1, \dots \dots, N$$

$$S - conv(Y, K, L) = D - conv(Y, K) \circledast L$$

Algorithm Notation:

Y: is the input matrix
K: is the convolution kernel
P: is the kernel size
y, and k: are the rows on matrix Y and K, respectively.
L: is the convolution kernel with size 1.
$\circledast$: is the convolution operation.
D-conv(.): each row of the input Y convolves with the corresponding row of matrix K.
S-conv(.): standard convolution which is a separable convolution.

Figure 2 clarify how the separation process of Conv-TasNet algorithm is in signals form. It shows the separation results on a signal form of Conv-TasNet of (Music &Female) category
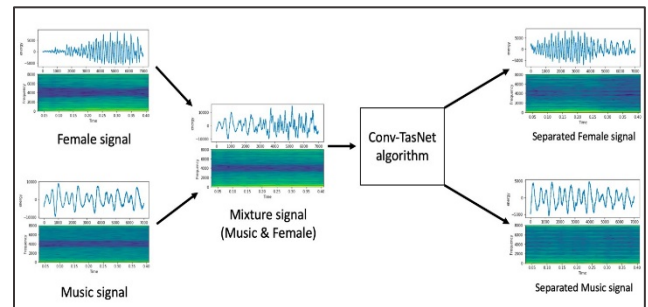
Figure 2: Conv-TasNet Algorithm

On the other hand, Demucs algorithm applies the convolution models, which are encoders, decoders, and LSTM models. Demucs is based on a U-network structure which is identical to the Wave-U-Net structure with differences. However, U-network performs a skip connection between the models, encoder, and decoder to supply a straight connection to the initial signal. Although, U-network does not use linear interpolation but uses a transposed convolution. That helps reducing computational cost and the size of the memory, which enhances the utilization of memory and process four times less in general. Nevertheless, running out of memory is the limitation of the Demucs multi-channel. Figure 3 shows the Demucs process.
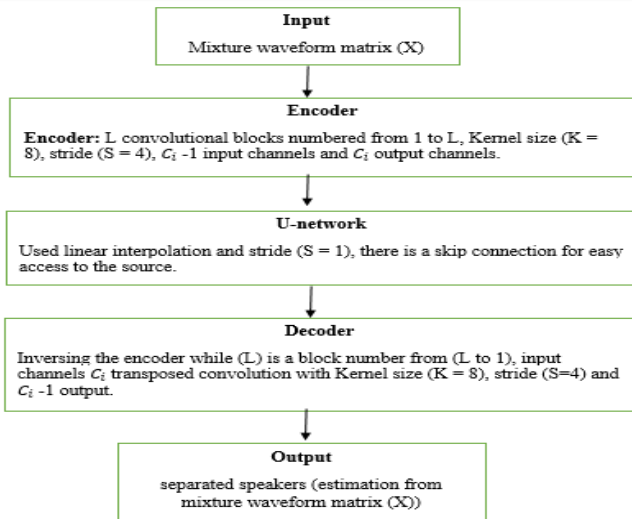
Figure 3: Demucs Stages

There is discrepancy between Conv-TasNet and Demucs. Demucs has a powerful decoder, that increments the kernel size 1 to 3, and the capacity as shown in table 1. Which gives advantage of preserve the lost information while mixing the instruments and unable to retain from the masking. Consequently, the disadvantage of retain information is execution time; because of the subsequent operations which could happen. Ultimately, time-domain and accuracy needs development.

3.1. Demucs General Equation

Basic approach of Demucs algorithm is (U-network structure), it uses Loss function to skip the association between encoder and decoder blocks to get straight transform to the initial input signal phase.

$$L_1(\widehat{X}_S, X_S) = \frac{1}{T} \sum_{t=1}^{T} |\widehat{X}_{s,t} - X_{s,t}|$$

$$L_2(\widehat{X}_S, X_S) = \frac{1}{T} \sum_{t=1}^{T} (\widehat{X}_{s,t} - X_{s,t})^2$$

Algorithm Notation:

$X_{s,t}$: $Waveform\ of\ source\ s, t\ is\ a\ subscript\ which\ is\ the\ t^{th}\ sample$

$\widehat{X}_{s,t}$: $Predicted\ waveform, t\ is\ a\ subscript\ which\ is\ the\ t^{th}\ sample$

$\frac{1}{T}$: $Number\ of\ Sampling$

$L(\widehat{X}_s, X_s)$: $is\ a\ reconstruction\ loose\ uses\ the\ average\ meant\ squrare\ error$

$of\ average\ absolute\ error\ between\ waveforms.$

Figure 4 clarify how the separation process of Demucs algorithm is in signals form. shows the separation results on a signal form of Demucs of (Music &Female) category.

Table 1 shows the comparison between both algorithms depend on four features. These features are channel, kernel, time-domain, and speed.
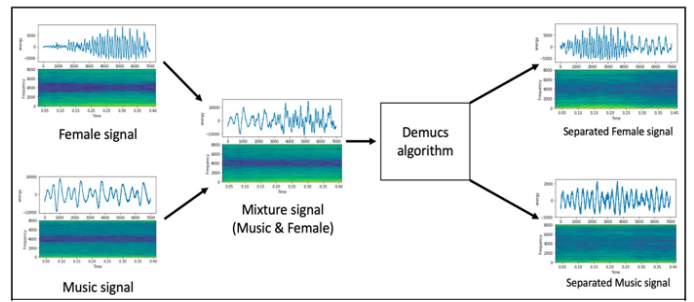
Figure 4: Demucs Algorithm

Table 1: Conv-TasNet & Demucs Coperasion

Features	Conv-TasNet	Demucs
channel	2 multiple channels	(2 or more) Multiple channels
kernel	(K = 1) low capacity	Bigger capacity kernel size ($K = 3$)
time-domain	offline/online	offline/online
speed	high speed	low speed

The comparison is based on some measurements. First, the SciPy Python library has been used to compute the standard deviation (SD) for the reference source (mixture matrix) and the estimation source (separated matrix). Additionally, the signal-to-noise-ratio (SNR) is a measurement tool used to measure the noise ratio of the signals. Second, the Scikit Learn Python library has been used to calculate the coefficient of determination for constant input arrays instead of accuracy. Moreover, , there is a root mean square error (RMSE) which is a management tool used to measure the difference between predicted and actual values. The mean absolute error tool (MAE) has been used to measure the prediction error for the testing and training datasets. Finally, the actual execution time has been calculated using the Python Time package. Since, this study is based on the BSS approach, it uses a performance evaluation package (mir_eval). It includes evaluation functions for music or audio signals. It has 4 measurement tools: (1) Signal-to-distortion-Ratio (SDR): it computes how is the similarity between the original signal and the produced signals, (2) Signal-to-Artifacts-Ratio (SAR): using artefacts helps in differentiate between the assessment errors, (3) Signal-to-Interference-Ratio (SIR) is the best measure of Gaussian noise and (4) Spatial-Distortion-Ratio (ISR) is A measurement tool usually used in image signals not in audio signal processing field.
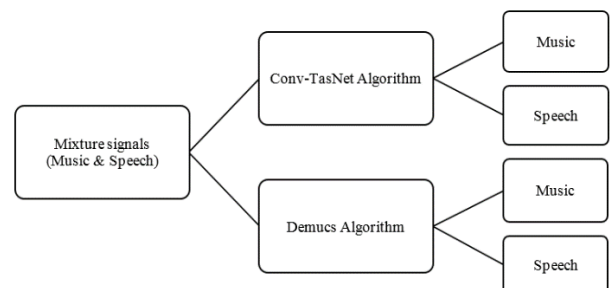
Figure 5: Data Processing for two algorithms

The mixture sound signal could be a mix of Drum, Vocals, Bass, and others. Demucs and Conv-TasNet can be tested on this mixture to get the estimation sound signal. The objective of this research is to study two estimated sources: music and vocals. Thus, the methodologies of these algorithms have changed to conform to the needs, as shown in figure 5. Then, the file of input and output stored the quantitative data as digital metrics at txt file. The digital metric files reduce the execution time and memory process; because they reduce the cost to process digital NumPy array rather than the audio signal.

4. Dataset and Pre-Processing

The dataset in this research project was obtained from youtube.com to compare and evaluate the Demucs and Conv-TasNet algorithms [25]. The data contained in a file with a WAV format. The dataset grouped within five classification experiments: Music with Male, Female, Child, and Conversation. The total size of experiment files is 10GB which contains a duration time of each experiment. The duration time is 3 hours, 28 minutes, and 20 seconds equal to 2GB for every single group. Data selection depends on the negativity of noise ratio, the clarity, and the suitability for instance, pure classic music such as Jazz, Piano, and more. Moreover, pure male sounds such as speech in international conferences like TED talk AND Apple. These sounds have no background noise or interfering sounds. They are a continuous speech extracted from Youtube.com. The factors (SD, SNR) of the pre-processing data were applied on all three languages mentioned before.

4.1. Data Types and signals

- Music (e.g., classical, piano, jazz): Figure 6 shows the jazz music acoustic vibrations.
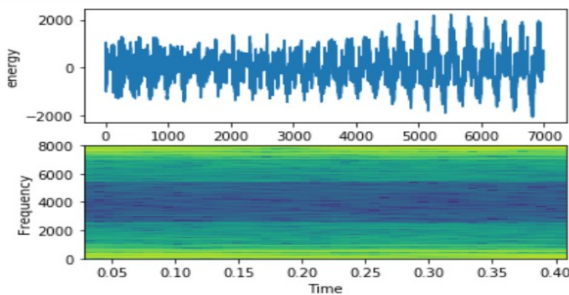
Figure 6: Music Signal on top energy on bottom frequency
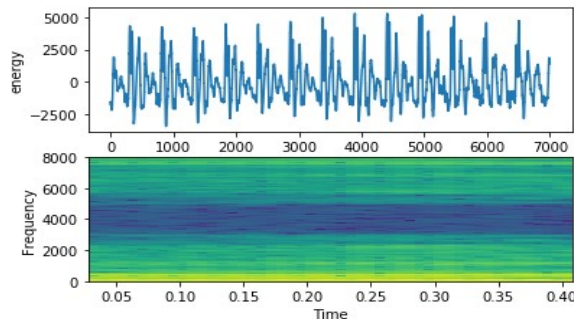
Figure 7: English Female Signal on top energy on bottom frequency

- Female (e.g. Michelle Obama speech): Figure 7 shows a high frequency of female signal like seasonal indices.
- Female (e.g., Balqees Fathi speech in women power conference): Figure 8 shows a high English female frequency of signal like seasonal but sequence indices.
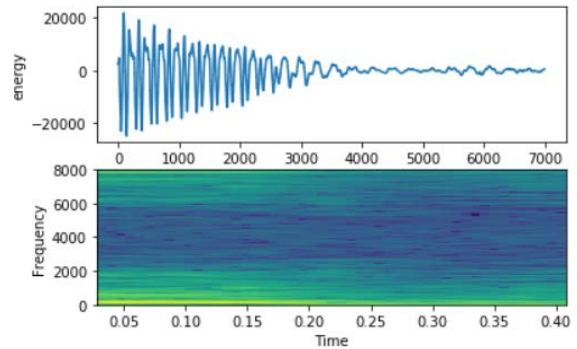
Figure 8: Arabic Female Signal on top energy on bottom frequency

- Chinese Female Figure 9 shows a mid-frequency of signal comparing with other signals.
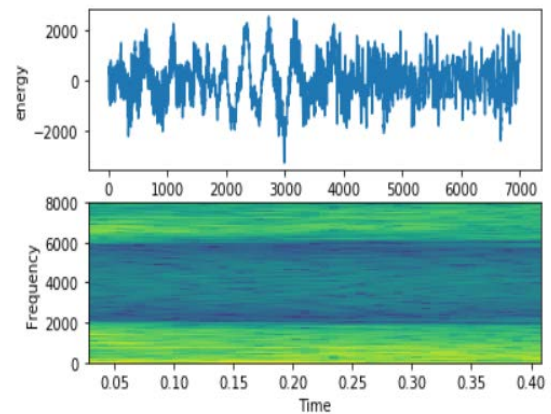
Figure 9: Chinese Female Signal on top energy on bottom frequency

- Male (e.g., male speech from Apple conference): Figure 10 shows a low frequency of male signal. Males' voices have lower intensity than females' voices. This figure shows acoustic vibrations likes a screaming sound.
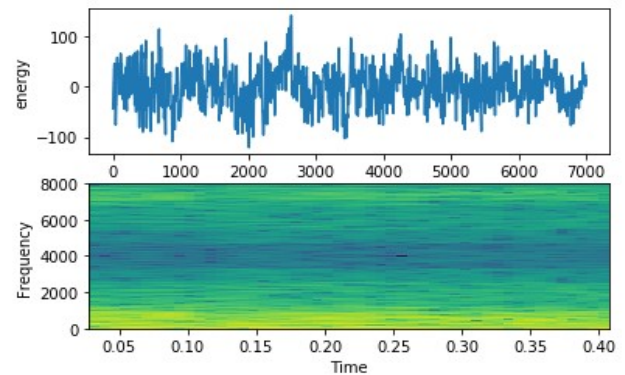
Figure 10: English Male Signal on top energy on bottom frequency

- Male (e.g., male speech from TEDX speech): Figure 11 shows the lowest frequency of male signal. Males' voices. This figure shows acoustic vibrations likes a screaming sound.
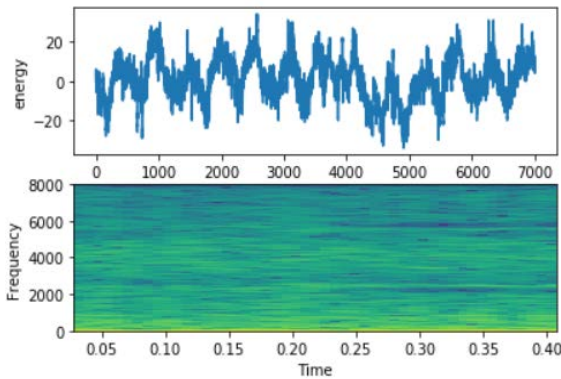
Figure 11: Arabic Male Signal on top energy on bottom frequency

- Chinese Male: Figure 12 shows the highest frequency of male signal. Males' voices. This figure shows more organized vibrations likes a seasonal.
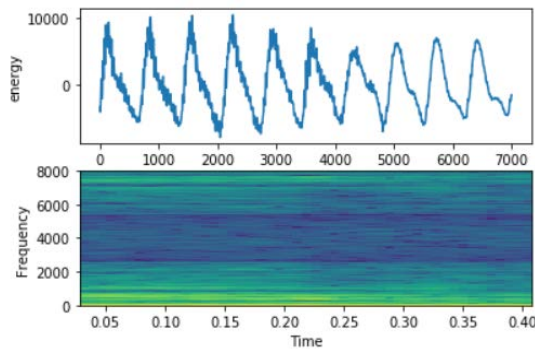
Figure 12: Chinese Male Signal on top energy on bottom frequency

- Conversation (e.g., male and female discussion): Figure 13 shows a mix between male and female voices. The first part is a male vocal while the other part is a female vocal.
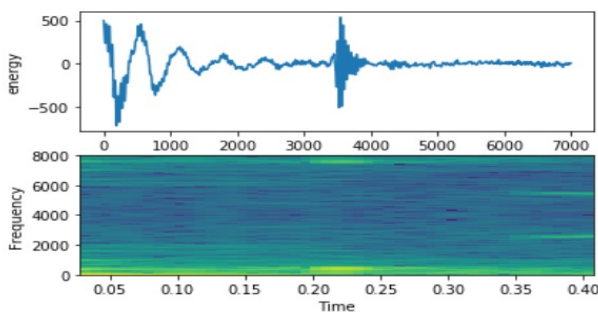
Figure 13: English Conversation Signal on top energy on bottom frequency

- Conversation (e.g., male and female discussion): Figure 14 shows a mix between male and female voices. The first part is a male vocal while the other part is a female vocal. It looks unorganized signal.
- Conversation (e.g., male and female discussion): Figure 15 shows a mix between male and female voices. The first part

is a male vocal while the other part is a female vocal. It looks unorganized signal.
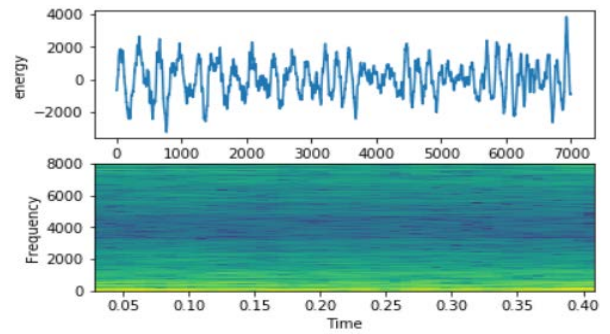
Figure 14: Arabic Conversation Signal on top energy on bottom frequency
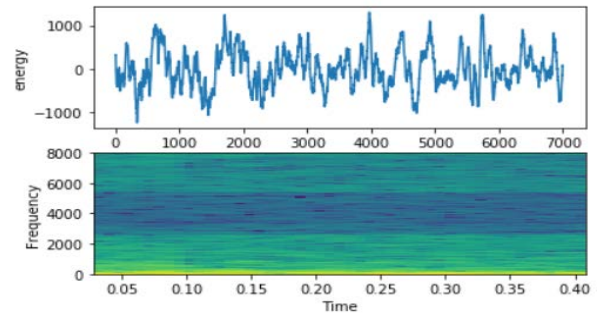
Figure 15: Chinese Conversation Signal on top energy on bottom frequency

- Child (e.g., child stories): Figure 16 shows a high vibration at the beginning, but slightly decreasing at the end.
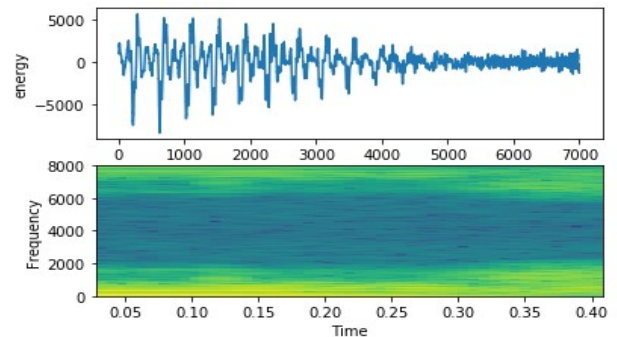
Figure 16: English Child Signal on top energy on bottom frequency

- Child (e.g., child stories): Figure 17 shows a high frequency.
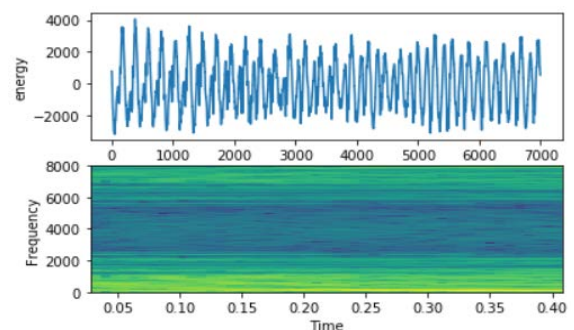
Figure 17: Arabic Child Signal on top energy on bottom frequency

- Child (e.g., child stories): Figure 18 shows a high frequency and heigh vibration
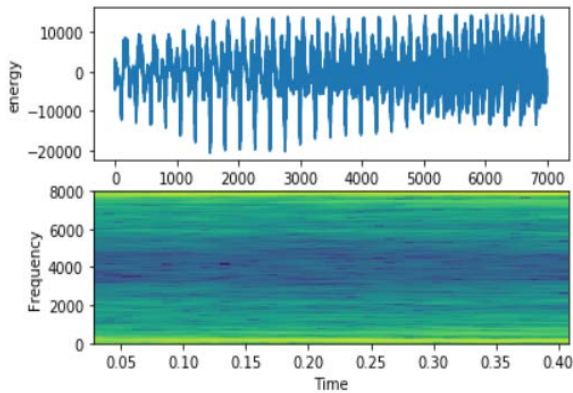
Figure 18: Chinese Child Signal on top energy on bottom frequency

4.2. Data pre-processing

Jupyter and Python libraries have helped produce the input segments of the separation process by a pre-processing, which contains two operations split and merge. The first operation is splitting, which splits the 3 hours audio file of each category into 1250 segments. The second operation is merging, which merges 1250 Music segments with each category one by one. Figure 19 shows the steps of pre-processing data segments. The mixture segment resources have 10 seconds of duration time per segment. Each file consists of two original segment resources (music & male), (music & female), (music & child) or/and (music & conversation).
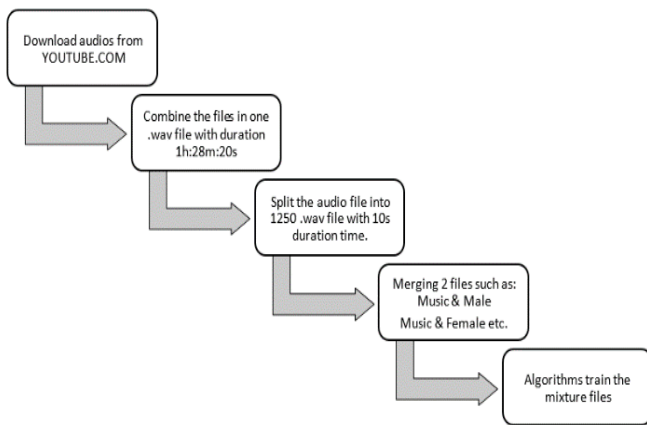
Figure 19: Steps of pre-processing data

The experiments are based on six dataset files. The main file is an Excel sheet contains four columns: Youtube URL (object), start time (float), end time (float), and label (object). There are five files for each category, one of them is a music file contains 1250 Records with IDs and music names [26]. Moreover, Male, Female, Child, and Conversation have files with IDs and their names. This research study evaluated an audio dataset of ten gigabytes. each category had two gigabytes of mixture segments with a music dataset. In addition, these files related to the original files by IDs. For each category of the dataset 70% was used for training and 30% for testing. Numpy is a Python library used for

randomizing splitting by randomize function. The experiment started with training data and then testing data of each mixture dataset. Finally, measurement tools were used to evaluate the separation results.

4.3. Audio signals description in terms of SD and SNR

Standard Deviation (SD) is a measurement tool used to compute the data spreads (distribution) via computing the mean of squared deviations, then square root of the result. [25].

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Xi - \mu)^2}$$

Table 2: The mean of standard deviation of all data

Mean	Music	Male	Female	Child	Conversation
English	0.111309	0.052772335	0.0523693	0.078172052	0.0620758
Arabic	0.111309	0.046679	0.022318	0.232547	0.083778
Chinese	0.111309	0.103396	0.107852	0.134417	0.086935

In table 2, the value of all columns = 0.1 SD. The results are closing to zero value, which indicates all the estimation values are near to the mean.

Signal-Noise-Ratio (SNR): is a measuring tool that calculates the ratio between the original signal with the artifact signal such as a noise. It computes the mean of a signal, then dividing the result by the SD [27], [28].

$$SNR = SD(MEAN(X))$$

Table 3: Signal-Noise-Ratio

Mean	Music	Male	Female	Child	Conversation
English	0.000257	-2.16667E-06	-4.53405E-6	-9.37E-06	-9.2014E-06
Arabic	0.000257	-3.80E-05	-9.94E-05	0.000168	-3.26E-05
Chinese	0.000257	0.000142	7.59E-06	-9.47E-05	0.000594

Table 3 shows the noise ratio of all categories. All SNR values are negative; this signifies a minimum noise ratio and considers as a pure signal. The negativity of SNR indicates as a high-quality score.

5. Experiments and Results

The audio datasets run on personal computer OMEN hp, i7 Intel core and 16GB of RAM. Using Python language for data processing via Jupyter and Pycharm. Using multiple packages such as: mir_eval Version: 0.5, Sklearn Version: 0.0, Scipy Version: 1.4.1, Pandas Version: 0.25.1, NumPy Version: 1.18.2, youtube_dl Version: 2020.3.24, Pydub Version: 0.23.1 and more. The audio files downloaded from Youtube.com via youtube_dl libraries. This downloading process took more than 2 hours per label. Merging Music files with all labels such as Male files via Overlay function this process takes 1 hour per label. After that, applying the data on Conv-TasNet and Demucs algorithms. For Demucs the process takes 6 hours and for Conv-TasNet the process takes 3 hours per label. The output of each algorithm

produced 12.3 GB augments on the memory process, 1250 folders composed of 3750 files produced from each algorithm. Evaluation of each output for each algorithm takes 2 hours. this calculation for each 1250 WAV mixture files. Four experiments were applied for training and testing datasets, to study the effectiveness or speech classification for comparing the algorithms. Four different data types were used in these algorithms: (Music & Male), (Music & Female), (Music & Conversation) and (Music & Child).

5.1. Evaluation Results

The comparison between Conv-TasNet and Demucs algorithms was based on seven measurement tools using three packages Scikit-Learn, mir_eval and time. Scikit-learn package supports three accuracy metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R square (R2). Mir_eval package affords four tools to evaluate the performance: Spatial-Distortion-Ratio (ISR), Signal-to-Artefacts-Ratio (SAR), Signal-to-Interference-Ratio (SIR), and Signal-to-distortion-Ratio (SDR). Ultimately, the time package was used to calculate the execution time of the separation process.

The three tools were implemented to calculate the accuracy of the Conv-TasNet and Demucs algorithms. They support the quality score evaluation pertaining for observation inputs and estimation outputs. The observations cannot be calculated directly with accuracy because continuous metrics cannot apply to accuracy tools. RMSE, R2 and MAE were used to describe the results of algorithms.

a- R Square (R^2)

It is a statistical tool used to measure the proportion of how close the data points to the fitted regression line. Via subtract the observed data with the data fitted on the line regression. In other words, it tells how data goodness fit the regression model [29] [30].

$$R^2 = \frac{Sum\ of\ Squares_{regression}}{Sum\ of\ Squares_{total}}$$

Training experiments:

Table 4: the experiments results pertaining to the (Music & Male and Music & Female) training dataset of all three languages.

Algorithm	Mean	Music & Male Experiment		Music & Female Experiment	
		Music	Male	Music	Female
Demucs	English	0.594750505	0.554857091	0.652751885	0.706623375
	Arabic	0.906899	0.730184	0.817224	0.252617
	Chinese	0.911638	0.924542	0.890167	0.924055
Conv-TasNet	English	0.673642922	0.611274307	0.725606045	0.782790036
	Arabic	0.939895	0.913543	0.91162	0.155181
	Chinese	0.937934	0.936259	0.898883	0.925415

Table 4 shows that Music has an R-squared (R^2) value of 59% and (Male) has an R^2 value of 55%. The R^2 value can be more accurate if the noise ratio is low. For the R^2 values described between 0% and 100%, 0% indicates that the observed data are

far from the mean and have a low error estimation, while 100% indicates that the observed data are close to the mean [31].

As it shows that (Music & Male) experiments of both Arabic and Chinese are close to 100% which indicates the observed data are close to the mean. In another experiment, (Music & Female) of Arabic language is close to 0% which indicates the observed data are far from the mean and have a low error estimation.

Table 5: R Square of Conversation and child experiments of Training Dataset of all three languages.

Algorithm	Mean	Music & Conversation Experiment		Music & Child Experiment	
		Music	Conversation	Music	Child
Demucs	English	0.641354444	0.682678489	0.624809687	0.763154352
	Arabic	0.938469	0.91497	0.67352	0.922591
	Chinese	0.865062	0.799423	0.884608	0.937521
Conv-TasNet	English	0.724467042	0.753158555	0.662892685	0.759285571
	Arabic	0.658099	0.000153	0.857881	1.16E-05
	Chinese	0.875329	0.822089	0.962708	0.477527

Tables 4 and 5 show the English language experiment results of Demucs algorithm, pertaining to the (Music & Female), (Music & Conversation) and (Music & Child) training datasets. The R^2 values of observation and prediction signals of each algorithm, are (Music) 65% and (Female) 70%, (Music) 64% and (Conversation) 68% and (Music) 62% and (Child) 76%, respectively. These results show that the (Music & Female) experiments achieved high R^2 values and that (Child) also achieved a higher value of Demucs of training experiment. While (Music & Male) experiment of Demucs has the lowest value 59% and 55% respectively. On the other hand, Conv-TasNet achieved the highest value of R^2, where (Female) $R^2 = 78\%$, that means Female signals are the most signals close to the mean. In this evaluation it is noticeable that the Conv-TasNet always has the higher value compared to Demucs. Demucs training experiment during Arabic language has the highest value of R square which is 93% at Music (Music & Male) experiment, while it has the lowest speech value 25% at Female (Music & Female) experiment. On another side, Conv-TasNet experiment during Chinese language has the highest value 96% at Music (Music & Child) experiment and Arabic language has lowest value at speech 0.0001% at (Music & Conversation) and (Music & Child) experiments.
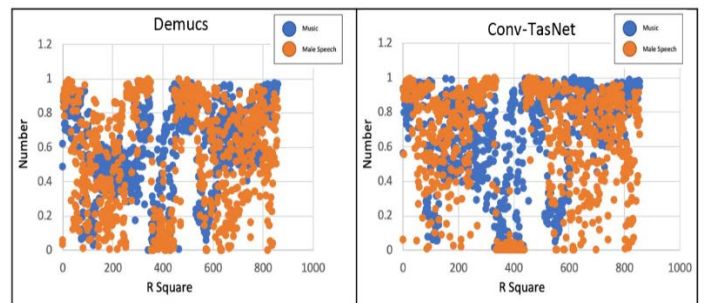
Figure 20: $R^2$of (Music & Male) experiment of Demucs & Conv-TasNet of English language.

On another side, Arabic language with Demucs algorithm achieved the highest value of R-square at speech side (Male, Conversation and Child). Also, Chinese language usually has high value of R-square near to 100% which means better performance, close to the mean and low estimation errors on signals.

The right plot in figure 20 represents the R square of observation and prediction data of (Music & Male) of Demucs. The left plot shows the spread points of observation and prediction data of (Music & Male) of Conv-TasNet of training data. The blue points show the music data and orange points show the males speech data. It explains the spread of data point of observation and prediction data. X-axis represent the R square values of (Music & Male) experiment. The sample data is not the reason for the increased R^2 value because all the experiments used the same number of segments. However, R^2 result may be an effect of the lower noise ratio of observation.

Testing experiments:

Like the training experiment, all values in tables 6 and 7 are above 50%. The highest value of R^2 is 77% of Conv-TasNet pertaining to the (Female) experiment and 67% of (Female) of Demucs algorithm. In training and testing data experiments, training data has the same R^2 with testing data which inference that data size sample has no affection of R^2 value.

Table 6: R^2 of male and female experiments of testing dataset

Algorithm	Mean	Music & Male Experiment		Music & Female Experiment	
		Music	Male	Music	Female
Demucs	English	0.598482775	0.527509406	0.651269443	0.672236996
	Arabic	0.929927	0.90909	0.872887	0.54528
	Chinese	0.917021	0.908275	0.888142	0.911523
Conv-TasNet	English	0.677168022	0.659364504	0.712894379	0.774478377
	Arabic	0.927393	0.796538	0.958407	0.882355
	Chinese	0.966593	0.479971	0.85849	0.885474

Table 7: R^2 of Conversation and Child experiments of Testing Dataset

Algorithm	Mean	Music & Conversation Experiment		Music & Child Experiment	
		Music	Conversation	Music	Child
Demucs	English	0.645854253	0.681211663	0.624127646	0.756858895
	Arabic	0.8902	0.866712	0.744405	0.669279
	Chinese	0.865062	0.799423	0.893974	0.94875
Conv-TasNet	English	0.718706834	0.767224614	0.670057154	0.75645193
	Arabic	0.899566	8.36E-06	0.676389	0.000152
	Chinese	0.510896	0.393926	0.812849	0.910011

Demucs testing experiment during Arabic language has the highest value of R square which is 93% at Music (Music & Male) experiment, while English language has the lowest speech value 53% at speech (Music & Male) experiment. Also, at Arabic language with the value of speech 54% at (Music & Female)

experiment. On another side, Conv-TasNet experiment during Chinese language has highest value 97% at Music (Music & Male) experiment; and Arabic language has lowest value at speech 0.000008% at (Music & Conversation) experiment and 0.0001 at (Music & Child) experiment.

b- Root Mean Square Error (RMSE)

where R^2 is the relative measure of the fit of the model, RMSE is the absolute measure of the fit. It is also a tool to measure how accurately the algorithm predicts the response. It can be used instead of accuracy tools because the accuracy tool does not support the continuous array of reference resources [32] [33].

$$RMSE = \sqrt{1 - r^2 SD_y}$$

Training experiments:

Tables 8 and 9 show the RMSE values for the (Music & Male), (Music & Female), (Music & Conversation) and (Music & Child) training datasets. The RMSE values are (Music) 0.06 and (Male) 0.04, (Music) 0.06 and (Female) 0.03, (Music) 0.06 and (Conversation) 0.03 and (Music) 0.06 and (Child) 0.03, respectively of Demucs. Lower values of RMSE indicate absolute fit. Overall, both algorithms have positive values from (0.3 to 0.6) which means that the observations and prediction data are close.

Arabic and Chinese language have lower value relatively with English language. Mostly, the value of RMSE is around 0.3 which indicates the absolute fit of the output signals comparing with English separation of both algorithms.

Table 8: Root Mean Square Error of Male and Female experiments

Algorithm	Mean	Music & Male Experiment		Music & Female Experiment	
		Music	Mal	Music	Female
Demucs	English	0.060387951	0.035588642	0.055773658	0.026737967
	Arabic	0.028117	0.028008	0.039305	0.031757
	Chinese	0.030426	0.028518	0.034205	0.029987
Conv-TasNet	English	0.052309023	0.031107757	0.048070573	0.021558262
	Arabic	0.021331	0.021132	0.025871	0.025568
	Chinese	0.026394	0.025694	0.031611	0.030064

Table 9: Root Mean Square Error of conversation and child experiments.

Algorithm	Mean	Music & Conversation Experiment		Music & Child Experiment	
		Music	Conversation	Music	Child
Demucs	English	0.058033883	0.027021274	0.057891011	0.032158433
	Arabic	0.027798	0.024634	0.062237	0.064626
	Chinese	0.039827	0.038917	0.036068	0.033564
Conv-TasNet	English	0.04975772	0.020641371	0.053277448	0.032526892
	Arabic	0.061912	0.247073	0.042143	0.173703
	Chinese	0.033942	0.036129	0.019014	0.015994

Testing experiments:

Tables 10 and 11 show the RMSE values for the (Music & Male), (Music & Female), (Music & Conversation) and (Music & Child) testing datasets. Similarly, with training dataset, the values of Conv-TasNet and Demucs are almost the same. Overall, both algorithms have positive values from (0.3 to 0.6) which indicates that observations and prediction data are almost fit.

Like training data, the testing experiment also shows better results of RMSE of both languages Arabic and Chinese.

Table 10: Root Mean Square Error of Male and Female experiments

Algorithm	Mean	Music & Male Experiment		Music & Female Experiment	
		Music	Male	Music	Female
Demucs	*English*	0.061387858	0.038965244	0.058082848	0.02998228
	Arabic	0.024837	0.018296	0.03204	0.031994
	Chinese	0.024996	0.027324	0.034895	0.029049
Conv-TasNet	*English*	0.052586705	0.029218817	0.05128763	0.020946208
	Arabic	0.023592	0.023737	0.01534	0.014384
	Chinese	0.017425	0.013958	0.032243	0.031023

Table 11: Root Mean Square Error of conversation and child experiments.

Algorithm	Mean	Music & Conversation Experiment		Music & Child Experiment	
		Music	Conversation	Music	Child
Demucs	*English*	0.05703156	0.026870191	0.059313099	0.032708388
	Arabic	0.033467	0.03027	0.041961	0.047127
	Chinese	0.039827	0.038917	0.030169	0.029733
Conv-TasNet	*English*	0.049186272	0.021074537	0.0546344	0.033307789
	Arabic	0.029296	0.156674	0.054295	0.24063
	Chinese	0.071935	0.080743	0.040955	0.040211

c- Mean Absolute Error (MAE)

MAE is a tool for measuring the accuracy of training algorithms. It handles continuous NumPy arrays and can be used to enhance the results of RMSE when measuring the algorithms' predictions. It computes the average magnitude of the errors that occur between predictions and actual observations.

$$MAE = \frac{1}{n} \sum_{j=1}^{n} |y_j - \hat{y}_j|$$

Training experiments:

Notably, tables 12 and 13 show (Female) and (conversation) of training dataset experiments, via Conv-TasNet get the smallest value of error 0.01 between the reference and estimation resources. Demucs has 0.04 of (Music) of all experiments and 0.02 of each speech (Male, Female, Conversation and child).

Table 12: Mean Absolute Error of male and female experiments.

Algorithm	Mean	Music & Male Experiment		Music & Female Experiment	
		Music	Male	Music	Female
Demucs	*English*	0.043319834	0.023187058	0.039918792	0.016335086
	Arabic	0.015769	0.015549	0.02708	0.022367
	Chinese	0.018334	0.016945	0.021714	0.019184
Conv-TasNet	*English*	0.036686502	0.020443172	0.034022512	0.01306448
	Arabic	0.012075	0.011286	0.015174	0.014449
	Chinese	0.016413	0.015721	0.019738	0.018671

Table 13: Mean Absolute Error of conversation and child experiments.

Algorithm	Mean	Music & Conversation Experiment		Music & Child Experiment	
		Music	Conversation	Music	Child
Demucs	*English*	0.041430066	0.016731485	0.040630954	0.018055779
	Arabic	0.017914	0.015582	0.045522	0.047743
	Chinese	36.54131796	14.75716936	35.2270372	15.65436042
Conv-TasNet	*English*	0.034715515	0.012663916	0.035056178	0.016750811
	Arabic	0.045405	0.18224	0.029647	0.121688
	Chinese	30.34135987	11.05559904	0.012242	0.010555

Notably Demucs algorithm at training experiment shows the highest performance of Music separation 0.02 MAE in Arabic language of both (Music & Male) and (Music & Conversation) experiments as well as Chinese language at (Music & Male) experiment. Chinese language has significant values - highest value of error - through (Music & Conversation) and (Music & Child) experiments, these values are (36.5 & 14.8) and (35.2 & 15.7), respectively. Additionally, Conv-TasNet algorithm in training experiment shows the highest performance of Music separation 0.01 MAE in Arabic language at (Music & Male) experiment. However, Chinese language has significant values (30.3 & 11.1) at (Music & Conversation) sequentially which considered as the lowest performance in separation process.

Testing experiments:

Comparably, tables 14 and 15 of testing dataset experiment are like training dataset experiment of MAE values.

Table 14: Mean Absolute Error of male and female experiments.

Algorithm	Mean	Music & Male Experiment		Music & Female Experiment	
		Music	Male	Music	Female
Demucs	*English*	0.044283635	0.025880494	0.041179648	0.018042131
	Arabic	0.013289	0.009229	0.020964	0.020834
	Chinese	0.015598	0.016635	15.89534399	6.964262591
Conv-TasNet	*English*	0.0370836	0.019214275	0.035862557	0.012616786
	Arabic	0.013402	0.013194	0.008488	0.007984
	Chinese	46.01418709	25.07100812	13.37673393	4.706061233

Table 15: Mean Absolute Error of Conversation and Child experiments

Algorithm	Mean	Music & Conversation Experiment		Music & Child Experiment	
		Music	Conversation	Music	Child
Demucs	*English*	0.040552242	0.016646969	0.041642397	0.018309848
	Arabic	0.020308	0.017798	0.029497	0.033534
	Chinese	0.010246	0.007809	0.019657	0.019533
Conv-TasNet	*English*	0.034867332	0.013045045	0.036287197	0.017151002
	Arabic	0.019533	0.102707	0.038631	0.174129
	Chinese	0.056936	0.064269	0.024952	0.024837

In Demucs testing experiment, Arabic language has 0.01 at Music (Music & Male) experiment. Chinese language has 0.01 at Music (Music & Conversation) experiment. Nevertheless, Chinese language has highest value (15.9 & 6.10) at (Music & Female) experiment that indicates a high error rate through separation process. Conv-TasNet testing experiment has lowest error 0.01 in Chinese language at Music (Music & Conversation) experiment. On another side, Chinese language has a high error after separation process; the values are 46.0 & 25.1 at (Music & Male) experiment and 13.4 & 4.7 at (Music & Female) experiment.

5.2. Mir_eval of (SDR, SAR, SIR, ISR)

It is a Python package used to evaluate the results, which retravel machine learning algorithms. It extracts the music separation information pertaining to reference and estimation resources. Mir_eval package quantitatively compares the signals to algorithm implementations. The original mir_eval is a bss_eval package, which has been developed to evaluate the audio and image signals [34].

Training experiments:

Table 16: mir_eval of male and female experiments

Algorithm	SDR	Music & Male Experiment		Music & Female Experiment	
		Music	Male	Music	Female
Demucs	English	6.365845562	2.261560323	7.724350679	5.891797598
	Arabic	14.34474	4.531938	11.95328	-4.01775
	Chinese	12.42069	11.31496	12.1474	11.05148
Conv-TasNet	English	7.768486164	2.785485859	9.208170136	7.873549345
	Arabic	14.66376	11.28408	14.29149	-5.51784
	Chinese	13.49518	12.34095	12.57517	11.18431
SIR					
Demucs	English	6.365845562	2.261560323	7.724350679	5.891797598
	Arabic	14.34474	4.531938	11.95328	-4.01775
	Chinese	12.42069	11.31496	12.1474	11.05148
Conv-TasNet	English	7.768486164	2.785485859	9.208170136	7.873549345
	Arabic	14.66376	11.28408	14.29149	-5.51784
	Chinese	13.49518	12.34095	12.57517	11.18431

SDR is the signal distortion ratio. It describes how similar the estimation resources are to the reference resources. SDR provides a global performance measure; but the other three measures are also important. Tables 16 and 17 show the average SDR values, taken from all the tracks. In these tables, Conv-TasNet produces a higher SDR score for the (Female & Music) experiment, and for the (Child) training data set where (Child) = 8.3 SDR'. In turn, Conv-TasNet produces a lower SDR score for the (Child & Music) training data set where (Music) = 6.6 SDR for the (Music) training dataset and a smaller SDR value where (Male) = 2.8 SDR for the (Male & Music). On the other side, Demucs produces a higher value where (Music) = 7.7 SDR for the (Female) training dataset and where (Child) = 8.1 SDR for the (Music & Child) experiment. However, it produces a lower SDR score where (Music) = 6.4 SDR and 6.5 SDR, respectively, for the (Male) and (Child) training datasets and where (Male) = 2.3 SDR for the (Male) training dataset.

Comparing between the language experiments at Table 19, The highest SDR is in an Arabic experiment. Where Music values around 11 to 14 SDR and Male speech, but Female speech has the lowest value -4 and -5 SDR in both algorithms. While Chinese experiment has a heigh value 11 to 12 at Music and Male or Female speech.

Table17: mir_eval of conversation and child experiments

Algorithm	SDR	Music & Conversation Experiment		Music & Child Experiment	
		Music	Conversation	Music	Child
Demucs	English	7.332337909	4.889571299	6.460674567	8.145132574
	Arabic	13.08034	10.58806	3.800309	11.54092
	Chinese	10.14399	6.673121	9.924395	12.20906
Conv-TasNet	English	8.941261981	7.486695151	6.626825047	8.292722159
	Arabic	3.972427	3.972427	8.91058	-26.6596
	Chinese	10.63976	7.20355	16.42239	-8.05199
SIR					
Demucs	English	7.332337909	4.889571299	6.460674567	8.145132574
	Arabic	13.08034	10.58806	3.800309	11.54092
	Chinese	10.14399	6.673121	9.924395	12.20906
Conv-TasNet	English	8.941261981	7.486695151	6.626825047	8.292722159
	Arabic	3.972427	3.972427	8.91058	-26.6596
	Chinese	10.63976	7.20355	16.42239	-8.05199

At Table 17, there are more significant results of SDR affected by NL. Arabic experiment of Demucs algorithm has 13 at Music of conversation speech separation, but it has 3.8 SDR at Music of Child speech separation. On another side, Conv-TasNet algorithm during Arabic separation of conversation experiment is 4 SDR. While the lowest value in training experiment is -26.7 SDR.

Testing experiments:

One the other side, Conv-TastNet and Demucs have the higher values of (Child) testing dataset experiment, both values are 8.02 SDR. Then, (Conversation), (Female) and (Male) of testing dataset experiments respectively. SIR and SDR have the same values, which means a high correlation between these two measures. This is further indicated by the fact that they dominated in terms of estimation error. However, SIR is the best

measure of Gaussian noise, while SDR focuses more on interference-level intrusiveness and separation [35]-[37]. The experiments in this research study illustrated that Conv-TasNet is a preferable algorithm from the perspective of SDR and SIR. The reason is Demucs using a larger window-frame size'. Increasing weighting for noise power can cause potential modification of the SDR measure.

Table 18 and 19, in the context of comparing the Conv-TasNet algorithm with other research in terms of SDR, in [2], the authors have trained four models using the MusDB dataset [38], with 100 songs through Demucx, Conv-TasNet, Open-Unmix, MMDnseLSTM and Wave-U-Net. The average SDR of the vocals was 6.81 for the Conv-TasNet algorithm (the higher score) and 7.61 for the MMDnseLSTM algorithm. Their research study focused on music separation, while our research focused on music and speech separation in four categories. The highest value of music SDR separation for the (Music) and Female datasets was in Conv-TasNet. Moreover, there are some classification experiments that produced scores even higher than those found in [9], such as the Child, Female and Conversation training datasets. For Demucs algorithm, the score of SDR was 8.15 in the Child dataset; this was higher than the reported in [9], (6.08 for MusDB and 7.08 for 150 extra song tracks).

Table 18: mir_eval of Male and Female experiments

Algorithm	SDR	Music & Male Experiment		Music & Female Experiment	
		Music	**Male**	**Music**	**Female**
Demucs	English	6.303056998	1.461799732	7.607939697	4.703374514
	Arabic	12.92352	10.95636	11.55566	1.264586
	Chinese	11.78847	10.37697	10.47468	10.54035
Conv-TasNet	English	6.303056998	1.461799732	7.607939697	4.703374514
	Arabic	15.76899	6.375536	14.3024	9.198612
	Chinese	17.21237	-7.92302	9.808399	10.60297272
SIR					
Demucs	English	6.303056998	1.461799732	7.607939697	4.703374514
	Arabic	12.92352	10.95636	11.55566	1.264586
	Chinese	11.78847	10.37697	10.47468	10.54035
ConvTasNet	English	6.303056998	1.461799732	7.607939697	4.703374514
	Arabic	15.76899	6.375536	14.3024	9.198612
	Chinese	17.21237	-7.92302	9.808399	10.60297272

Table 19: mir_eval of conversation and child experiments

Algorithm	SDR	Music & Conversation Experiment		Music & Child Experiment	
		Music	**Conversation**	**Music**	**Child**
Demucs	English	7.390943771	4.908684372	6.583362429	8.02266854
	Arabic	11.81469	8.604822	6.909017	6.11558
	Chinese	10.14399	6.673121471	10.3703	13.03493
Conv-TasNet	English	8.912469304	7.531863904	6.66059257	8.02425869
	Arabic	10.93189	-26.1081	4.572872	-26.5985
	Chinese	1.261381	-0.95342	8.354678	10.88152
SIR					
Demucs	English	7.390943771	4.908684372	6.583362429	8.02266854
	Arabic	11.81469	8.604822	6.909017	6.11558
	Chinese	10.14399	6.673121471	10.3703	13.03493
Conv-TasNet	English	8.912469304	7.531863904	6.66059257	8.02425869
	Arabic	10.93189	-26.1081	4.572872	-26.5985
	Chinese	1.261381	-0.95342	8.354678	10.88152

In [8], the author studied the application of the Conv-TasNet algorithm to the WSJ0-3mix dataset for speech separation. Their study focused on how causality for Conv-TasNet that is a causal configuration due to causal convolution and/or layer normalization operations, leads to drops in Conv-TasNet performance. The SDR of causal Conv-TasNet value is 8.2, while the corresponding value is 13.1 for non-causal Conv-TasNet. The high SDR score for non-causal Conv-TasNet refers to the high signal noise ratio that is equal to 12.7, while -0.00001 SNR of music tracks of this research.

Notably, the signal-to-artifacts ratio (SAR) value has no correlation to the SIR and SNR values because it calculates the quantization and interferences of signals. It assists in distinguishing between the estimation errors using artefacts. For all the experiments of different languages and algorithms, SAR has zeros value. Spatial-Distortion-Ratio (ISR) is more often used for image signal processing than audio signal processing. For all the experiments of different languages and algorithms, ISR has infinite value. The mir_eval package is limited when it comes to auditory signification. For instance, SIR can hardly distinguish between two different variables. In addition, SDR does not compute the total perceived distortion in accurate manner.

5.3. Execution time

Figure 21 shows that Demucs separated (Music & Male) faster than (Conversation & Music), (Child & Music) and (Female & Music) of training and testing datasets respectively. Conv-TasNet separated (Female & Music) faster than (Music & Male), (Child & Music) and (Conversation & Music) in training dataset respectively. A slightly different in ms is between (Child & Music) and (Conversation & Music), but (Conversation & Music) was faster than (Child & Music) in testing dataset. Also, figure 21, shows the average execution time in nanoseconds per track of Conv-TasNet that was seven times faster than Demucs, for testing and training datasets in all languages' experiments. Conv-TasNet could be applied in real-time better than Demucs. Demucs needs further development, comparing with Conv-TasNet delay time. In [2], the author computed the speed of training of Demucs as 1.6 seconds and of Conv-TasNet as 0.7 seconds. However, this comparison is not accurate enough for consideration. The reason is that the Conv-TasNet was applied to only two seconds of track duration; also, for Demucs using a different sample size significantly affected the actual processing time. Noticeable, that there was a slightly different in execution time between different natural languages around 0.1 which is not necessary to be mentioned, because the separation process of algorithms does not affect by NL. The results from each algorithm were based on some equations to perform a separation process. The equations of each algorithm and their notations have mentioned before in research methodology section.

Mean of Excution Time of Traning and testing experiment of English Language

Mean of Excution Time of Traning and testing experiment of Chinese Language

Mean of Excution Time of Traning and testing experiment of Arabic Language
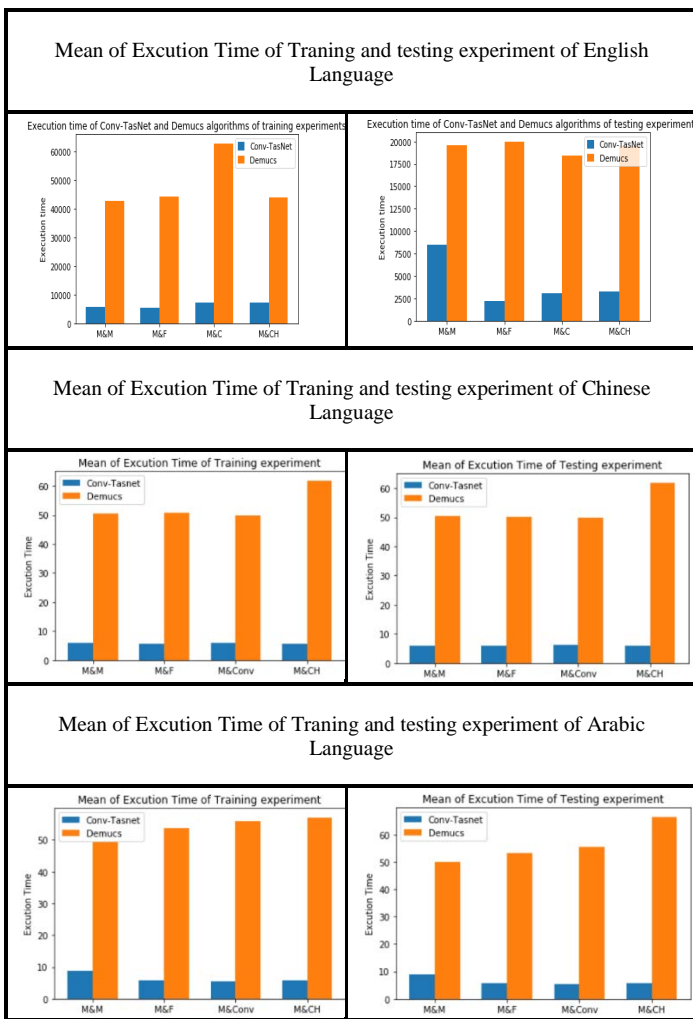
Figure 21: The mean of execution time of English, Arabic and Chinese languages.

In general, after calculating the sum of execution time in nanoseconds (ns) of each separation process of languages. training and testing Demucs experiment, English is 202.3 ns and 203.2 ns the fastest at separation process, respectively. However, Arabic is 216.1 ns and 225.1 ns the slowest at separation process, respectively. Training and testing Conv-TasNet experiment, English is 29.1 ns and 29.4 ns the slowest at separation process, respectively. Nevertheless, Chinese is 23 ns faster than English and Arabic at Training experiment. While Arabic is 25.9 ns faster than English and Chinese at Testing experiment, See Appendix Table 20-25.

6. Conclusion

This research study compared the Conv-TasNet and Demucs algorithms. Both algorithms use BSS approach during the separation process. Conv-TasNet and Demucs estimate the number of blind sources to be separated into four sources: drums, bass, vocals and other. However, this research reduces the number of sources to be two, which are vocal and music. These algorithms are based on two approaches: real-time and blind-source separation. A random 10GB audio set was taken from Youtube.com. These data separated into five categories: Male, female, Music, Conversation and Child. Using three Natural languages, which is English, Arabic and Chinese [25]. The

performance with different and specific sources has been tested. The results show Conv-TasNet has excellent performance in separating music and speech sources. The highest SDR score of music is 9.21 in the female experiment, while the highest SDR score of speech at child experiment is 8.14. In addition, the average execution time of Conv-TasNet algorithm is seven times faster than Demucs algorithm. With the NL extend, in general, Chinese language has high performance at source separation process during Demucs and Conv-TasNet algorithms in training and testing experiments. However, there are some significant values of separation performance at (Music & Conversation) experiment. The values are (15.9 & 6.10) indicate high MAE error. Both algorithms have high error during source separation process of Chinese (Music & Conversation) experiments. The mixture signal contains Music and Male & Female Conversation.

Supplementarily, in general, Arabic language has high performance during Demucs separation process at (Music & Male). Normally, English language shows minor results comparing with Arabic and Chinese languages during both separation algorithms as it mentioned from table 6 to table 12.

There are some limitations in this research study. Noise reduction was attempted with multiple algorithms such as (spectral gating of noise reduction [38], Deep Convolutional Neural Networks for Musical Source Separation [39], aasp-noise-red [40], Deep Audio Prior - Pytorch Implementation [41], Joint Audio Correction Kit (J.A.C.K.) [42] and more). Nonetheless, those attempts were unsuccessful because the environment needs to be python 2. Old libraries are no longer supported for instance: J.A.C.K algorithm for noise reduction implemented on Python 3 via Pycharm. There was a lot of errors fixed by the researcher but after loading the input files there was no results because there is a runtime errors, which is one of the difficult errors to be tracked and fixed. The Demucs algorithm could not support large amount of data. The dataset was 10GB, but the Demucs algorithm started to lose values without separating after training 8GB. The Authors have environment limitation to run Demucs GPU version, because it needs to be run on a wider range of hardware.

The future work of this research will focus on training more audio sets or different audio sets, using different algorithms like deep learning approach to separate the data and increasing the number of data categories. In addition, noise reduction can be a function inside the BSS algorithms. The Conv-TasNet algorithm requires evolution in terms of actual time separation, computational cost and development of the separation process to achieve 100% interference separation. In addition, use the NL process functions in both algorithms could lead to a different and significant results.

7. References

[1] A. Alghamdi, G. Healy, H. Abdelhafez "Real Time Blind Audio Source Separation Based on Machine Learning Algorithms", IEEE 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 35-40, Oct, 2020, doi:10.1109/NILES50944.2020.9257891

[2] M. Pal, R. Roy, J. Basu, M. S. Bepari, "Blind Source Separation: A Review and Analysis," 2013 International Conference Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation, 1-5, 2013, doi:10.1109/ICSDA.2013.6709849.

[3] K. Ball, N. Bigdely-Shamlo, T. Mullen, K. Robbins, "PWC-ICA: A Method for Stationary Ordered Blind Source Separation with Application

to EEG," Computational intelligence and neuroscience **2016**(73), 1-20, 2016, doi:10.1155/2016/9754813.

[4] E. M. Grais, H. Wierstorf , D. Wa, R. Mason and M. D. Plumbley, "Referenceless Performance Evaluation of Audio Source Separation using Deep Neural Networks.," 27th European Signal Processing Conference (EUSIPCO). IEEE, 1-5, 2019.

[5] N. Makishima, S. Mogami, N. Takamune, D. Kitamura, H. Sumino, S. Takamichi , H. Saruwatari and . N. Ono, "Independent Deeply Learned Matrix Analysis for Determined Audio Source Separation," EEE/ACM Transactions on Audio, Speech, and Language Processing, **27**(10), 1601-1615, 2019.

[6] V. Leplat, N. Gillis, M. Shun Ang, "Blind Audio Source Separation with Minimum-Volume Beta-Divergence NMF," IEEE Transactions on Signal Processing, **68**, 3400 – 3410, 2020, doi: 10.1109/TSP.2020.2991801.

[7] Y. Luo and N. Mesgarani, "TASNET: TIME-DOMAIN AUDIO SEPARATION NETWORK FOR REALTIME, SINGLE-CHANNEL SPEECH SEPARATION," IEEE/ACM Transactions on Audio, Speech, and Language Processing, **27**(8), 1256-1266, 2019, doi: 10.1109/TASLP.2019.2915167.

[8] G. Wichern, M. Flynn, J. Antognini, E. McQuinn, "WHAM!: Extending Speech Separation to Noisy Environments," Sep. 2019. [Online]. Available: http://wham.whisper.ai/.

[9] Lab Neural Acoustic Processing, "NEURAL ACOUSTIC PROCESSING LAB," 2019. [Online]. Available: http://naplab.ee.columbia.edu/tasnet.html.

[10] D. Fourer G. Peeters, "Fast and adaptive blind audio source separation using recursive Levenberg-Marquardt synchrosqueezing," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 766-770, 2018, doi: 10.1109/ICASSP.2018.8461406.

[11] B. Pardo, "Interactive Audio Lab," 2018. [Online]. Available: https://interactiveaudiolab.github.io/.

[12] B. Loesch, B. Yang, "Blind source separation based on time-frequency sparseness in the presence of spatial aliasing," International Conference on Latent Variable Analysis and Signal Separation. Springer, Berlin, Heidelberg, 1-8, 2010.

[13] Informatics, UCD School of Computer Science and, "CHAINS characterizing Individual Speakers," 2006. [Online]. Available: https://chains.ucd.ie/.

[14] Y. Mori, T. Takatani, H. Saruwatari, K. Shikano, T. Hiekata, T. Morita, "High-presence hearing-aid system using DSP-based real-time blind source separation module.," IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, IV-609, 2007.

[15] S. Rickard, R. Balan and J. Rosca, "Real-time time-frequency based blind source separation," AJE, pp. 1-2, 2001.

[16] H. Abouzid and O. Chakkor, "Blind source separation approach for audio signals based on support vector machine classification.," Proceedings of the 2nd international conference on computing and wireless communication systems, pp. 1-6, 2017, doi: 10.1109/ICASSP.2007.366986.

[17] A. Ferreira, D. Alarcão, "Real-time blind source separation system with applications to distant speech recognition.," Applied Acoustics, 113, 170-184, 2016, doi: 10.1016/j.apacoust.2016.06.024.

[18] D. J. Watts. Small Worlds: The Dynamics of Networks between Order and Randomness, Princeton Univ. Press, Princeton, USA, 1999.

[19] D. Stoller, S. Ewert, S. Dixon, "Wave-U-Net: A Multi-Scale Neural Network for End-to-End Audio Source Separation," 19th International Society for Music Information Retrieval Conference (ISMIR 2018), 334-340, 2018.

[20] A. Défossez, U. Nicolas, B. Léon and B. Francis, "Music Source Separation in the Waveform Domain," 1-16, 2019, arXiv preprint arXiv:1907.02404.

[21] Y. Luo, N. Mesgarani, "Conv-TasNet: Surpassing Ideal Time-Frequency Magnitude Masking for Speech Separation," IEEE/ACM transactions on audio, speech, and language processing **27**(8), 1256-1266, 2019, doi: 10.1109/TASLP.2019.2915167.

[22] C. Biemann "Chinese Whispers - an Efficient Graph Clustering Algorithm and its Application to Natural Language Processing Problems", Workshop on TextGraphs New York City, 73–80, June 2006.

[23] M. M. Awais; WaqasAhmad; S. Masud; S. Shamail, "Continuous Arabic Speech Segmentation using FFT Spectrogram", Innovations in Information Technology conference, 2006, doi: 10.1109/INNOVATIONS.2006.301939.

[24] D. Pani, A. Pani, L. Raffo, "Real-time blind audio source separation: performance assessment on an advanced digital signal processor," The Journal of Supercomputing **70** (3), 1555-1576, 2014, doi:10.1007/s11227-014-1252-4.

[25] Arwa-Data-Analytics, Machine-Learning-Algorithms-for-Real-Time-Blind-Audio-Source-Separation-with-Natural-Language-Detect, Retrieved from GitHub: https://github.com/ArwaDataAnalytics/Machine-Learning-Algorithms-for-Real-Time-Blind-Audio-Source-Separation-with-Natural-Language-Detect, July 23, 2021.

[26] "scipy.org," ENTHOUGHT, 26 Jul 2019. [Online]. Available: https://docs.scipy.org/doc/numpy/reference/generated/numpy.std.html. [Accessed 12 April 2019].

[27] W. Gragido, J. Pirc, D. Molina, Blackhatonomics, Science Direct .com, 2013.

[28] Vishal3096, "geeksforgeeks," geeksforgeeks, 2019. [Online]. Available: https://www.geeksforgeeks.org/scipy-stats-signaltonoise-function-python/. [Accessed 12 April 2019].

[29] A. HAYES, "investopedia," investopedia, 18 Mar 2020. [Online]. Available:https://www.investopedia.com/terms/r/r-squared.asp. [Accessed 14 April 2020].

[30] M. B. Editor, "Regression Analysis: How Do I Interpret R-squared and Assess the Goodness-of Fit?", 30 May 2013. [Online]. Available: https://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-do-iinterpret-r-squared-and-assess-the-goodness-of-fit. [Accessed 14 April 2020].

[31] T. Bock, "8 Tips for Interpreting R-Squared," display, 10 April 2020. [Online]. Available: https://www.displayr.com/8-tips-for-interpreting-r-squared/. [Accessed 14 April 2020].

[32] Stephanie, "What is Root Mean Square Error (RMSE)?," 25 October 2016. [Online]. Available: https://www.statisticshowto.com/rmse/. [Accessed 14 April 2020].

[33] K. GRACE-MARTIN, "Assessing the Fit of Regression Models," theanalysisfactor, [Online]. Available: https://www.theanalysisfactor.com/assessing-the-fit-of-regression-models/. [Accessed 14 April 2020].

[34] Craffel, "mir_eval github," Aug 2016. [Online]. Available: https://github.com/craffel/mir_eval/blob/master/mir_eval/transcription.py. [Accessed 11 March 2020].

[35] L. Roux, . J. S. Wisdom, H. Erdogan, "SDR – HALF-BAKED OR WELL DONE?" IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 626-630, 2019, doi: 10.1109/ICASSP.2019.8683855.

[36] J. M. Kates, "Cross-correlation procedures for measuring noise and distortion in AGC hearing aids," The Journal of the Acoustical Society of America, 107(6), 3407-3414, 2000, doi:10.1121/1.429411.

[37] S. Datasets, "#MUSDB18," Github, 9 10 2019. [Online]. Available: https://sigsep.github.io/datasets/musdb.html. [Accessed 27 Feb. 2020].

[38] K. amrr, "Noise reduction in python using spectral gating," Feb. 2020. [Online]. Available: https://github.com/timsainb/noisereduce. [Accessed March 2020].

[39] Nkundiushuti, "DeepConvSep," Github, 2015. [Online]. Available: https://github.com/MTG/DeepConvSep. [Accessed March 2020].

[40] lcolbois, "aasp-noise-red," 2019. [Online]. Available: https://github.com/lcolbois/aasp-noisered. [Accessed March 2020].

[41] Y. Tian, C. Xu and D. Li, "Deep Audio Prior - Pytorch Implementation," Jan. 2020. [Online]. Available: https://github.com/adobe/Deep-Audio-Prior. [Accessed March 2020].

[42] C. Barth, "Joint Audio Correction Kit (J.A.C.K.)," Github, 2019. [Online]. Available: https://github.com/cooperbarth/Joint-Audio-Correction-Kit. [Accessed March 2020].

Appendix

Training experiments:

Table 20: Execution time of Training Dataset of English languages

Experiments	Execution time of Demucs (ns)	Execution time of Conv-TasNet (ns)
Male & Music (M&M)	49.84469477	6.823944947
Female & Music (M&F)	51.39934172	6.003973062
Music & Conversation (M&C)	50.43340774	8.179383581
Music & Child (M&CH)	50.70981113	8.14998204

Table 21: Execution time of Training Dataset of Arabic languages

Experiments	Execution time of Demucs (ns)	Execution time of Conv-TasNet (ns)
Male & Music (M&M)	49.421875	8.90625
Female & Music (M&F)	53.703125	5.8046875
Music & Conversation (M&C)	55.8671875	5.4921875
Music & Child (M&CH)	57.109375	5.84375

Table 22: Execution time of Training Dataset of Chinese languages

Experiments	Execution time of Demucs (ns)	Execution time of Conv-TasNet (ns)
Male & Music (M&M)	50.421875	5.75
Female & Music (M&F)	50.890625	5.7265625
Music & Conversation (M&C)	49.8046875	6
Music & Child (M&CH)	61.8515625	5.5390625

Testing experiments

Table 23: Execution time of Testing Dataset English languages

Experiments	Execution time of Demucs (ns)	Execution time of Conv-TasNet (ns)
Male & Music (M&M)	50.24387019	6.735885814
Female & Music (M&F)	51.80440415	6.00008378
Music & Conversation (M&C)	50.3566125	8.143201463
Music & Child (M&CH)	50.78504406	8.641241776

Table 24: Execution time of Testing Dataset Arabic languages

Experiments	Execution time of Demucs (ns)	Execution time of Conv-TasNet (ns)
Male & Music (M&M)	50.0859375	8.7734375
Female & Music (M&F)	53.1015625	5.7578125
Music & Conversation (M&C)	55.4921875	5.53125
Music & Child (M&CH)	66.40625	5.828125

Table 25: Execution time of Testing Dataset Chinese languages

Experiments	Execution time of Demucs (ns)	Execution time of Conv-TasNet (ns)
Male & Music (M&M)	50.53125	5.9296875
Female & Music (M&F)	50.0546875	5.921875
Music & Conversation (M&C)	49.8046875	6.234375
Music & Child (M&CH)	61.8828125	8.669216296

Real-time Measurement Method for Fish Surface Area and Volume Based on Stereo Vision

Jotje Rantung*, Frans Palobo Sappu, Yan Tondok

Department of Mechanical Engineering, Faculty of Engineering, Sam Ratulangi University, Manado, 95115, Indonesia

A R T I C L E I N F O	A B S T R A C T
	In the automation of the fish processing industry, the measurement surface-area and volume of the fish requires a method that focuses on processing automation. The creation of a stereo-vision based on real-time measurement method is one of the most essential aspects of this work. To do this task, we completed two steps. The first, the acquisition of the image of the fish using a stereo camera and calibrating the image for size using sample of the image acquisition. Second, by applying image processing techniques and vision system, the fish surface area and fish volume is obtained in real-time. The experimental results of the proposed method have good results for fish surface area and fish volume. The measuring process using stereo-vision only takes a short time, making it suitable for the real-time method.

1. Introduction

Knowing the volume size is useful in the fish processing industry for size sorting, quality assessment, and microbial concentration [1]. After calculating the volume, other physical parameters such as mass and density are also easy to investigate. Surface area and volume are important physical parameters in fish[2,3]. Surface-area and fish volume are important parts in processing into fish products that must be known. In the salting process of fish, the area and thickness of the salt are affected by the surface area. Furthermore, understanding the surface area of fish is needed to calculate heat and mass transfer, determine other physical parameters including gas permeability, weight per unit surface area, and respiration rate. Fish-volume and other physical properties are appropriate for calculating water content, heat transfer, and respiration rates. Likewise, the volume of fish affects the rate of cooling and freezing, which is ultimately useful for determining the heat load on the cooling system and calculating cooling costs [4, 5].

Measurement of surface area and volume of fish is currently generally done manually by using the eyes of people or workers using a manual meter. Manual measurements can result in inaccurate, ineffective, and time-consuming measurements, especially when measuring fish in large numbers. The most commonly used for measuring surface area is the tape method, while the volume measurement used the water displacement method [6]. The tape method has drawbacks, such as timing problems, and may cause measurement errors due to human error. Similarly, because the fish has holes in its gills and mouth, volume measurement using the gas transfer method and the water displacement method is not practical for fish shapes. Analytical estimates based on principal dimensions and weight have been studied [7], but the results of this method are very time-consuming because the sensing process is carried out by human.

In the entire fish processing process, measuring the surface area and volume is critical. To increase production and minimize processing time and costs, the fish processing sector requires technology solutions that focus on processing automation. Innovative advances in image processing have empowered the use of new methods to quantify the surface region and volume of fish correctly, rapidly, and precisely. With the availability of image processing techniques, digital image analysis has begun to be used for simple measurement systems such as line measurements [8]. A partitioning technique utilizing the image processing has been completed to ascertain the surface region and volume of fish [9]. The working principle of this measurement method is offline. Real-time image segmentation to determine the coordinates of image objects has been carried out [10]. The segmentation principle employs stereo-vision as a measuring instrument, which is typically unstable and slow. With the availability of image processing methods and stereo-vision as measuring tools, the automation of the fish processing industry is feasible because to the development of a stereo-vision-based real-time measuring.

*Corresponding Author: Jotje Rantung, Faculty of Engineering, Sam Ratulangi University, Manado, 95115, Indonesia, +6285298392179, jrantung@unsrat.ac.id

Based on the previous description, the researcher is interested in investigating and developing new methods for measuring the surface area and volume of fish, specifically real-time measurements, and creating a stereo-vision-based instrumentation system for real-time measurements.

The main objective of this research is to develop a method of measuring the surface area and volume of fish using a stereo camera as a vision instrument so the measurements can be applied in real-time. The development is implemented by combining the ellipsoid approach method and the real-time image segmentation method using a stereo camera as stereo-vision.

2. Related Work

As previously described, the fish handling industry requires mechanical arrangements that attention on mechanization of preparing to expand efficiency and lessen handling time and expenses. Estimating the surface region and volume of fish is a significant stage in the entire process of the fish handling industry. In order to automate the process of measuring the surface area and volume of fish, new methods need to be developed. Due to technological advances in image processing methods and the principle of real-time image segmentation using stereo-vision, research on image processing and stereo-vision applications for measurement is increasingly being carried out.

The first step in real-time measurement using stereo-vision is segmenting the image of the object to be measured. Studies on image segmentation methods are still being carried out at present, all of which aim to improve the quality of the image segmentation of objects. The selection of automatic thresholding for image segmentation based on genetic algorithms was done in [11]. Research on image segmentation of fish objects using the K-means cluster enhancement algorithm to obtain fish body contours by separating the fish image from the background in complex background conditions has been carried out [12]. All the studies mentioned above showed satisfactory results. However, the proposed studies were limited to static imagery.

Researchers have conducted several preliminary investigations on fish image segmentation in addition to the studies described above. Research to determine the amount of fish skin injury has been proposed [13]. The proposed research was to realize the form of injury to fish based on the L*A*B* color space and the HSV color space. Subsequent research proposed a new approach to measuring 2D injury rates in fish with a modified K-means cluster algorithm based on the L*A*B* color space [14]. The experimental results of this study indicate that the proposed new approach is closer to the level of injury and actual injury to fish than the results of the manual threshold method on L*A*B* color images. The two studies above resulted in fish image segmentation, which effectively measures the level of injury to fish skin. However, in practice, it was still carried out offline. An offline image segmentation development is real-time image segmentation. The real-time image segmentation method for online measurement purposes was developed [10]. This real-time image segmentation method uses a stereo camera as stereo-vision to determine the coordinates of the image object. The emphasis was on determining the coordinate image of the object.

Image processing techniques are still being used to develop methods for measuring surface area and volume. Some of them was the determination of the volume and surface area of bubbles [15], estimating the volume and weight of apples using the 3D reconstruction method [16], and the determination of the surface area and volume of the fish using the ellipsoid approach method applied to image processing [9]. However, the determination and estimation of volume and surface area of image objects were still done offline in these studies. The strategy to decide surface region and volume of axisymmetric farming items was created [17]. In this study, the image captured by the CCD camera was processed using adobe photoshop. The use of an inexpensive 3D scanner to measure the surface area and volume of sweet potatoes was proposed [18]. The results of the method were satisfied for the identification of features related to shape using 3D scanner-based measurement.

In this study, we will develop a method of measuring the surface-area and volume of fish in real-time using a stereo camera as stereo-vision. Development is done by combining the ellipsoid approach method that researchers have reported in [9] and the real-time image segmentation method using a stereo-camera as stereo-vision reported in [10]. In the elliptical approach, the fish object is considered an ellipsoid shape by making partitions on three sides of the view according to the analytical model. Partitions on three sides are analyzed as image pixel shape are estimated as surface area and volume. The real-time segmentation process is carried out on a stereo-vision video frame that will capture the 3D coordinate value of the fish object to be measured. Real-time 3D coordinate values in the stereo-vision video frame corresponding to the image pixel values on the partitions of each viewpoint of the fish. The pixel value will be converted to millimeter size according to stereo-vision calibration.

3. Proposed Real Time Measurement Method

This section presents a real-time estimating technique for the surface region and volume of fish by using a vision system. The stereo camera mounted on the highest point of the fish object is used to capture the images. Image calibration for fish size was carried out using one image sample taken with a stereo camera. The stereo camera performs the segmentation procedure in real time on the video frame. The calibration of the stereo camera yielded intrinsic and extrinsic parameters. Open-source computer vision libraries (OpenCV, C++) were used to create an image segmentation algorithm that works in real-time. To extract objects in a "graph-based image," it utilizes a combination "HSV color space, threshold value, mathematical morphological transformations, and contour detection techniques." This research provides a method for performing real-time image segmentation. This process uses open-source computer vision (OpenCV, C++) to determine the surface area and volume of the fish. The utilization of a stereo camera to quantify the surface region and volume of fish in 3D directions is introduced as well in this paper. To test the validity of the proposed real-time measurement methodology, the experimental findings of the proposed real-time measurement technique are compared to those of the analytical measurement approach.

3.1. Stereo Camera Image Acquisition

The experimental arrangement for this proposed measuremet is shown in Figure 1. A stereo camera was utilized to capture the images in this study. Figure 2 depicts the target measurement experimental setup. The cross-line serves as a guide for precisely positioning the target fish. A stereo camera can automatically determine the fish's location.
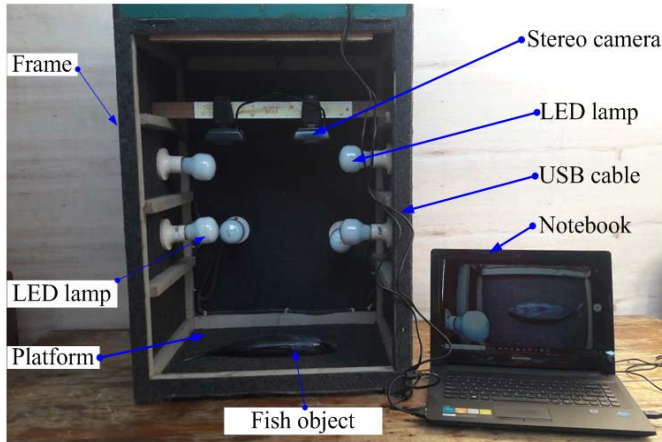
Figure 1: Experimental arrangement of proposed measurement
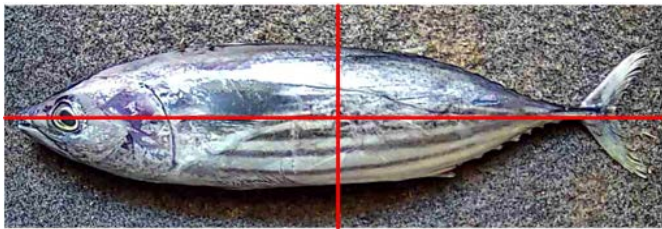
Figure 2: Experimental setup for target measurement

3.2. Image Calibration

A sample image taken with a stereo camera is used to calibrate the size of the fish image. A caliper was used to measure the physical dimensions of the fish in millimeters, as shown in Figure 3. The value of the calibration constant for the measurement of surface area and fish volume was based on the findings of image calibration of one fish object. The number of pixels from the outermost distance of the fish pictures is the unit of measurement in millimeters
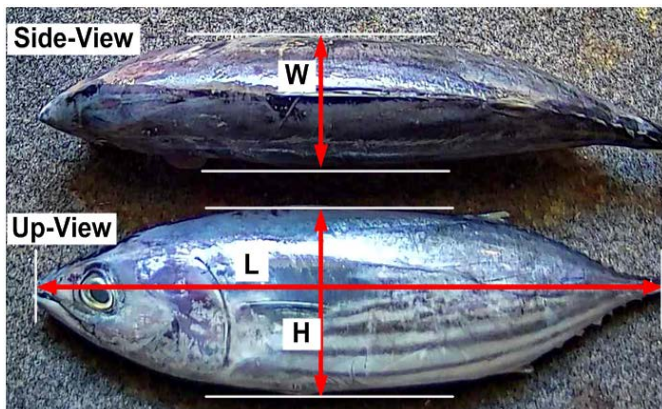
Figure 3: Fish measured by using micrometer

3.3. Partition for image processing method

Figure 4 depicts a fish surface region and volume picture preparation technique. Each circle area and volume may be calculated as follows:

$$S_i = \pi(b_i + c_i)\Delta t_i \qquad (1)$$

$$S = \sum_{i}^{n/2} S_i \quad \text{for } i = 2,3,\cdots,n/2 \qquad (2)$$
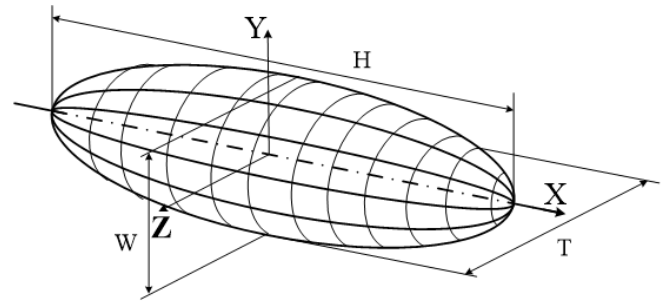
(b) Prolate spheroid orientation
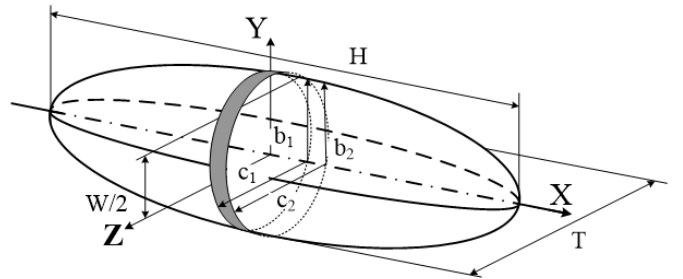
(b) General ellipsoid

Figure 4: Model used for image processing method

Area of each disc is calculated as average area of left and right planes, and dt_i is very thin and set as a pixel. The volume of each disc can be calculated as follows:

$$V_i = \pi\Delta t_i(b_i c_i) \qquad (3)$$

$$V = \sum_{i}^{n/2} V_i \quad \text{for } i = 2,3,\cdots,n/2 \qquad (4)$$

where $b_{1i} = W_{1i}/2$ and $b_{2i} = W_{2i}/2$ denote heights of the discs of the left and right planes in top view, $c_{1i} = H_{1i}/2$ and $c_{2i} = H_{2i}/2$ denote widths of the discs of left and right planes in side view, respectively, dt_i denotes as the thickness between disc of the left and right planes, and n denotes the number of boundary point of the fish contours.

3.4. Stereo Camera Calibration

The focal length of a stereo camera is an important parameter in a measurement algorithm. This value is obtained by calibrating a stereo camera. This parameter determines whether a camera

lens's ability to focus on an object through distorted images is strong or weak. There are four intrinsic parameters: f_x and f_y as the focal lengths of the camera in terms of pixel dimensions in the x and y direction, and (u_0, v_0) is the principal point. The camera usually represents lens distortion that is a radial distortion given as follows:

$$\mathbf{u}_d = \begin{cases} u_d = (u_u - u_0)(1 + k_1 r_u^2 + k_1 r_u^4) \\ v_d = (v_u - v_0)(1 + k_1 r_u^2 + k_1 r_u^4) \end{cases} \tag{5}$$

distortion-free is expressed as $p(u_u, v_u)$ and distortion-normalized image coordinates is expressed as $p(u_d, v_d)$. The radial distortion coefficients are expressed as k_1 and k_2, and $r_u^2 = u_u^2 + v_u^2$. The focal length is determined by the camera model.

$$f = \frac{1}{2}(f_x / m_x + f_y / m_y) \tag{6}$$

In terms of pixel size in the x and y directions, the focal length of the camera are denoted by f_x and f_y.
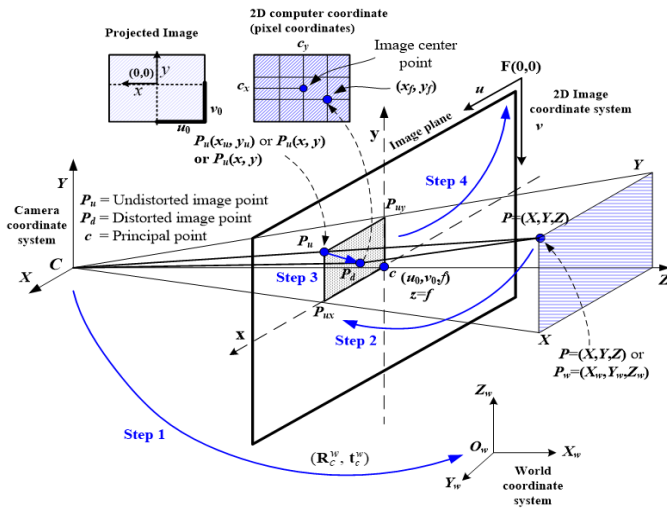
Figure 5: Steps of camera calibration

The most significant role of camera calibration is to identify the four intrinsic parameters and the two distortion coefficients.

The recommended calibration procedure as follows:

- Setup camera, print a pattern and attach it to a planar surface.
- Take a few images of the model plane under different orientations by moving either the plane or the camera.
- Detect the feature points in the images.
- Estimate the intrinsic and extrinsic parameters of the camera.

There are four steps to convert a point from the world coordinates to the computer memory image coordinate, shown in Figure 5.

3.5. Real-Time Image Segmentation

The segmentation process is done in real-time on the video frame by using a stereo camera. In reducing complexity and computation time, hue and value feature spaces are segmented. After that, they combine as a feature image segmentation. A stereo camera is applied to capture the images of the fish. In this work, use contour-based segmentation and mathematical morphological method for real-time segmentation. Image segmentation is done by combining the *HSV* color space, threshold, mathematical morphological transformations, and contour detection techniques to extract objects in graphics-based images. Software design is implemented by using the C++ programming language. A library available in open-source computer vision (OpenCV, C++) is used to implement real-time image segmentation by loading images, creating windows to hold an image in real-time, and saving the image.

3.6. 3D Measurement

The 3D position is obtained from stereo triangulation. First, two images are obtained from a set of three-dimensional test points whose three-dimensional coordinates are known. Secondly, the estimated 3D coordinate of the same points is computed from their projections using the calibrated parameters. Finally, the discrepancy between real and estimated positions is compared. In this case, the accuracy depends on the calibration of both cameras. Figure 6 shows the principle of 3D measurement using a stereo camera. The projective transformation for the basic stereo camera image is sought with the epipolar constraint that the epipolar line is horizontal, and 3D measurements are made from the information about the corresponding point by a stereo camera. The pairs of baseline stereo images are generated from ordinary images with the projective transformation of the axes of X, Y, and Z.
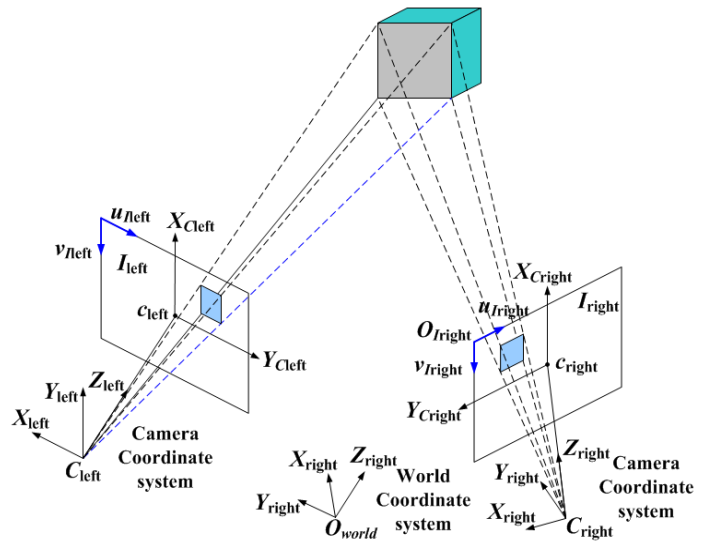
Figure 6: 3D measurement principle using stereo camera

The triangulation of stereo camera model is shown in Figure 7. Generally, stereo systems may have optical axes with the fixation point at a finite distance from the cameras. The left and right image planes are represented by the segments I*left* and I*right*, respectively, and O*l* and O*r* are the centers of projection, or optical centers in the left and right of the camera, respectively. Because the optical axes are parallel, their point of centroid called the fixation point lies intimately far from the cameras.
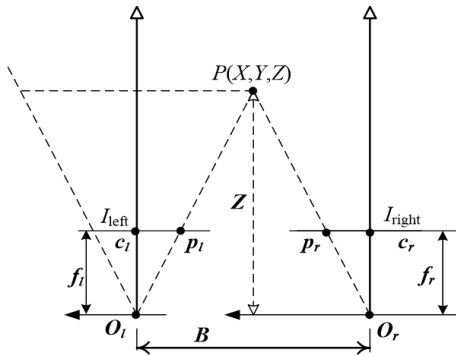
Figure 7: Triangulation of stereo camera mode

By assuming that the 3D coordinate frame has its origin in the optical center of the left camera, the perspective projection from the 3D camera coordinate (X, Y, Z) to the ideal image coordinate (x, y) is as follows:

$$\begin{cases} x_l = \dfrac{X \cdot f}{Z} \\[2mm] x_r = \dfrac{(X - B) \cdot f}{Z} - B \\[2mm] y_l = y_r = \dfrac{Y \cdot f}{Z} \end{cases} \quad (7)$$

From Equation (7), the Z coordinate can be determined as follows:"

$$Z = f \cdot \frac{B}{B - d} \quad (8)$$

where $x_l = p_l\text{-}c_l$, $x_r = p_r\text{-}c_r$, and $d = x_r\text{-}x_l$ is a disparity that is the difference in image position between corresponding points in the two images. Once Z is determined, it is straightforward to calculate X and Y using similar triangles:

$$X = x_l \cdot \frac{B}{B - d} \quad (9)$$

$$Y = y_l \cdot \frac{B}{B - d} \quad (10)$$

Following the segmentation procedure, the 3D object coordinates are determined. "First, the object is recognized to determine its center, and then the principle of determining 3D coordinates is implemented. The following is how the distance error rate is calculated.

$$e_d = \left(\frac{Z - \text{real distance}}{\text{real distance}} \right) \cdot 100\% \quad (11)$$

3.7. Fish Surface Area and Volume Measurement

Figure 8 shows the flow chart of the software developed to calculate the surface area and volume of the fish. The process

begins with obtaining two synchronous images of the cameras and their subsequent correction using the parameters obtained from the calibration. Once the images are corrected, the fish real-time image segmentation algorithm is executed individually for the up-view image and side-view image. It is done for the up-view image to obtain fish length and width in pixels, and the same process is done for the side-view image to obtain fish length and height in pixels.

The first process is done for the up-view image so that length (L) and width (W) are obtained, and these results are stored as a saved model. The length (L) and height (H) is calculated in a side view, and the result is saved as a saved model. For the last process, the values of LWH are used for the calculation process of surface area and volume.
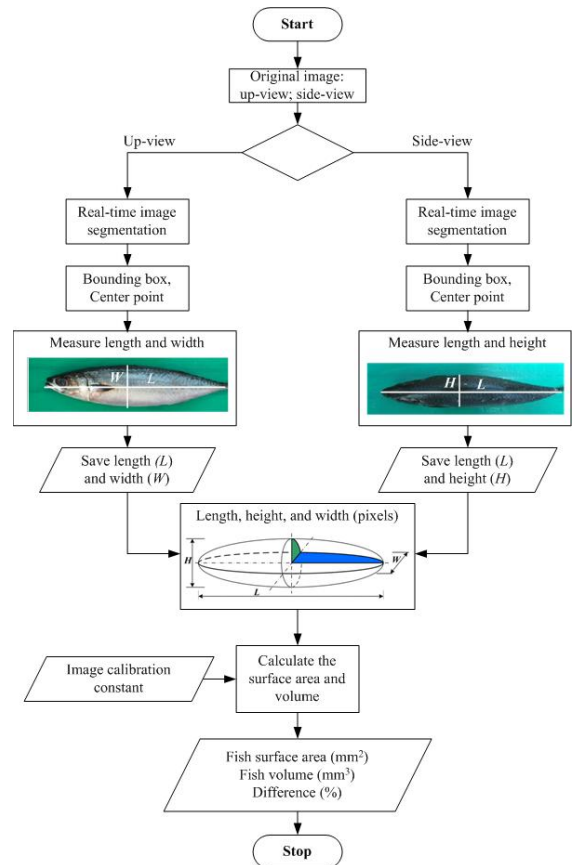
Figure 8: Flowchart to calculate fish surface are and volume

4. Experiment Results

4.1. Experiment Result for Image Calibration and Stereo Camera Calibration

In this experiment, for image calibration, we used a vernier caliper to measure the physical dimension of the fish in millimeters. We use the image calibration result of one fish object as a calibration constant value for measuring fish surface area and volume measurement. The size in millimeters is obtained by measuring the number of pixels from the outermost distance of the fish images. Figure 3 shows the fish image processing results used for image calibration. An example of one fish has a dimension of 290 mm (L) x 51.1 mm (H) x 38.22 mm (W), and the calibration result is obtained as 1 pixel = 0.0251 mm for (L), 1 pixel = 0.00557 mm for (H), and 1 pixel = 0.00475 mm for (W).

Camera intrinsic parameters are calibrated at a working distance of about 1000 mm. The chessboard image is used as a calibration pattern, and a 19" LCD monitor is used to display the image pattern. Table 1 shows the calibration results of the left camera and right camera, respectively.

Table 1: Calibration results of the intrinsic camera parameter

Camera	Focal length (pixel)		Principal point (pixel)	
	f_x	f_y	u_0	v_0
Left	940.49	769.00	674.65	180.35
Right	941.49	769.50	675.65	181.25

Camera	Distortion coefficients		pixel/mm		Focal length (mm)
	k_1	k_2	P_{ux}	P_{uy}	
Left	-98.91	11.34	674	180	2.8289
Right	-99.01	11.84	675	181	2.8311

4.2. Experiment result for Real-Time Image Segmentation

The proposed segmentation method is designed to detect fish objects useful for fish processing. The proposed image segmentation method is applied to different scanning objects before it is applied to the fish. Various common objects, such as simple objects, small and large objects, objects of different shapes, and various objects, are used in this experiment. Once the stereo camera finds the center position of the image fish it wants to measure, the open-source computer vision will segment the fish image, calculating it to prepare for real-time surface area and volume measuring.

The vision targets tested in the experiments are fish as a real detected vision object. Figure 9 shows fish target detection. Figure 9(a) shows the original RGB color image, Figure 9(b) shows the threshold image, and Figure 9(c) shows a segmented image with a centroid position.

(a) Thresholded image

(b) Original RGB color image with the contour image
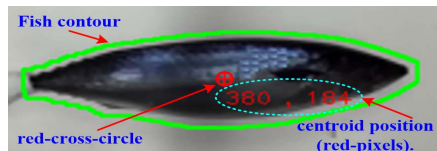
(c) Original RGB color image with the contour image and bounding box

Figure 9: Fish target detection

The image processing stage involves image segmentation for the fish object target, the object centroid, and the bounding box in pixels. The white background used in the experimental setup simplifies the background detection procedure in both RGB color images obtained with a stereo camera.

The results of segmenting fish objects using the lowest threshold (Th = 0) and a bounding box are shown in Figure 10. The best results were found in this research, as shown in Figure 10(a). The feature segmentation findings from this experiment will be used in the next step because of their high accuracy. The results of feature segmentation in Figure 10 are used in the next test. By applying the graph cut algorithm, the results are shown in Figure 11. Finally, image segmentation is obtained from the object's bw image, as shown in Figure 11(a).

(a) Thresholded image
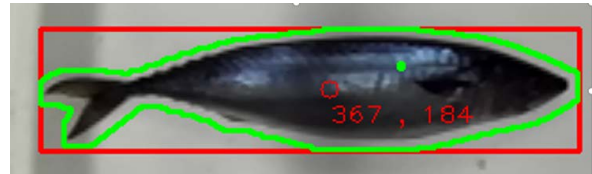
(b) Original RGB color image

Figure 10: Segmentation of the fish object with Th = 0 and bounding box

(a) Thresholded image

(b) Original RGB color image with the contour image

(c) Original RGB color image with the contour image and bounding box

Figure 11: Segmentation of the fish object with Th=0

4.3. Experimental results of fish surface area and volume measurement

Figure 12 shows the representation of the right camera and the left camera in real-time image segmentation. The white area of the fish threshold image in Figure 12(a) is represented by the inner area of the green line in the original RGB color image of the fish in Figure 12(b). The white area in the threshold image and the green line in the original RGB color image are visible for the left camera and right camera. In other words, image segmentation works well. It is also indicated by the bounding box in the green

line area, the center point of the original left-right image (red), and the center point between the left camera and the right camera.

(a). Left and right thresholded images
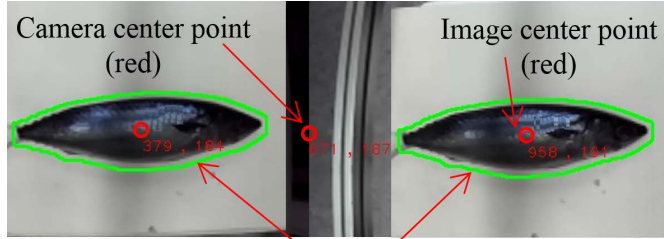
(b). Left and right original RGB color images
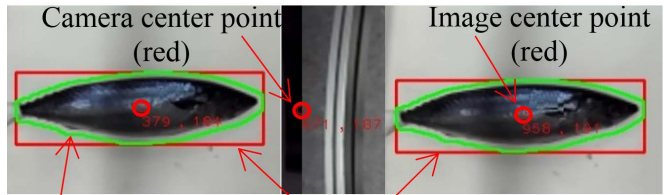
(c). Left and right original RGB color images with bounding box

Figure 12: Representation of the right camera and left camera

The length, width, and height of the bounding box (L, W, and H) correspond to the fish image's length, width, and height in pixels. Because calculating the surface area and volume can be limited to one image, the centroid will be calculated for both images obtained with the stereo camera system. The thresholded image of the target analyzed in one experiment, as shown in Figure 12(a). Figure 12(b) shows the results of the detection of fish to be analyzed, namely the location of the centroid (red circle) and the contour (green line). Figure 12(c) shows the results of fish detection, which will be used to calculate the surface area and volume of fish. The result obtained in this case shows the centroid location (red circle), contour (green line), and bounding box in the stereo camera image. One example of the object experiment result is shown in Figure 12(b) and Figure 12(c). In this result, the image centroid for the left camera in pixels is (379,184), the center of the image for the right camera in pixels is (958,191), and the image center is for stereo cameras in pixels (671,187). The distance between the planes of the two cameras and the plane of pixels can be determined using this value.

The 2D image coordinates (pixels) of the center of mass in a stereo camera are shown in Table 2. Table 3 shows the results of the 3D coordinates (mm). Based on experimental results for 3D coordinate measurements, the least distance error rate, as listed in Table 3, is at a distance of 550 mm. The distance between the stereo camera and the object will be adjusted to 550 mm to measure the surface area and volume of the fish. The results of using stereo-vision to measure the surface area and volume of four fish in real-time are shown in Table 4. In Table 4, the measurement results

using the analytical method are used as a comparison to obtain the effectiveness of the proposed real-time measurement method.

Table 2: Image coordinate of the centroid point

Trial	Centroid in 2D coordinate (pixels)		Real distance Z (mm)
	X	Y	
1	687	203	780
	680	194	660
	671	187	550
2	687	203	780
	680	194	660
	671	187	550
3	687	203	780
	680	194	660
	671	187	550

Table 3: Real-time measurement results of surface area and volume of four fishes

Trial	Centroid in 3D coordinate (pixels)			Real distance Z (mm	Distance error rate e_d (%)
	X	Y	Z		
1	16	20	793.2	780	1.69
	16	20	667.7	660	1.17
	16	20	553.4	550	0.62
2	16	20	792.9	780	1.65
	16	20	667.5	660	1.14
	16	20	553.2	550	0.58
3	16	20	791.6	780	1.49
	16	20	667.1	660	1.08
	16	20	552.6	550	0.47

Table 4: Real-time measurement results of surface area and volume of four fishes

	Fishes sample size (mm)			Surface area (mm²)		
	L	H	W	Real-time method	Analytic method	Error e_d (%)
1	290	38.22	51.1	173,159.92	164,428.75	5.31
2	287	37.12	49.0	167,981.93	160,028.51	4.97
3	289	38.20	50.5	170,715.11	163,020.54	4.72
4	287	37.00	48.9	166,746.44	159,963.96	4.24

	L	W	H	Volume (mm³)		
1	290	38.22	51.1	311,256.65	296,406.67	5.31
2	287	37.12	49.0	286,384.78	273,189.71	4.83
3	289	38.20	50.5	305,330.52	291,763.51	4.65
4	287	37.00	48.9	282,648.04	271,750.83	4.01

5. Conclusions

A method for measuring fish surface area and volume in real-time using a stereo camera as a stereo-vision was proposed. The measurement process was done in real-time on the video frame. Image object segmentation is done first before measurement. The hue (H), saturation (S), and value (V) were separately segmented before they are combined. The result of this segmentation was the targeting object.

Experiment results show that real-time image segmentation of the proposed method had a good result. The experiments demonstrated that the calibration process could quickly detect the chessboard corners. After the calibration results, the focal lengths of the left camera and right camera were about 2.8289 mm and 2.8311 mm, respectively. The focal length difference between the two cameras was about 0.00221 mm. For desired object distances of 780 mm, 660 mm, and 550 mm, the distance error rate was less than 2% for both distances of 650 mm and 550 mm in three trial times. However, for 780 mm, the distance error rate was bigger than the two distances of 550 mm and 660 mm and was more than 3%. The 3D coordinate measurement results revealed that increasing the distance increased the distance error in the Z coordinate, which was caused by the camera's vision field of view. As a result, a distance of 550 mm is used in the next analysis, which involves measuring the surface area and volume of the fish.

The surface area and volume of fish were measured by the proposed real-time measurement and compared to the analytic measurement method. Experimental results using a sample of four fishes show that the differences in surface area and volume were 4.24%~5.31% and 4.01%~5.01%, respectively. The process of real-time image segmentation, feature extraction, and measurement of fish surface area and volume takes about 0.0018 milliseconds. The results show that the proposed method can accurately measure the surface area and volume of fish in real-time.

The real-time image segmentation, 3D information, and measurement methods for surface area and volume proposed in this research could be applicable in fish recognition and sorting applications. The 3D information-based stereo camera could be applied to an automated fish processing system to handle fish on a conveyor belt.

The measurement method developed by this research can be utilized to automatically measure the surface area and volume of fish using stereo-vision. Furthermore, by employing this approach, measuring will become easy, faster, more effective, and efficient. The findings of this study can also be accustomed to other fields of research, such as real-time assessment of fruit product dimensions, fish sorting processes on moving conveyors, and moving goods robots.

Acknowledgment

References

[1] E.S. Bridge, R.K. Boughton, R.A. Aldredge, T.J.E. Harrison, R. Bowman, S.J. Schoech, "Measuring egg size using digital photography: Testing Hoyt's method using Florida Scrub-Jay eggs," Journal of Field Ornithology, **78**(1), 109–116, 2007, doi:10.1111/J.1557-9263.2006.00092.X.

[2] R. Badonia, A. Ramachandran, T. V Sankar, "Quality Problems in Fish Processing," Journal of the Indian Fisheries Association, **18**, 283–287, 1988.

[3] A. Getu, K. Misganaw, "Post-harvesting and Major Related Problems of Fish Production," Fisheries and Aquaculture Journal, **06**(04), 2015, doi:10.4172/2150-3508.1000154.

[4] D. J. Lee, X. Xu, J.D. Eifert, P. Zhan, "Area and volume measurements of objects with irregular shapes using multiple silhouettes," Optical Engineering, **45**(2), 027202, 2006, doi:10.1117/1.2166847.

[5] A.B. Koc, "Determination of watermelon volume using ellipsoid approximation and image processing," Postharvest Biology and Technology, **45**(3), 366–371, 2007, doi:10.1016/J.POSTHARVBIO.2007.03.010.

[6] S.M.A. Moustafa, "Theoretical Prediction of Volume, Surface Area, and Center of Gravity for Agricultural Products," Transactions of the ASAE, **14**(3), 549–0553, 1971, doi:10.13031/2013.38336.

[7] T.Y. Wang, S.K. Nguang, "Low cost sensor for volume and surface area computation of axi-symmetric agricultural products," Journal of Food Engineering, **79**(3), 870–877, 2007, doi:10.1016/J.JFOODENG.2006.01.084.

[8] F. Pla, J.M. Sanchiz, J.S. Sánchez, "An integral automation of industrial fruit and vegetable sorting by machine vision," IEEE International Conference on Emerging Technologies and Factory Automation, ETFA, **2**, 541–546, 2001, doi:10.1109/ETFA.2001.997731.

[9] J. Rantung, M.T. Tran, H.Y. Jang, J.W. Lee, H.K. Kim, S.B. Kim, "Determination of the Fish Surface Area and Volume Using Ellipsoid Approximation Method Applied for Image Processing," Lecture Notes in Electrical Engineering, **465**, 334–347, 2017, doi:10.1007/978-3-319-69814-4_33.

[10] J. Rantung, J.M. Oh, H.K. Kim, S.J. Oh, S.B. Kim, "Real-Time Image Segmentation and Determination of 3D Coordinates for Fish Surface Area and Volume Measurement based on Stereo Vision," Journal of Institute of Control, Robotics and Systems, **24**(2), 141–148, 2018, doi:10.5302/J.ICROS.2018.17.0213.

[11] 병 룡이, Q. Bao Truong, V. Huy Pham, 김 형 석 Byung-Ryong Lee, H.-S. Kim, "Automatic Thresholding Selection for Image Segmentation Based on Genetic Algorithm," **17**(6), 587–595, 2011, doi:10.5302/J.ICROS.2011.17.6.587.

[12] H. Yao, Q. Duan, D. Li, J. Wang, "An improved K-means clustering algorithm for fish image segmentation," Mathematical and Computer Modelling, **58**(3–4), 790–798, 2013, doi:10.1016/J.MCM.2012.12.025.

[13] M.T. Tran, H.H. Nguyen, J. Rantung, H.K. Kim, S.J. Oh, S.B. Kim, "A New Approach of 2D Measurement of Injury Rate on Fish by a Modified K-means Clustering Algorithm Based on L*A*B* Color Space," Lecture Notes in Electrical Engineering, **465**, 324–333, 2017, doi:10.1007/978-3-319-69814-4_32.

[14] M.T. Tran, J. Rantung, T.H. Nguyen, H.K. Kim, S.B. Kim, "Measurement of injury rate on fish skin and performance comparison based on L*A*B* and HSV color spaces," Aug. 2021, doi:10.1051/matecconf/201815902010.

[15] J. Wen, Q. Sun, Z. Sun, H. Gu, "An improved image processing technique for determination of volume and surface area of rising bubble," International Journal of Multiphase Flow, **104**, 294–306, 2018, doi:10.1016/J.IJMULTIPHASEFLOW.2018.02.004.

[16] B. Zhang, N. Guo, J. Huang, B. Gu, J. Zhou, "Computer Vision Estimation of the Volume and Weight of Apples by Using 3D Reconstruction and Noncontact Measuring Methods," Journal of Sensors, **2020**, 2020, doi:10.1155/2020/5053407.

[17] C.M. Sabliov, D. Boldor, K.M. Keener, B.E. Farkas, "Image Processing Method to Determine Surface Area and Volume of Axi-Symmetric Agricultural Products," International Journal of Food Properties, **5**(3), 641–653, 2002, doi:10.1081/JFP-120015498.

[18] A. Villordon, J.C. Gregorie, D. LaBonte, "Direct Measurement of Sweetpotato Surface Area and Volume Using a Low-cost 3D Scanner for Identification of Shape Features Related to Processing Product Recovery," HortScience, **55**(5), 722–728, 2020, doi:10.21273/HORTSCI14964-20.

A S T E S

Coupled Apodization Functions Applied to Enhance Image Quality in Ultrasound Imaging using Phased Arrays

Wided Hechkel[1,*], Brahim Maaref [2], Néjib Hassen[1]

[1]*Microelectronic and Instrumentation Laboratory, FSM, University of Monastir, Monastir, 5000, Tunisia*

[2]*Electronic and Microelectronic Laboratory, FSM, University of Monastir, Monastir, 5000, Tunisia*

A B S T R A C T

The remarkable presence of side lobes levels in ultrasound B-mode imaging significantly decreases the image quality. Therefore, the use of an apodization function is of great importance. Linear windowing functions are among the most efficient techniques used to optimize antenna directivity by suppression of the side lobes. However, the apodization causes the degradation of lateral resolution by elevating the main lobe width. A new apodization approach that couples a linear windowing technique, and a non-linear windowing technique has been suggested. In this paper, a new dynamic apodization technique called "Dynamic Triangular Apodization" (DTA) is proposed. It enables non-linear windowing for each imaging direction. In addition, the effectiveness of a hybrid method that combines the Hanning window with the DTA algorithm is evaluated. In order to validate the proposed method, two types of simulation are carried out; namely point spread function and cyst phantom simulation. In the point spread function simulation, the main lobe width of the dual-apodization algorithm is very similar to that of the rectangular window and at the same time, the side lobes levels are significantly reduced. In the cyst target simulation, the Contrast to Noise Ratio (CNR) values of the dual apodization are significantly improved. The performance of the non-linear apodization is numerically investigated. In comparison with the rectangular window, the non-linear apodization method called DTA-Hann maintains low side lobes levels without altering the main lobe width. Consequently, it is a promising technique that aims to ameliorate the image quality.

1. Introduction

Performed by a professional, ultrasound imaging is virtually harmless: it is the only technique that provides an image of the fetus safely and it can be repeated without any problem. Ultrasound is relatively an inexpensive medical imaging technique: it requires only one machine and the price of consumables is negligible.

The ultrasound scanner is portable, allowing the examination to be performed at the patient's bed, in an intensive care unit for example. It is non-irradiating. It is among rare real-time imaging techniques, which offer complete clinical diagnosis of the patient during examination. It allows the use of several modalities to specify an anomaly: 2D, 3D, planar reconstructions, pulsed or colour doppler, etc.

The kernel of the ultrasound imaging system is the beamforming block. It defines the precision and resolution of the final output image. The beamforming stage has two functions: the first deals with the delay of the signals to enable steering and focusing in transmission and reception. The second function consists of the design of an appropriate apodization function to apply corresponding weighting coefficients to the transmission and reception channels. The main purpose of apodization is to increase the energy of the main lobe and eliminate the energy of side lobes.

To eliminate side lobes levels, researchers suggested multiplying transmission ultrasound waves and reception radio-frequency data by weighting coefficients. This technique was called apodization and it was very recommended to enhance ultrasound image resolution [1]. In [2], the authors proposed to reduce the side lobe levels by combining two transmitted waveforms. They are focused on the same focal depth. The first was generated by the odd transducer elements and had a single

*Corresponding Author: Wided Hechkel, Email: hechkelwided@gmail.com

focus. The second was generated by the even transducer and had two foci. The interlacing of the two shapes suppressed the side lobes of the transmitted wave. In [3] the authors introduced the concept of Null subtraction imaging (NSI). This technique applies three apodization functions to the received echoes. As a result, side lobe levels were largely reduced and consequently, a great improvement of the image quality was shown. In [4] the authors used a combination of two windowing functions to amplify received ultrasound waves. This approach was applied for vector flow ultrasound imaging. In [5] the authors proposed the use of the apodization stage after the delay beamforming stage. Their method was applied to the Synthetic Transmit Aperture (STA) technique. As a result, it enhanced the resolution and contrast. In [6] the authors applied a non-linear apodization function for real time 3D ultrasound imaging. They weighted signals obtained from Rectangular Boundary Array (RBA). In [7] the authors created a new computer aided software for the ultrasound image quality assessment. Images obtained from different apodization filters were compared. Their technique contributed to the amelioration of ultrasound images. In [8] the authors introduced the concept of singular value decomposition (SVD). It consists of an apodization filter applied to a sequence of images. The images have the same background signals. They were processed in a very short time. The best image quality was then selected, and the others were neglected. This technique was used for coherent plane wave compounding (CPWC) imaging. In [9] the authors, created a new apodization filter and they used it in a passive ultrasound mapping beamformer. The filter was obtained by the convolution between two non-linear tapering functions. In [10] the authors proposed a new side lobe suppression technique. The adoption of this apodization method improves the frame rate of the system and enables plane wave compounding imaging. In [11] the authors introduced the expanding apodization approach with 3D ultrasonography. This method uses a larger memory storage than static apodization techniques, but it leads to a significant improvement of the final image quality. In [12] the authors evaluated the compromise between computation cost and the type of the linear apodization window used to acquire the image. Authors have also used the Hanning window in the beamforming stage for a new ultrafast imaging technique [13,14] . Apodization was used in the most recent beamforming systems [15–18].

Although the apodization methods mentioned above enhance image quality in terms of contrast resolution and spatial resolution, they suffer from multiple deficiencies. The major drawbacks are the complex computational algorithms, the hard circuitry implementation, the high memory requirement and sometimes the lack of efficiency due to the less dynamic selection of the apodization window.

This paper is an extension of the work originally presented in "2020 17th International Multi-Conference on Systems, Signals _ Devices (SSD)" [12]. In this paper, a non-linear Dynamic Triangular Apodization (DTA) window is proposed. It applies specific weights to the multi-element transducers for each imaging point. The window function is obtained using a triangular shape where the triangle base is the aperture of transmitted/received beams and its top is the imaging point itself. This technique provides an acceptable compromise between image quality and computational complexity while keeping a very simplified algorithm for implementation. The use of a dual apodization

algorithm that combines DTA with Hanning window in the original work led to enhanced image quality.

The following section introduces the idea of Dynamic Triangular Apodization (DTA) technique and establishes the theoretical basis in the context of side lobes levels. It presents afterwards the mathematical algorithm based on the approach of dynamic directivity of the apodization window. Material and methods are presented in section 3. Simulation results of PSF (Point Spread Function) and CP (Cyst Phantom) are presented in section 4. Both simulations are performed using DTA, DTA-Hann and conventional apodization windows such as Hanning and rectangular techniques. Section 5 discusses the DTA, DTA-Hann functions in comparison with other windows. The inferences and conclusions along with the future scope of study are presented in Section 6.

2. Materials and Methods

2.1. Side lobes levels

The heterogeneity of the near field generates complex acoustic phenomena such as diffraction, which creates "side lobes". These acoustic phenomena also result in aberrations of the ultrasound beam, in the form of "side lobes" arranged in concentric cones around the axis of the wave. Consequently, artifacts may be generated when the ultrasound is reflected by targets located in the path of these side lobes, outside the axis of the beam. Although the acoustic energy is less intense in side lobes than the main lobe, the presence of these beam imperfections causes "acoustic noise" that degrades the ultrasound image quality (Figure 1.a.).

The "side lobes" transmit a part of the acoustic energy to both sides of the wave axis according to various angles. The echoes reflected by these side lobes result in the appearance of noise in the image, altering the effective contrast. The orientation of these side lobes in relation with the axis of the beam is a function of the wavelength and an inverse function of the probe diameter.

Smaller transducer elements or transducer functional groups have low directivity, especially when the beam is tilted by phase shift. The smaller the group of piezoelectric elements (i.e. the smaller the size of the probe), the greater the importance of side lobes. It is even greater when the beam is oblique (Figure 1.b.).
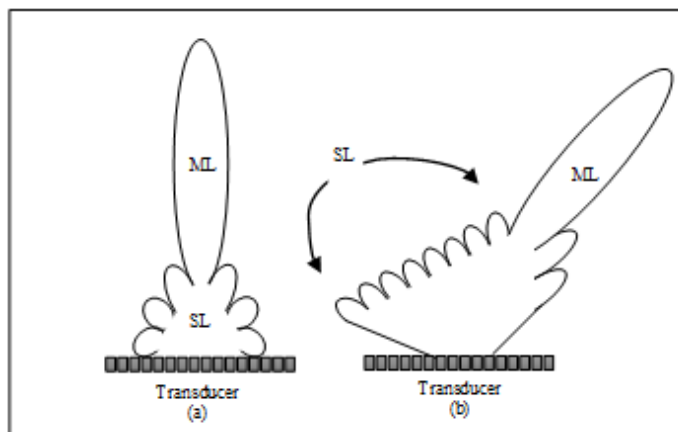
Figure 1: (a) Main Lobe (ML) and Side Lobe (SL) of a vertical beam. (b) Main Lobe (ML) and Side Lobe (SL) of a directed beam.

2.2. Dynamic Triangular Apodization (DTA)

The proposed algorithm calculates the apodization values for each channel of every single imaging point, i.e. at each scan angle and each scan depth. In order to combine both windowing and dynamic aperture growth, the algorithm first calculates the required aperture described by three parameters, H, K and a which are derived from the Cartesian coordinates of the adequate point and the width of the transducer. The second part of the algorithm basically fits a Triangular window inside the aperture, while granting the multi-element transducer $(0, K)$ with the maximum weight. zeros are simply assigned to the elements falling outside the aperture. After compressing the apodization values in $[0.5, 1]$ range, all the points that was located at the same direction, will exhibits the same DTA window, hence, the same apodization values. Figure 2 shows the apodization window directivity for two angles, 0° and 5°.
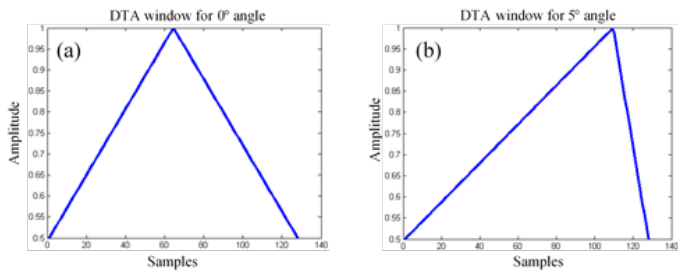
Figure 2: (a) Shape of the DTA window for 0° angle (b) Shape of the DTA window for 5° angle

The requested algorithm must find for each multi-element transducer the adequate apodization value. The apodization window is modelled with a triangle. The point M with coordinates (H, K) of plan P, is the top of this triangle. It is also the targeted point to be scanned. The width of the transducer is the base of the triangle. Therefore, for each point of the plan P, the algorithm provides a new triangular window, and hence new apodization values for every multi-element transducer. The working principle is illustrated in Figure 3.
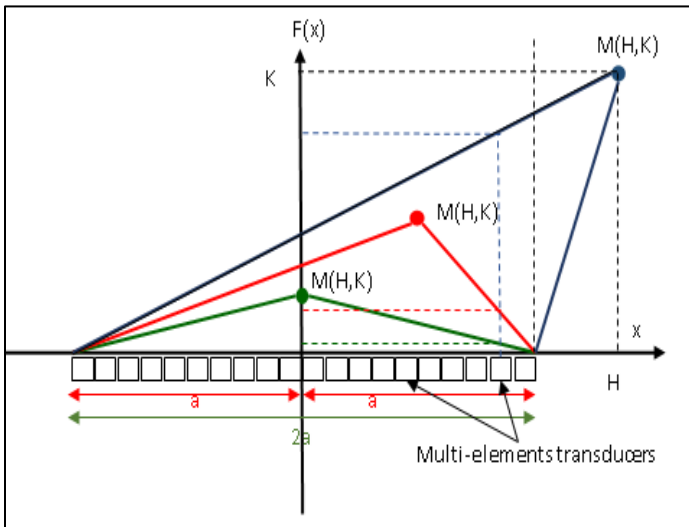
Figure 3: Schematic of the dynamic shape of the DTA window

The algorithm solve the problem as follows:

Algorithm 1: DTA algorithm

```
if K ≥ 0 then
    if H ≥ 0 then
        if H = 0 then
            if X ≥ 0 then
                F(X) = (-K / a).X+K;
            else if X < 0 then
                F(X) = (K / a).X+K;
            end
        else if H < a then
            if X ≤ H then
                F(X) = (K/(H+a)).X+(a.K)/(H+a);
            else if X > H then
                F(X) = (K/(H-a)).X+(a.K)/(a-H);
            end
        else if H ≥ a then
                F(X) = (K/(H+a)).X+(a.K)/(H+a);
        end
    else if H < 0 then
                G(X) = - F(X);
    end
end
```

where:

- H is the abscissa of the point to be processed
- K is the ordinate of the point to be processed
- X is the abscissa of the transducer elements (referring to their length and position)
- a is the length of half aperture of the transducer

In order to simplify the implementation, the proposed algorithm was alleviated by reducing computational complexity and avoiding potential cumbersome circuits. Therefore, the whole apodization equations mentioned above can be summarized in only one equation described by the following code.

$$F(X)=(K.X)/(H-sign(X-H).a)+(a.K)/(a-sign(X-H).H) \quad (1)$$

- If H = 0 then,

$$F(X) = \frac{K.X}{-sign(X).a} + \frac{a.K}{a}$$

$$= \begin{cases} \frac{K.X}{-a} + \frac{a.K}{a}, If X \geq 0, \\ \frac{K.X}{a} + \frac{a.K}{a}, If X < 0, \end{cases}$$

$$= \begin{cases} \frac{-K.X}{a} + K, If X \geq 0, \\ \frac{K.X}{a} + K, If X < 0, \end{cases}$$

- If H < a then,

For $X \leq H$, the obtained equation is:

$$F(X) = (K.X)/(H + a) + (a.K)/(a + H)$$

For $X > H$, the obtained equation is:

$$F(X) = (K.X)/(H - a) + (a.K)/(a - H)$$

- If $H \geqslant a$ then $H \geqslant X$. Consequently:

$$F(X) = (K.X)/(H + a) + (a.K)/(a + H)$$

where the sign function is defined for all real t by:

$$sign\ (t) = \begin{cases} 1, if\ t > 0, \\ 0, if\ t = 0, \\ -1, if\ t < 0, \end{cases}$$

In order to keep the apodization values within the interval [0.5, 1], the final equation Y(X) of the DTA algorithm becomes:

$$Y(X)=F(X)/(2*K)+0.5 \qquad (2)$$

2.3. DTA-Hann approach

To take the advantages of both non-linear and linear apodization techniques, the combination of DTA with the Hanning window was suggested. As a consequence, apodization values for the different multi-element transducers for the same aperture are obtained from DTA and Hanning window, each of them contributing by 50%. The new approach is called DTA-Hann. Figure 4 shows the apodization window directivity obtained with DTA, Hanning and DTA-Hann for 128 multi-element transducers.

The Hanning window is defined by:

$$W(X) = 0.5 - 0.5*cos\ (2*pi*X/N),\ X = 0,\ 1,...,N-1 \qquad (3)$$

Where N is the number of multi-element transducers.

Therefore, $D(X) = (W(X) + Y(X)\ )\ /\ 2$, is the final equation of the DTA-Hann algorithm.
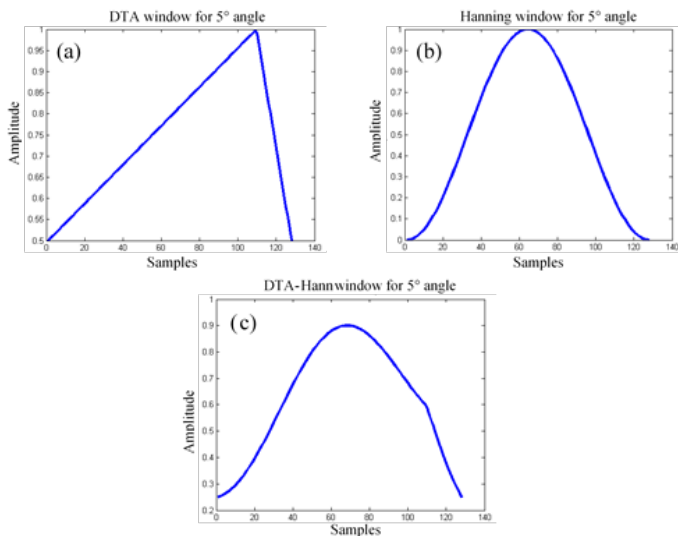
Figure 4: (a) Shape of the DTA window for 5° angle (b) Shape of the Hanning window for 5° angle (c) Shape of the DTA-Hann window for 5° angle

2.4. Simulation Setup Configuration

Each of the point spread functions and the cyst phantoms are modelled. All the simulations are implemented in a virtual ultrasound scanning software called Field_II. In the ultrasound software simulator [10], RF (Radio Frequency) echo data are produced using a 1D phased array transducer, which is composed of 128 elements. Each element is characterized by a length of 193 μm, a height of 5 mm, a width of 183 μm and a kerf between adjacent elements of 10 μm. This transducer is able to scan a sector of 60° using a speed of sound of 1540 m/s. Through transmission, the transducer generates a Hanning-windowed 2-cycle sinusoid which bursts at 4 MHz. A relative bandwidth of 60% is used in transmission. The returned signals for each channel are sampled at 100 MHz. The RF data are exported to Matlab for beamforming. Dynamic receive focusing is employed in all algorithms. The Hilbert transformation is applied to the raw received RF data to detect the envelope and the logarithmic compression. To prevent a coarse display due to a large image line spacing, a low-pass interpolation was performed to the raw received RF data laterally by a factor of 5. The scan conversion is the last step to form the image. All images are represented over a 60-dB dynamic range.

2.4.1. Image Quality Metrics

Good image quality is the main challenge in the reconstruction of an ultrasound system. Therefore, it is assumed that good image quality options are the most important factors in the performance evaluation of an ultrasound imager. A precise diagnosis relies heavily on the obtained image quality, which can be characterized in terms of spatial resolution, contrast resolution, and signal to noise ratio.

2.4.1.1. Point Spread Function (PSF)

Metrics used to evaluate the quality of the Point Spread Function (PSF) are adopted from different fields such as radar and communications. The most commonly used metric to measure the resolution is the Full Width at Half Maximum (FWHM), while the side lobe performance is usually measured by the Peak Side lobe Level (PSL). In this evaluation, two new performance evaluation metrics for ultrasound imaging are utilized: Full Width at Half Dynamic Range (FWHDR) and Doubled Main-lobe Level (DML). Choosing a metric based on image dynamic range is well justified. Indeed, when the same image is plotted with a larger dynamic range, the image resolution drops. As a consequence, it becomes harder to distinguish between the closely placed scatterers. Therefore, the FWHDR is effectively the same as the FWHM for log-compressed data. The DML corresponds to the level where the projection of point spread function's (PSF) area at half dynamic range is doubled, which is equivalent to $\sqrt{2} \times$ FWHDR. The DML is based on a search algorithm that finds the point, in terms of dB, where the width of the main lobe plus side lobes $\geqslant \sqrt{2} \times$ FWHDR. The DML search algorithm starts from the tip of the main-lobe and moves vertically down [19] . The PSL does neither quantify the main lobe widening nor estimate the proximity of side lobes. However, the DML reflects the main lobe widening and side lobes behaviour. Nevertheless, it does not pick up the near side lobes, which do not reduce the image quality as much as far side lobes, since they are visualized as a part of the image.

2.4.1.2. Cyst Phantom

Ten cysts with radiuses varying from 8.5 mm to 2 mm are located at five penetration depths to evaluate the beamformers under the cyst targets. The cyst phantom consists of a collection of point targets, five cyst regions, and five highly scattering regions. This can be used for characterizing the contrast-lesion detection capabilities of an imaging system. The scatterers in the phantom are generated by finding their random position within a volume of 100 x 100 x 10 mm, and then ascribing a Gaussian distributed amplitude to each scatterer. If the scatterer resides within a cyst region, the amplitude is set to zero. Within the highly scattering region, the amplitude is multiplied by 20. The phantoms typically consist of 50,000 scatterers simulated over 128 scan lines. Figure 5 shows the cyst phantom morphology and the position of each cyst.
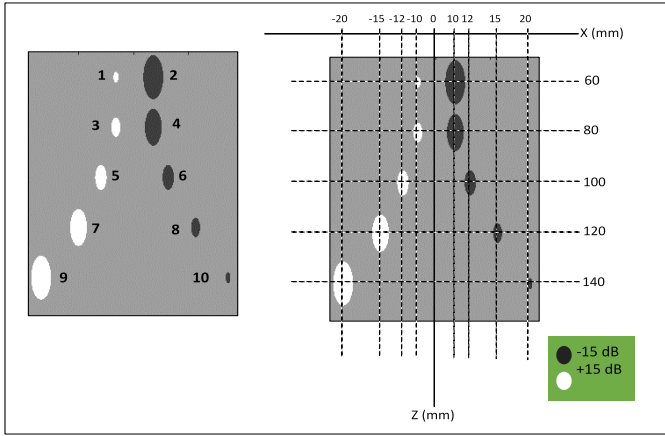
Figure 5: Cyst Phantom structures and dimensions

Metrics used to evaluate the quality of the Cyst Phantoms (CP) are:

- Signal to Noise Ratio (SNR)

The system sensitivity and echo detection capability depend on the determination of a system signal-to-noise ratio (SNR). The SNR can be measured using a tissue-mimicking phantom, in which case the SNR as a function of depth, z, can be determined as the ratio of the signal power, S, to the noise power, N [20]:

$$\text{SNR}(z) = \frac{S(z)}{N(z)} = \frac{E\{x(z)\}^2}{E\{x(z) - E\{x(z)\}^2\}} \qquad (4)$$

where E is the expectation operator, and x is the signal RF value.

In this article, the signal power will be estimated from M measurements as:

$$S(z) = \left(\frac{1}{M}\sum_{m=1}^{M} x_m(z)\right)^2 \qquad (5)$$

and the noise power will be estimated as:

$$N(z) = \frac{1}{M}\sum_{m=1}^{M} \left(x.\left|m(z) - \frac{1}{M}\sum_{m=1}^{M} x_m(z)\right|\right)^2 \qquad (6)$$

The SNR expressed in dB is calculated as follows:

$$\text{SNR}_{dB}(z) = 10 * \log_{10}(\text{SNR}(z)) \qquad (7)$$

penetration depths to evaluate the beamformers under the cyst targets.

- Contrast to Noise Ratio (CNR)

To quantify the contrast in the final image, the Contrast-to-Noise (CNR) values can be used, where the contrast is defined as the logarithmic difference between the mean values of the image kernels fully within and outside the cysts.

The CNR is expressed as [21]:

$$\text{CNR} = \frac{(\mu_B - \mu_T)^2}{\sigma_B^2 + \sigma_T^2} \qquad (8)$$

where μ denotes the mean and σ denotes the standard deviation of the log-compressed B-mode image. The subscripts T and B stand for the target and the background, respectively.

3. Results

3.1. PSF results

Figure 6 shows the Point Spread Function simulation results of DTA, DTA-Hann, Hanning and rectangular apodization algorithms. Results are generated for 6 different steering angles which are 0°, 5°, 10°, 15°, 20° and 25° for points located at the same depth which is 100 mm. Only the right side of the steering is considered, because the results obtained for negative steering angles are symmetric to positive ones.

It can be noticed from the figure that the DTA algorithm and DTA-Hann exhibit better main lobe characteristics compared to other algorithms. However, the side lobe levels are relatively higher than those obtained with the Hanning window.
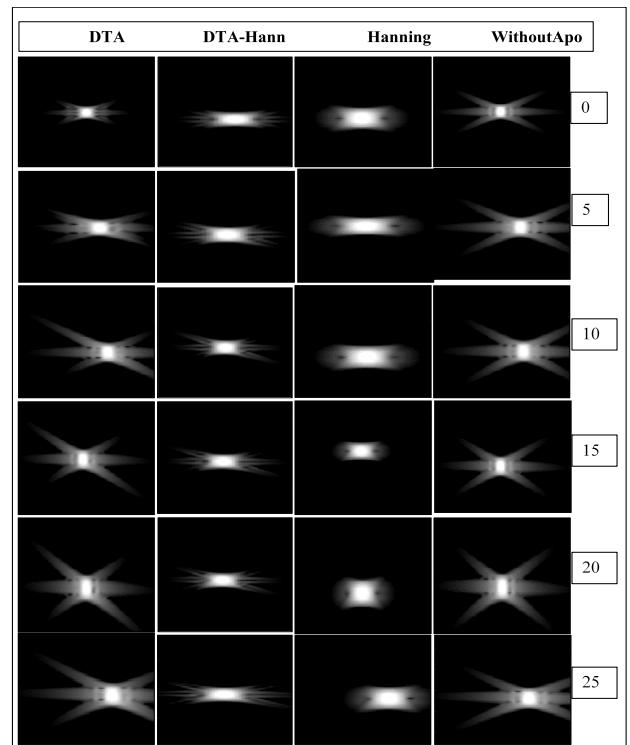
Figure 6: Point Spread Function simulation results for 0,5,10,15,20 and 25 steering angles, using three apodization algorithms; DTA, DTA-Hann, Hanning, and without apodization

Table 1 presents the Full Width Half Maximum (FWHM) for axial resolution evaluation of the PSFs. The DTA technique provides excellent results at 0°, 15° and 25° steering angles with 82 µm, 80.4 µm and 82.2 µm respectively. The rectangular apodization also offers some of the best results and shows an optimal main lobe width.

In order to quantify uncertainties between the two most accurate techniques, which are DTA and rectangular window, the Relative Difference (RD) method was calculated in the last column of the table, with the equation:

$$RD(\%)=100\times(B- A)/A \qquad (9)$$

where A and B are two different scenarios. The relative difference shows a very negligible difference between them regarding FWHM for axial resolution. Indeed, the percentage of the difference is within [-3, 3].

Table 1: FWHM Axial Resolution (µm)

Algorithm Angles	A = DTA	DTA-Hann	Hanning	B = Without_apo	RD(%)
0°	**82**	**82**	82.9	83.3	1.58
5°	**81.6**	82.2	82.8	**79.7**	-2.32
10°	82.4	83.2	**81.6**	**80.5**	-2.3
15°	**80.4**	**82.3**	83.2	82.7	2.86
20°	**81.5**	83.8	82.1	**80.5**	-1.22
25°	**82.2**	82.5	83.3	**81.9**	-0.36

Table 2 presents the Peak Side Lobe (PSL) for axial resolution evaluation of the four algorithms. The Hanning window shows the best values for all simulations. It exhibits the smallest levels of PSL. The DTA-Hann window also gives optimal PSL values, a little bit higher than Hanning results. The DTA algorithm offers PSL values very close to those provided by rectangular apodization.

Contrary to the FWHM, the relative difference obtained between the two most accurate techniques, which are DTA-Hann and Hanning window in this case, shows an important difference between them regarding PSL for axial resolution. The difference increases for large opening angle. Therefore, the Hanning window is more performant in the suppression of side lobe levels.

Table 2: PSL Axial Resolution (dB)

Algorithm Angles	A = DTA	DTA-Hann	Hanning	B = Without_apo	RD(%)
0°	-320.36	**-397.62**	**-436.18**	-289.75	9.69
5°	-309.51	-309.12	**-383.61**	**-312.27**	24.09
10°	-326.20	**-363.04**	**-678.30**	-344.25	86.84
15°	-306.23	**-345.35**	**-675.75**	-319.92	95.65
20°	-274.49	**-315.76**	**-672.48**	-288.89	113
25°	-217.74	**-265.18**	**-323.27**	-229.36	21.62

Table 3 details the PSF performance for axial resolution. It shows the FWHDR image quality option. The best parameters for each apodization algorithm are highlighted in the table below. The best performances are obtained for the DTA and rectangular windows. Obviously, the DTA apodization gives a wider FWHDR than the rectangular window, for 0°, 15° and 25° angles, indicating a better image resolution.

The last column of Table 3 shows the relative difference between DTA and rectangular window. The difference is very negligible regarding FWHDR for axial resolution. In fact, the percentage of the difference is within [-3, 1].

Table 3: FWHDR Axial Resolution (µm)

Algorithm Angles	A = DTA	DTA-Hann	Hanning	B = Without_apo	RD(%)
0°	**171.0**	**171.0**	171.5	172.4	0.8
5°	168.2	172.0	171.8	**163.1**	-2.9
10°	165.1	172.2	170.9	**163.4**	-1.0
15°	**169.6**	171.4	172.2	**170.9**	0.76
20°	**171.1**	173.0	171.6	**168.2**	-1.69
25°	166.9	172.4	172.7	**168.5**	0.95

Table 4 details the PSF performance for axial resolution. It shows the DML image quality option. The best performances are obtained with the DTA-Hann and Hanning windows. Therefore, the DTA-Hann technique provides better main lobe characteristics than the Hanning, for 10° and 20° angles, indicating a better image resolution especially for more deviated directions.

The relative difference obtained between the two most accurate techniques, which are DTA-Hann and Hanning window, shows a little difference between them regarding DML for axial resolution. Therefore, both techniques are suitable in this case.

Table 4: DML Axial Resolution (µm)

Algorithm Angles	A = DTA	DTA-Hann	Hanning	B = Without_apo	RD(%)
0°	-68.6	-68.6	**-69.4**	**-70.3**	1.16
5°	-68.1	**-68.9**	**-69.7**	-65.9	1.16
10°	-66.7	**-69.1**	**-68.5**	-66.1	0.86
15°	-68.1	-68.5	**-69.7**	**-68.9**	1.75
20°	**-69.1**	**-69.3**	-68.5	-67.5	1.37
25°	-65.5	**-68.5**	**-69.4**	-66.9	1.31

Table 5 presents the Full Width Half Maximum (FWHM) for lateral resolution evaluation of the PSFs generated by the four apodization techniques. Obviously, the DTA and DTA-Hann algorithms provide in this case the best values of FWHM. These results are optimal for an ultrasound imaging application. The rectangular apodization shows very close FWHM values. However, the Hanning gives very large FWHM lateral resolution values, especially for a very large opening angle.

Table 5: FWHM Lateral Resolution (µm)

Algorithm Angles	A = DTA	DTA-Hann	Hanning	B = Without_apo	RD(%)
0°	**48**	**44.5**	68.2	49.6	3.33
5°	**44**	52.8	68.8	**33.5**	-23.86
10°	**36.6**	53.6	69.3	**34.2**	-6.55
15°	50.5	**47.2**	71.2	**50.1**	-0.79
20°	**51**	52.1	73.6	**43.0**	-15.68
25°	**43.6**	56	76.7	**43.3**	-0.68

The relative difference between the most accurate techniques, which are DTA and rectangular window, is more or less significant regarding FWHM for lateral resolution. The rectangular window provides better main lobe especially for 5° and 20°.

Table 6 presents the Peak Side Lobe (PSL) for lateral resolution evaluation of the four algorithms. The Hanning and DTA-Hann window show the best values for almost all simulations, with a little difference between them. They exhibit the smallest levels of PSL lateral resolution. The DTA algorithm provides acceptable values for less deviated angles (0°, 5° and 10°), but values at higher steering angles are very similar to those of rectangular apodization.

An important difference appears between the two best techniques regarding PSL for lateral resolution, which are DTA-Hann and Hanning window. The difference increases for large opening angle. Consequently, the Hanning window is more performant in the suppression of side lobe levels. However, only for 10° opening angle, the DTA-Hann gives better results than Hanning.

Table 6: PSL Lateral Resolution (dB)

Algorithm Angles	A = DTA	DTA-Hann	Hanning	B = Without_apo	RD(%)
0°	**-381.22**	-335.85	**-377.52**	0	12.40
5°	-137.22	**-375.38**	**-454.77**	-142.94	21.14
10°	-145.81	**-346.99**	**-216.21**	-149.02	37.68
15°	-121.32	**-203.59**	**-558.17**	-124.07	174.16
20°	-120.20	**-317.82**	**-579.26**	-124.83	114.07
25°	-121.47	**-198.54**	**-586.63**	-125.03	195.47

Table 7 presents the FWHDR performance for lateral resolution of the PSF. Obviously, the DTA technique exhibits a very similar profile to the rectangular windowing technique. However, the DTA apodization gives a wider FWHDR for 15° angle.

The relative difference between the two best techniques, which are DTA and rectangular window, reveals an important difference between them for less deviated angles. The FWHDR lateral resolution provided by DTA is better for highly deviated angles.

Table 7: FWHDR Lateral Resolution (μm)

Algorithm Angles	A = DTA	DTA-Hann	Hanning	B = Without_apo	RD(%)
0°	**102.4**	121.1	149.1	**93.2**	**-8.98**
5°	**96.6**	117.0	152.1	**81.5**	**-15.63**
10°	**86.4**	115.3	153.7	**82.1**	**-4.97**
15°	**95.2**	120.6	156.8	**95.4**	**0.21**
20°	**99.6**	123.1	162.2	**94.6**	**-5.02**
25°	**100.8**	125.6	169.1	**100.2**	**-0.59**

Table 8: DML Lateral Resolution (μm)

Algorithm Angles	A = DTA	DTA-Hann	Hanning	B = Without_apo	RD(%)
0°	**-63.8**	-60.9	-55.8	**-71.0**	11.28
5°	**-63.9**	-62.4	**-58.6**	-56.9	-10.95
10°	-57.7	**-64.7**	**-58.8**	-56.5	-2.07
15°	**-70.3**	-61.8	-56.5	**-71.0**	0.99
20°	**-68.2**	-61.6	-58.6	**-62.8**	-7.91
25°	-58.2	-56.2	**-58.6**	**-59.0**	1.37

Table 8 describes the DML performance for lateral resolution of the PSF. Also here, the DTA and rectangular windowing

techniques provide the best performances. In addition, the DTA apodization shows the best profile for 5° and 20° angles.

A little difference is obtained between the two most accurate techniques regarding DML for lateral resolution, which are DTA and rectangular window. As a consequence, the two techniques are similar in this case.

3.2. Cyst Phantom Results

The reconstructed images are provided in Figure 7. Obviously, the cyst targets are not easily detected in the image generated by rectangular apodization. Moreover, the reconstructed images suffer from the noise. Indeed, the DTA-Hann technique results in a higher quality image and more detectable cyst targets.
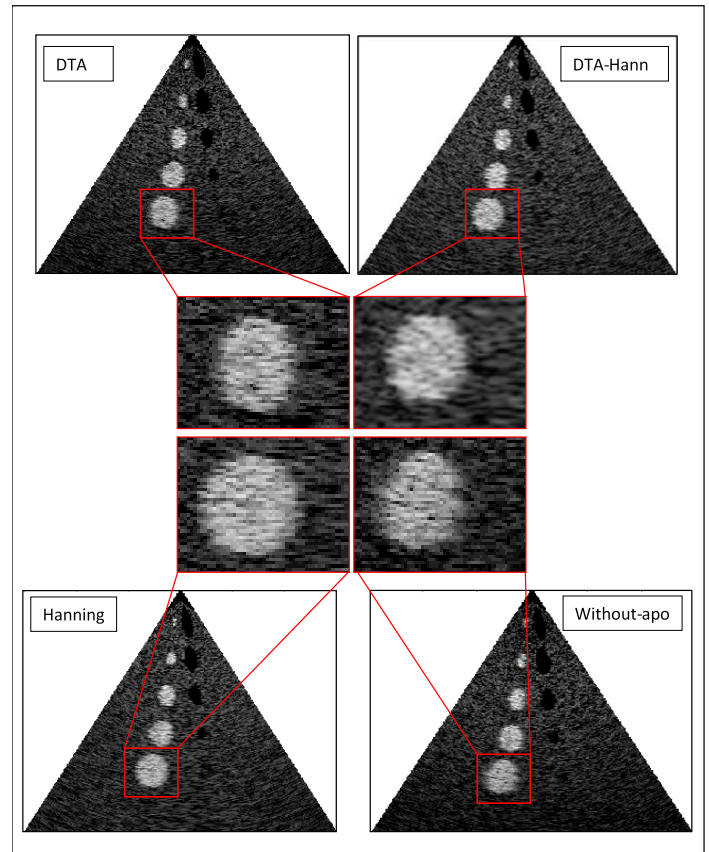
Figure 7: Cyst Phantom simulations results for three apodization algorithms; DTA, DTA-Hann, Hanning and without apodization.

Table 9 presents the Signal to Noise Ratio (SNR) and Contrast to Noise Ratio (CNR) evaluation of the cyst phantom technique for three apodization algorithms; DTA, DTA-Hann, and Hanning, as well as the algorithm without apodization. The best SNR is obtained with the DTA method, reaching 14.34 dB. The rectangular apodization also provides acceptable SNR values with 14.32 dB. However, the DTA-Hann and Hanning methods show lower SNR values with 13.98 dB and 13.88 dB, respectively.

The CNR results demonstrate that the DTA and DTA-Hann methods offer the best values, which are 42.42 dB and 42.66 dB, respectively. The Rectangular apodization and Hanning apodization show less accurate CNR results, with 41.35 dB and 42.25 dB, respectively.

A little difference is obtained between the two most accurate techniques regarding CNR and SNR (RD = -056% for CNR and RD = 2.57% for SNR), which are DTA and DTA-Hann. Consequently, the two techniques are similar in this case.

Table 9: Signal to Noise Ratio (SNR) and Contrast to Noise Ratio (CNR) evaluation of the cyst phantom

Apodization Type	SNR (dB)	CNR (dB)
B = DTA	**14.34**	42.42
A = DTA-Hann	13.98	**42.66**
Hanning	13.88	42.25
Without	14.32	41.35
R D (%)	2.57%	-0.56%

4. Discussion

The dual apodization technique DTA-Hann offers excellent image quality, which is revealed by PSF and Cyst phantom simulation results. For tissue structures located at less deviated angles, the DTA gives results very close to those obtained with linear windows. For tissue structures located at more deviated angles, the DTA-Hann provides the best image quality in terms of SNR and CNR. This is due to the fact that DTA-Hann eliminates the side lobes produced by the field generated in the same directivity. The DTA-Hann algorithm gives an optimal FWHM compared to Hanning and rectangular windows. Indeed, this is justified by the high selectivity of the undamaged transmitted and received echoes. However, the PSL levels are relatively high especially for very deviated angles, and this is generated because of the non-linearity of the dynamic algorithm.

When DTA-Hann and Hanning algorithms are compared qualitatively and quantitatively as shown in Figure 7 and Table 10, the FWHM shows a relative difference of 29.29% between them (DTA-Hann: scenario A with 53.6 µm, Hanning: scenario B with 69.3 µm). In fact, the scenario A is much better than the scenario B. The scenario A provides 130 dB and 5.9 dB more than the scenario B in terms of PSL and DML, respectively. When the FWHDR is used to measure the lateral width, there is a difference of 33.3% between these two cases (A = 115.3 µm, B = 153.7 µm). Here also, the scenario A is much better than the scenario B.

Table 10: Performance metrics comparison for lateral resolution at 10° angle between DTA-Hann and Hanning windows

	FWHM	FWHDR	DML	PSL
A = DTA-Hann	53.6	115.3	-64.7	**-346.99**
B = Hanning	69.3	153.7	**-58.8**	-216.21
RD(%)	29.29 %	33.30%	-9.11%	37.68%

The DTA-Hann technique is more selective than linear functions, providing the best main lobe characterization. Side lobes reductions are more or less significant using DTA-Hann since they depend on the deviation angle. Linear methods are more efficient for side lobes elimination, but at the cost of less characterization of the main lobe.

Finally, the DTA-Hann algorithm is more suitable to ultrasound imaging than linear apodization windows. It is the most performant and selective. It also offers the most accurate image options for cyst phantom simulation.

The DTA is used in transmit and receive apodization. Dynamic receive apodization methods are possible and can be implemented easily in the digital beamforming ultrasound software. However, dynamic transmit apodization methods are not very known for the time being, and the exploration of the DTA-Hann in transmission is among prospects of future work for more evaluation and study.

5. Conclusion

In this paper, a new dynamic apodization technique (DTA) that enables windowing for each imaging direction was proposed. Then, the efficiency of a non-linear windowing technique that combines the Hanning window with the DTA algorithm was evaluated.

In the point target simulation, the main lobe widths of the dual-apodization algorithm were very similar to those of the rectangular window without removing side lobes. In the cyst target simulation, CNR values of the dual apodization were significantly improved. The Dynamic Triangular Apodization is our new suggested method. It is based on following the shape of the transmitted and received ultrasound field from a focal point. Thanks to this method, the apodization weight becomes more sensitive to the undamaged transmitted and received echoes. This apodization method can be used in transmission and reception. Actually, the implementation of the DTA-Hann algorithm in digital beamforming ultrasound software is among prospects of future work.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] K. Sumaiya, S.S. Kawathekar, "Drawbacks of Poor-Quality Ultrasound Images and its Enhancement," International Journal of Computer Applications, **175**, 47–55, 2020, doi:10.5120/ijca2020920880.

[2] A. Ilovitsh, T. Ilovitsh, K.W. Ferrara, "Multiplexed ultrasound beam summation for side lobe reduction," Scientific Reports, **9**(1), 13961, 2019, doi:10.1038/s41598-019-50317-7.

[3] A. Agarwal, J. Reeg, A.S. Podkowa, M.L. Oelze, "Improving Spatial Resolution Using Incoherent Subtraction of Receive Beams Having Different Apodizations," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, **66**(1), 5–17, 2019, doi:10.1109/TUFFC.2018.2876285.

[4] A.N. Madhavanunni, M.R. Panicker, "Directional Beam Focusing Based Dual Apodization Approach for Improved Vector Flow Imaging," in 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 300–303, 2020, doi:10.1109/ISBI45749.2020.9098494.

[5] A. Vayyeti, A.K. Thittai, "A Weighted Non-Linear Beamformer for Synthetic Aperture Ultrasound Imaging," in 2020 IEEE International Ultrasonics Symposium (IUS), 1–3, 2020, doi:10.1109/IUS46767.2020.9251413.

[6] J.T. Yen, E. Powis, "Boundary Array Transducer and Beamforming for Low-Cost Real-time 3D Imaging," in 2020 IEEE International Ultrasonics Symposium (IUS), 1–4, 2020, doi:10.1109/IUS46767.2020.9251670.

[7] Side lobe free medical ultrasonic imaging with application to assessing side lobe suppression filter | SpringerLink, Sep. 2021.

[8] W. Guo, Y. Wang, J. Yu, "A Sibelobe Suppressing Beamformer for Coherent Plane Wave Compounding," Applied Sciences, **6**(11), 359, 2016, doi:10.3390/app6110359.

[9] S. Lu, R. Li, Y. Zhao, X. Yu, D. Wang, M. Wan, "Dual apodization with cross-correlation combined with robust Capon beamformer applied to

ultrasound passive cavitation mapping," Medical Physics, **47**(5), 2182–2196, 2020, doi:10.1002/mp.14093.

[10] Sidelobe reduction for plane wave compounding with a limited frame number | BioMedical Engineering OnLine | Full Text, Sep. 2021.

[11] A. Ibrahim, P.A. Hager, A. Bartolini, F. Angiolini, M. Arditi, J.-P. Thiran, L. Benini, G. De Micheli, "Efficient Sample Delay Calculation for 2-D and 3-D Ultrasound Imaging," IEEE Transactions on Biomedical Circuits and Systems, **11**(4), 815–831, 2017, doi:10.1109/TBCAS.2017.2673547.

[12] W. Hechkel, B. Maaref, N. Hassen, "Evaluation of transmit/receive apodization scheme for hardware implementation used in B-mode ultrasound imaging," in 2020 17th International Multi-Conference on Systems, Signals Devices (SSD), 808–813, 2020, doi:10.1109/SSD49366.2020.9364236.

[13] W. Hechkel, B. Maaref, N. Hassen, "Evaluation of a novel ultrasound imaging methods for high frame rate echocardiography," in 2020 17th International Multi-Conference on Systems, Signals Devices (SSD), 185–190, 2020, doi:10.1109/SSD49366.2020.9364257.

[14] W. Hechkel, B. Maaref, N. Hassen, "New Sector Scan Geometry for High Frame Rate 2D-Echocardiography using Phased Arrays," International Journal of Advanced Computer Science and Applications (IJACSA), **11**(12), 2020, doi:10.14569/IJACSA.2020.0111266.

[15] G. Malamal, M.R. Panicker, "Towards A Pixel-Level Reconfigurable Digital Beamforming Core for Ultrasound Imaging," IEEE Transactions on Biomedical Circuits and Systems, **14**(3), 570–582, 2020, doi:10.1109/TBCAS.2020.2983759.

[16] M. Agarwal, A. Tomar, N. Kumar, "An IEEE single-precision arithmetic based beamformer architecture for phased array ultrasound imaging system," Engineering Science and Technology, an International Journal, **24**(5), 1080–1089, 2021, doi:10.1016/j.jestch.2021.03.005.

[17] J.T. Yen, M.M. Nguyen, Y. Lou, J.S. Shin, Y. Chen, H.L. Tarnoff, "Gated Transmit and Fresnel-Based Receive Beamforming With a Phased Array for Low-Cost Ultrasound Imaging," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, **68**(6), 2183–2192, 2021, doi:10.1109/TUFFC.2021.3062850.

[18] R.J.G. van Sloun, R. Cohen, Y.C. Eldar, "Deep Learning in Ultrasound Imaging," Proceedings of the IEEE, **108**(1), 11–29, 2020, doi:10.1109/JPROC.2019.2932116.

[19] S. Harput, J. McLaughlan, D.M.J. Cowell, S. Freear, "New performance metrics for ultrasound pulse compression systems," in 2014 IEEE International Ultrasonics Symposium, 440–443, 2014, doi:10.1109/ULTSYM.2014.0109.

[20] V. Papic, Z. Djurovic, G. Kvascev, P. Tadic, "On signal-to-noise ratio estimation," in Melecon 2010 - 2010 15th IEEE Mediterranean Electrotechnical Conference, 160–165, 2010, doi:10.1109/MELCON.2010.5476314.

[21] M. Welvaert, Y. Rosseel, "On the Definition of Signal-To-Noise Ratio and Contrast-To-Noise Ratio for fMRI Data," PLOS ONE, **8**(11), e77089, 2013, doi:10.1371/journal.pone.0077089.

Numeric Simulation on the Waves from Artificial Anti-gravity upon General Theory of Relativity

Yoshio Matsuki[1,*], Petro Ivanovich Bidyuk[2]

[1]*National University of Ukraine "Kyiv-Mohyla Academy", Kyiv 04070, Ukraine*

[2]*Educational and Scientific Complex "Institute for Applied System Analysis" of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv 03056, Ukraine*

A R T I C L E   I N F O

A B S T R A C T

This paper reports the algorithm, the input data and the result of the numeric simulation on the flows of the waves emitted from a rotating object that forms the artificial anti-gravity. First, an object with a heavy mass is placed in the 4-dimensional time and space, which is described by a fundamental tensor. Then the first-order derivative of the tensor describes the gravity, and the second-order derivative describes the waves. If the gravity created by the heavy mass is strong enough, time and space become dependent on each other. The input data for the simulation are discrete numbers that surrogate the infinity of the 4-dimensional time-space. The object is assumed to rotate and the tensor equations are solved. Then the coefficients are calculated, which present physical properties of the waves. The result of the simulation shows that the rotating object emits the waves with positive and negative energy intensities, and they have the spin angular momentum that changes its spinning direction upon the selection of the rotation-speed of the object.

1. Introduction

This research is an extended version of the work presented at the 2020 IEEE 2nd International Conference on System Analysis_Intelligent Computing (SAIC) held in Kyiv, Ukraine, in October 2020 [1] as well as the consequently published paper [2] from the Advances in Science, Technology and Engineering Systems Journal in May 2021. These papers reported that a rotating object would produce anti-gravity, if it held a heavy mass that would distort time and space. The last paper [2] informed of the waves emitted from the rotating object; however, the report of the analysis was forwarded to this new paper because the numeric simulation required a separate research.

An explicit image of the anti-gravity of the rotating object is a flying craft having a disc-shaped body for interstellar travel. Such an object should hold a compact nuclear fusion reactor inside, because the fusion reaction forms the heavier nuclei by fusing the lighter ones such as hydrogen, deuterium and tritium, and this process should distort time and space. However, this is the theory of a black hole [3], and the simulation of the time-space distortion in the nuclear fusion reactor is still the topic of the next research after this paper.

*Corresponding Author: Yoshio Matsuki, matsuki@wdc.org.ua

The waves emitted from the source of the anti-gravity are byproducts, but they can indicate the presence of the source if they have any physical properties such as the energy intensity and the spin angular momentum. The goal of this research is to calculate the energy intensity and the angular momentum of the waves.

In order to describe this phenomenon, the 4-dimensional curvilinear coordinates of the curved space are selected because the heavy object should warp the rectilinear coordinates of the flat space. To start the simulation, the curvature tensor, $R_{\mu\nu}$, is given to describe time and space where the object with a heavy mass is placed [3].

$$R_{\mu\nu} = 0 \tag{1}$$

$$R_{\mu\nu} = \Gamma^{\alpha}_{\mu\alpha,\nu} - \Gamma^{\alpha}_{\mu\nu,\alpha} - \Gamma^{\alpha}_{\mu\nu}\Gamma^{\beta}_{\alpha\beta} + \Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha} = 0 \tag{2}$$

$$\Gamma_{\mu\nu\sigma} = \frac{1}{2}\left(g_{\mu\nu,\sigma} + g_{\mu\sigma,\nu} - g_{\nu\sigma,\mu}\right) \tag{3}$$

This curvature tensor includes the first-order differentials of a fundamental tensor ($g_{\mu\nu}$), $g_{\mu\nu,\sigma} = \partial g_{\mu\nu}/\partial x^{\sigma}$, where x^{σ} is the σ-th vector (variable) in the given coordinates. Before describing the phenomenon in the curvilinear coordinates, the flat space provides

a significant introduction to the theory of the waves. Therefore, at first a series of the notations derives the tensor of the waves in an approximate flat space in this introductory section, and then the tensor of the waves is derived in the curved space later in the next section 2.1.

If the gravitational field is weak, the curvature is also weak and $g_{\mu\nu}$ is approximately constant; therefore,

$$-\Gamma_{\mu\nu}^{\alpha}\Gamma_{\alpha\beta}^{\beta}+\Gamma_{\mu\beta}^{\alpha}\Gamma_{\nu\alpha}^{\beta} = 0, \tag{4}$$

and

$$R_{\mu\nu}=\Gamma_{\mu\alpha,\nu}^{\alpha} - \Gamma_{\mu\nu,\alpha}^{\alpha} = 0. \tag{5}$$

On the other hand,

$$R_{\mu\nu\rho\sigma} = \frac{1}{2}\left(g_{\mu\sigma,\nu\rho} - g_{\nu\sigma,\mu\rho} - g_{\mu\rho,\nu\sigma}+g_{\nu\rho,\mu\sigma}\right)$$

$$+\Gamma_{\beta\mu\sigma}\Gamma_{\nu\rho}^{\beta} - \Gamma_{\beta\mu\rho}\Gamma_{\nu\sigma}^{\beta}. \tag{6}$$

By interchanging ρ and μ, and neglecting $\Gamma_{\beta\mu\sigma}\Gamma^{\beta}{}_{\nu\rho}$ - $\Gamma_{\beta\mu\rho}\Gamma^{\beta}{}_{\nu\sigma}$, we get

$$R_{\mu\nu}=g^{\rho\sigma}R_{\mu\nu\rho\sigma}$$

$$=g^{\rho\sigma}\left(g_{\rho\sigma,\mu\nu} - g_{\nu\sigma,\mu\rho} - g_{\mu\rho,\nu\sigma}+g_{\mu\nu,\rho\sigma}\right) = 0. \tag{7}$$

Then

$$g^{\mu\nu}\left(g_{\mu\nu,\rho\sigma} - g_{\mu\rho,\nu\sigma} - g_{\mu\sigma,\nu\rho}+g_{\rho\sigma,\mu\nu}\right) = 0. \tag{8}$$

In general, the moving particle in a scalar field of the potential energy *V* follows d'Alembert equation,

$$\nabla V=g^{\mu\nu}\left(V_{,\mu\nu} - \Gamma_{\mu\nu}^{\alpha}V_{,\alpha}\right) = 0. \tag{9}$$

In order to describe a particle moving in the 4-dimensional time and space, *V* is replaced by vectors x^{λ} of the 4-dimensional coordinates in which $x^{\lambda}{}_{,\alpha} = g^{\lambda}{}_{\alpha}$, then d'Alembert equation becomes

$$g^{\mu\nu}g_{\alpha,\nu}^{\lambda} - g^{\mu\nu}g_{\alpha}^{\lambda}\Gamma_{\mu\nu}^{\alpha} = 0. \tag{10}$$

Then

$$g^{\mu\nu}\Gamma_{\mu\nu}^{\lambda} = 0. \tag{11}$$

Meanwhile,

$$\Gamma_{\mu\nu}^{\lambda}=g^{\lambda\rho}\Gamma_{\rho\mu\nu} = \frac{1}{2}g^{\lambda\rho}\left(g_{\rho\mu,\nu}+g_{\rho\nu,\mu} - g_{\mu\nu,\rho}\right). \tag{12}$$

Therefore,

$$g^{\mu\nu}\Gamma_{\mu\nu}^{\lambda} = \frac{1}{2}g^{\lambda\rho}g^{\mu\nu}\left(g_{\rho\mu,\nu}+g_{\rho\nu,\mu} - g_{\mu\nu,\rho}\right)$$

$$=g^{\lambda\rho}g^{\mu\nu}\left(g_{\rho\mu,\nu} - \frac{1}{2}g_{\mu\nu,\rho}\right) = 0, \tag{13}$$

where

$$g_{\rho\nu,\mu}=y_{n,\rho\mu}y_{,\nu}^{n}+y_{n,\nu\mu}y_{,\rho}^{n}=y_{n,\rho\nu}y_{,\mu}^{n}+y_{n,\mu\nu}y_{,\rho}^{n}=g_{\rho\mu,\nu}, \tag{14}$$

where μ and ν are in symmetrical relation in the equation, therefore they are exchangeable; and,

$$y_{,\mu}^{n} = \frac{\partial y^{n}(x^{\mu})}{\partial x^{\mu}}, \tag{15}$$

where $\mu = 0$, 1, 2, 3, while x^{μ} are located in N-dimensional physical space of y^{n}, where $n = 1, 2, \ldots, N$.

Therefore,

$$g^{\mu\nu}\left(g_{\rho\mu,\nu} - \frac{1}{2}g_{\mu\nu,\rho}\right) = 0. \tag{16}$$

Then, in order to describe the waves moving in the gravitational field, it is differentiated by x^{σ}.

$$\frac{d}{dx^{\sigma}}g^{\mu\nu}\left(g_{\rho\mu,\nu} - \frac{1}{2}g_{\mu\nu,\rho}\right)$$

$$=g_{,\sigma}^{\mu\nu}\left(g_{\rho\mu,\nu} - \frac{1}{2}g_{\mu\nu,\rho}\right)+g^{\mu\nu}\left(g_{\mu\rho,\nu\sigma} - \frac{1}{2}g_{\mu\nu,\rho\sigma}\right)$$

$$=g^{\mu\nu}\left(g_{\mu\rho,\nu\sigma} - \frac{1}{2}g_{\mu\nu,\rho\sigma}\right) = 0 \tag{17}$$

In the equation (17),

$$g_{,\sigma}^{\mu\nu}\left(g_{\rho\mu,\nu} - \frac{1}{2}g_{\mu\nu,\rho}\right) = 0 \tag{18}$$

because $g_{\rho\mu}$ is approximately constant in the weak gravitational field. Therefore,

$$g_{\rho\mu,\nu} = 0 \tag{19}$$

By interchanging ρ and σ of (17),

$$g^{\mu\nu}\left(g_{\mu\sigma,\nu\rho} - \frac{1}{2}g_{\mu\nu,\sigma\rho}\right)=g^{\mu\nu}\left(g_{\mu\sigma,\nu\rho} - \frac{1}{2}g_{\mu\nu,\rho\sigma}\right) = 0. \tag{20}$$

By adding (8), (17) and (20), we get

$$g^{\mu\nu}g_{\rho\sigma,\mu\nu} = 0. \tag{21}$$

It satisfies d'Alembert equation, therefore it describes the waves that move in an empty space where there is only the weak gravitational field. (Note: This equation includes the second-order differentials, $g_{\mu\nu,\rho\sigma} = \partial^{2}g_{\mu\nu}/\partial x^{\rho}\partial x^{\sigma}$.)

The equation (21) describes the waves moving in the weak gravitational field that is approximately a flat 4-dimensional time and space; however, the general theory of relativity predicts the flow of the waves in curvilinear coordinates of the stronger gravitational field. In order to simulate the flow of the waves in the stronger gravitational field, this research includes the

following terms that have been neglected in the classic reference [3]:

$$-\Gamma_{\mu\nu}^{\alpha}\Gamma_{\alpha\beta}^{\beta}+\Gamma_{\mu\beta}^{\alpha}\Gamma_{\nu\alpha}^{\beta} \neq 0 \tag{22}$$

instead of the equation (4),

$$\Gamma_{\beta\mu\sigma}\Gamma_{\nu\rho}^{\beta} - \Gamma_{\beta\mu\rho}\Gamma_{\nu\sigma}^{\beta} \neq 0 \tag{23}$$

for the equation (6), and

$$g_{,\sigma}^{\mu\nu}\left(g_{\rho\mu,\nu} - \frac{1}{2}g_{\mu\nu,\rho}\right) \neq 0 \tag{24}$$

instead of the equation (18).

In this research, the energy intensity and the spin angular momentum of the flow of the waves are simulated as the calculated coefficients of the tensor equation in the spherical polar coordinate system, which is one of the curvilinear coordinate systems. The methodology and the result of the simulation are reported in the following sections.

2. Methodology

2.1. Algorithm

For the research of the strong gravitational field, $g_{\mu\nu}$, is not constant, but given by the following matrix, [3]:

$$g_{\mu\nu} = \begin{bmatrix} g_{00} & 0 & 0 & 0 \\ 0 & g_{11} & 0 & 0 \\ 0 & 0 & g_{22} & 0 \\ 0 & 0 & 0 & g_{33} \end{bmatrix} \tag{25}$$

$$g_{00} = 1 \tag{26}$$

$$g_{11} = -\frac{2m}{\mu(\rho-\tau)^{\frac{2}{3}}} \tag{27}$$

$$g_{22} = -\mu^2(\rho - \tau)^{\frac{4}{3}} \tag{28}$$

$$g_{33} = -\mu^2(\rho - \tau)^{\frac{4}{3}}\sin^2\theta \tag{29}$$

And the inverse of $g_{\mu\nu}$ is $g^{\mu\nu}$. For example,

$$g^{11} = -\frac{\mu(\rho-\tau)^{\frac{2}{3}}}{2m} \tag{30}$$

Then t is time, m is the mass of the object, r is the distance from the center of the object, θ is the angle between r and the rotation axis, z, and ϕ is the angle of the rotation around the rotation axis in the spherical polar coordinates shown in Figure 1.

τ is the distorted time and ρ is the distorted distance as defined by the following equations [3].

$$\tau = t + f(r) \tag{31}$$

$$\rho = t + g(r) \tag{32}$$
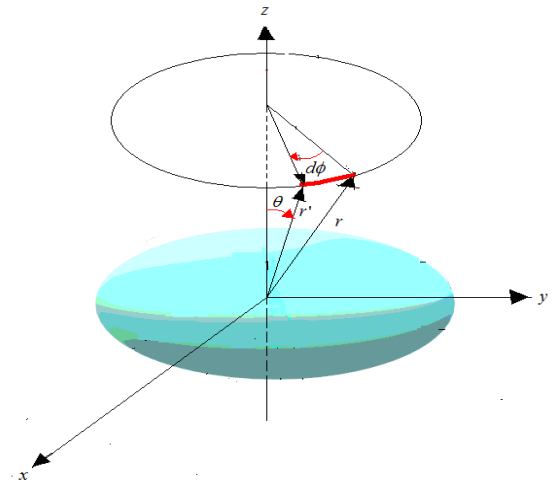
Figure 1: Spherical polar coordinates for the simulation

$f(r)$ and $g(r)$ are the functions later defined in the section 2.2. μ is given by the following equation with the mass, m, of the object ([3] p. 35):

$$\mu = \left(\frac{3}{2}\sqrt{2m}\right)^{\frac{2}{3}} \tag{33}$$

The equations below from (34) to (38) show how to derive the equation of the waves in the curved space. It follows the same logic from (7) to (21) in the introductory section above, which derives the equation of the waves in the approximate flat space. But, the equations below include the neglected terms in order to derive the equation in the curved space.

By adding the neglected terms of the equation (23) to the equation (7), we get:

$$R_{\mu\nu} = g^{\rho\sigma}R_{\mu\nu\rho\sigma} = g^{\rho\sigma}\left(g_{\rho\sigma,\mu\nu} - g_{\nu\sigma,\mu\rho} - g_{\mu\rho,\nu\sigma} + g_{\mu\nu,\rho\sigma}\right)$$
$$+ g^{\rho\sigma}\Gamma_{\beta\rho\sigma}\Gamma_{\mu\nu}^{\beta} - g^{\rho\sigma}\Gamma_{\beta\mu\rho}\Gamma_{\nu\sigma}^{\beta}$$
$$= g^{\rho\sigma}\left(g_{\rho\sigma,\mu\nu} - g_{\nu\sigma,\mu\rho} - g_{\mu\rho,\nu\sigma} + g_{\mu\nu,\rho\sigma}\right)$$
$$+ g^{\rho\sigma}g_{\beta\beta}\Gamma_{\beta\rho\sigma}\Gamma_{\beta\mu\nu} - g^{\rho\sigma}g_{\beta\beta}\Gamma_{\beta\mu\rho}\Gamma_{\beta\nu\sigma}$$
$$= g^{\rho\sigma}\left(g_{\rho\sigma,\mu\nu} - g_{\nu\sigma,\mu\rho} - g_{\mu\rho,\nu\sigma} + g_{\mu\nu,\rho\sigma}\right)$$
$$+ g^{\rho\sigma}g_{\beta\beta}\frac{1}{2}\left(g_{\beta\rho,\sigma} + g_{\beta\sigma,\rho} - g_{\rho\sigma,\beta}\right)\left(g_{\beta\mu,\nu} + g_{\beta\nu,\mu} - g_{\mu\nu,\beta}\right)$$
$$- g^{\rho\sigma}g_{\beta\beta}\frac{1}{2}\left(g_{\beta\mu,\rho} + g_{\beta\rho,\mu} - g_{\mu\rho,\beta}\right)\left(g_{\beta\nu,\sigma} + g_{\beta\sigma,\nu} - g_{\nu\sigma,\beta}\right)$$
$$= g^{\rho\sigma}\left(g_{\rho\sigma,\mu\nu} - g_{\nu\sigma,\mu\rho} - g_{\mu\rho,\nu\sigma} + g_{\mu\nu,\rho\sigma}\right)$$
$$+ \left(\frac{1}{2}g^{\rho\sigma}g_{\beta\beta}g_{\rho\sigma,\beta}g_{\nu\sigma,\beta}\right) = 0. \tag{34}$$

Interchanging ρ, μ, σ and ν, we get

$$g^{\mu\nu}\left(g_{\mu\nu,\rho\sigma} - g_{\mu\rho,\nu\sigma} - g_{\mu\sigma,\nu\rho} + g_{\rho\sigma,\mu\nu}\right)$$
$$+ \left(\frac{1}{2}g^{\rho\mu}g_{\beta\beta}g_{\rho\mu,\beta}g_{\nu\mu,\beta}\right) = 0. \tag{35}$$

By adding the neglected terms to the equation (17), we get

$$\frac{d}{dx^\sigma}\left\{g^{\mu\nu}\left(g_{\rho\mu,\nu}-\frac{1}{2}g_{\mu\nu,\rho}\right)+\frac{1}{2}g^{\rho\mu}g_{\beta\beta}g_{\rho\mu,\beta}g_{\mu\nu,\beta}\right\}$$

$$=g^{\mu\nu}_{,\sigma}\left(g_{\rho\mu,\nu}-\frac{1}{2}g_{\mu\nu,\rho}\right)+g^{\mu\nu}\left(g_{\mu\rho,\nu\sigma}-\frac{1}{2}g_{\mu\nu,\rho\sigma}\right)$$

$$+\left(\frac{1}{2}g^{\rho\mu}g_{\beta\beta}g_{\rho\mu,\beta}g_{\nu\mu,\beta\sigma}\right)+\left(\frac{1}{2}g^{\rho\mu}g_{\beta\beta}g_{\rho\mu,\beta\sigma}g_{\nu\mu,\beta}\right)$$

$$+\left(\frac{1}{2}g^{\rho\mu}_{,\sigma}g_{\beta\beta}g_{\rho\mu,\beta}g_{\nu\mu,\beta}\right)$$

$$+\left(\frac{1}{2}g^{\rho\mu}g_{\beta\beta,\sigma}g_{\rho\mu,\beta}g_{\nu\mu,\beta}\right)=0. \qquad (36)$$

Interchanging ρ and σ in (36), we get

$$g^{\mu\nu}_{,\rho}\left(g_{\sigma\mu,\nu}-\frac{1}{2}g_{\mu\nu,\sigma}\right)+g^{\mu\nu}\left(g_{\mu\sigma,\nu\rho}-\frac{1}{2}g_{\mu\nu,\sigma\rho}\right)$$

$$+\left(\frac{1}{2}g^{\sigma\mu}g_{\beta\beta}g_{\sigma\mu,\beta}g_{\nu\mu,\beta\rho}\right)+\left(\frac{1}{2}g^{\sigma\mu}g_{\beta\beta}g_{\sigma\mu,\beta\rho}g_{\nu\mu,\beta}\right)$$

$$+\left(\frac{1}{2}g^{\sigma\mu}_{,\rho}g_{\beta\beta}g_{\sigma\mu,\beta}g_{\nu\mu,\beta}\right)+\left(\frac{1}{2}g^{\sigma\mu}g_{\beta\beta,\rho}g_{\sigma\mu,\beta}g_{\nu\mu,\beta}\right)=0. \qquad (37)$$

Adding (35), (36) and (37), we get

$$g^{\mu\nu\rho\sigma,\mu\nu}+g^{\mu\nu}_{,\sigma}\left(g_{\rho\mu,\nu}-\frac{1}{2}g_{\mu\nu,\rho}\right)+g^{\mu\nu}_{,\rho}\left(g_{\sigma\mu,\nu}-\frac{1}{2}g_{\mu\nu,\sigma}\right)$$

$$+\left(\frac{1}{2}g^{\rho\mu}g_{\beta\beta}g_{\rho\mu,\beta}g_{\nu\mu,\beta\sigma}\right)$$

$$+\left(\frac{1}{2}g^{\rho\mu}g_{\beta\beta}g_{\rho\mu,\beta\sigma}g_{\nu\mu,\beta}\right)$$

$$+\left(\frac{1}{2}g^{\rho\mu}_{,\sigma}g_{\beta\beta}g_{\rho\mu,\beta}g_{\nu\mu,\beta}\right)$$

$$+\left(\frac{1}{2}g^{\rho\mu}g_{\beta\beta,\sigma}g_{\rho\mu,\beta}g_{\nu\mu,\beta}\right)$$

$$+\left(\frac{1}{2}g^{\sigma\mu}g_{\beta\beta}g_{\sigma\mu,\beta}g_{\nu\mu,\beta\rho}\right)$$

$$+\left(\frac{1}{2}g^{\sigma\mu}g_{\beta\beta}g_{\sigma\mu,\beta\rho}g_{\nu\mu,\beta}\right)$$

$$+\left(\frac{1}{2}g^{\sigma\mu}_{,\rho}g_{\beta\beta}g_{\sigma\mu,\beta}g_{\nu\mu,\beta}\right)+\left(\frac{1}{2}g^{\sigma\mu}g_{\beta\beta,\rho}g_{\sigma\mu,\beta}g_{\nu\mu,\beta}\right)=0. \qquad (38)$$

The equation (38) describes the waves that move in the strong gravitational field. The numbers, 0, 1, 2, and 3, are assigned to each of μ, ν, ρ, σ and β of the equation (38); then by the rule of tensor, all terms are summed up. Then the mathematical forms of the equation (38) are derived by the functions shown in the equations from (26) to (29). Then according to the equation (25), non-diagonal components such as g_{12} and g_{23} vanish; then only diagonal components remain as shown in the equations from (39) to (41). Hereinafter this research focuses on the spatial distribution of the waves but doesn't consider the time. Time is taken into account only by the distorted distance, ρ, of the equation (32). Therefore, only three diagonal components of a 3×3 matrix that represent the spatial distribution of the waves are considered.

$$W_{11}=\frac{16}{9(\rho-\tau)^2}+\frac{8}{3(\rho-\tau)^2}+\frac{64m}{27\mu(\rho-\tau)^{\frac{11}{3}}}$$

$$-\frac{144m}{27\mu(\rho-\tau)^{\frac{13}{3}}}+\frac{8\mu^3}{9m}sin^2\theta+\frac{16\mu^3}{18m}. \qquad (39)$$

$$W_{22}=8cos^2\theta-8sin^2\theta. \qquad (40)$$

$$W_{33}=\frac{128}{9(\rho-\tau)^2}+16cot^2\theta. \qquad (41)$$

And,

$$W_{\mu\nu}=W_{11}+W_{22}+W_{33}$$

$$=c_1X_1+c_2X_2+...+c_lX_l. \qquad (42)$$

$16\mu^3/18m$ of the equation (39) is a constant term, and it is not included in the equation (42) for the simulation.

The algorithm to simulate the relative energy intensity of the waves is as follows:

It is assumed that the stress energy tensor, T, reflects the energy intensities of the waves as shown in the equation (43), where k is a constant.

$$W_{\mu\nu}=kT \qquad (43)$$

The idea of the stress energy tensor is taken from the classical mechanics [4] (p. 327).

Each of the coefficients, c_1, $c_2 \cdots c_l$, relates to a component of the waves such as $1/(\rho-\tau)^2$, $1/(\rho-\tau)^{3/11}$, and so on. With the equation (44) we calculate the coefficients, c_1, $c_2 \cdots c_l$.

$$H=kT-W_{\mu\nu}=kT-(c_1X_1+c_2X_2+...+c_lX_l)$$

$$=kT-(c_1\frac{1}{(\rho-\tau)^2}+c_2\frac{1}{(\rho-\tau)^{\frac{11}{3}}}-c_3\frac{1}{(\rho-\tau)^{\frac{13}{3}}}$$

$$+c_4sin^2\theta+c_5(cos^2\theta-sin^2\theta)+c_6cot^2\theta) \qquad (44)$$

These coefficients make a column vector, c, which holds l rows. For calculating c, X (a n×l matrix) is set as shown in (45), where each of X_1, X_2, $\cdots X_l$ is a vector that holds n rows, and n is also the number of rows of each column of X. In this simulation the value of n is 23 as later explained in the section of the input data, and l is 6 for the equation (44).

The vectors of the matrix X should hold the projected image of the energy intensity on the surface of the sphere in the spherical polar coordinates. (Note: The surface of the sphere refers to Figure 14 of [2].

$$X=[X_1 \quad X_2 \quad ... \quad X_l] \qquad (45)$$

Then H is set for matrix algebra as shown below.

$$H=kT-Xc \qquad (46)$$

To solve this equation, a constraint is given as shown below, where X' is the transposed matrix of X.

$$X'H=X'(kT-Xc)=0 \qquad (47)$$

The matrix algebra continues as shown below to calculate the values of c.

$$X'Xc=X'kT \qquad (48)$$

$$c=(X'X)^{-1}X'kT \qquad (49)$$

here $(X'X)^{-1}$ is an inverse matrix of $X'X$.

If the object rotates as shown in Figure 1, its coordinate system is transformed by the transformation matrix D of the Euler's angles [4] as shown below.

$$D=\begin{pmatrix} cos\phi & sin\phi & 0 \\ -sin\phi & cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (50)$$

To simulate the rotation of the object around the axis of z, the matrix, $W_{\mu\nu}$, is multiplied by D; then it is transformed as shown below.

$$D \cdot W_{\mu\nu}$$

$$=\begin{pmatrix} cos\phi & sin\phi & 0 \\ -sin\phi & cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{bmatrix} W_{11} & 0 & 0 \\ 0 & W_{22} & 0 \\ 0 & 0 & W_{33} \end{bmatrix}$$

$$=\begin{bmatrix} cos\phi \cdot W_{11} & sin\phi \cdot W_{22} & 0 \\ -sin\phi \cdot W_{11} & cos\phi \cdot W_{22} & 0 \\ 0 & 0 & W_{33} \end{bmatrix} \quad (51)$$

To calculate the energy intensity of the waves, the diagonal components are taken from the above transformed matrix, $D \cdot W_{\mu\nu}$, then a matrix (52) is set.

$$\begin{bmatrix} dW_{11} & 0 & 0 \\ 0 & dW_{22} & 0 \\ 0 & 0 & W_{33} \end{bmatrix}$$

$$=\begin{bmatrix} cos\phi \cdot W_{11} & 0 & 0 \\ 0 & cos\phi \cdot W_{22} & 0 \\ 0 & 0 & W_{33} \end{bmatrix} \quad (52)$$

Then an equation of H is set as shown below to calculate the coefficients of the diagonal components.

$$H=kT- (c_1 \frac{cos\ \phi}{(\rho-\tau)^2} +c_2 \frac{cos\phi}{(\rho-\tau)^{\frac{11}{3}}} -c_3 \frac{cos\phi}{(\rho-\tau)^{\frac{13}{3}}}$$
$$+c_4\ cos\phi\ sin^2\theta + c_5\ cos\phi(cos^2\theta - sin^2\theta)$$

$$+c_6 \frac{1}{(\rho-\tau)^2} + c_7 cot^2\theta) \quad (53)$$

Hereinafter the same procedure follows as explained from the equation (44) to (49) to calculate the coefficients.

To calculate the spin angular momentum, a matrix is set as shown in (54), by taking out the anti-symmetrical components of $D \cdot W_{\mu\nu}$ from (51).

$$\begin{bmatrix} 0 & sin\phi \cdot W_{22} & 0 \\ -sin\phi \cdot W_{11} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (54)$$

Then a column vector is set as shown below.

$$\begin{bmatrix} sin\ \phi\ W_{22} \\ -sin\ \phi\ W_{11} \\ 0 \end{bmatrix} = \begin{bmatrix} W_{22} \cdot d\Omega_3 \\ -W_{11} \cdot d\Omega_3 \\ 0 \end{bmatrix} \quad (55)$$

With the above column vector, an equation of H is set as shown below for the calculation of the coefficients.

$$H=kT- \{- c_1 \frac{-sin\phi}{(\rho-\tau)^2} -c_2 \frac{-sin\phi}{(\rho-\tau)^{\frac{11}{3}}} +c_3 \frac{-sin\phi}{(\rho-\tau)^{\frac{13}{3}}}$$

$$-c_4(-sin\phi)sin^2\theta + c_5\ sin\phi(cos^2\theta - sin^2\theta) \quad (56)$$

The spin angular momentum is to be projected on the flat surface, which is perpendicular to the curved spherical surface in the spherical polar coordinate system. (Note: The flat surface refers to Figure 15 of [2].

2.2. Input data

Figure 2 and Figure 3 show the input data that simulate the distortion of time and space, which we used also for the previous research [1, 2]. In these figures, r is the distance from the center of the strong gravity; t is the time to travel for the distance; τ is the distorted time by the equation (31), which expands and shrinks depending on r and t; and, ρ is the distorted distance by the equation (32), which expands and shrinks depending on t and r.
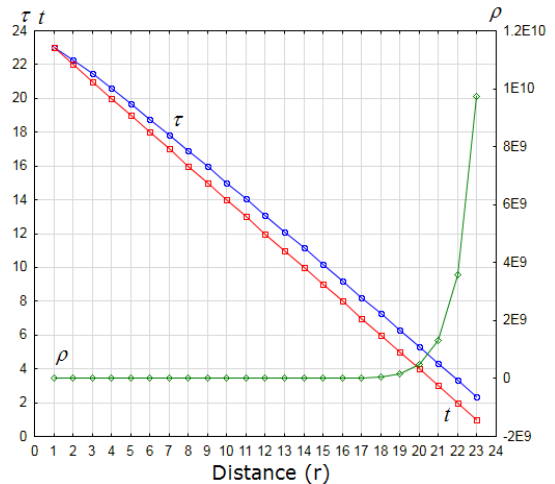
Figure 2: Time and distance from the center of the gravity, Case-1 (non-linear model): f(r)=log r and g(r)= e^r
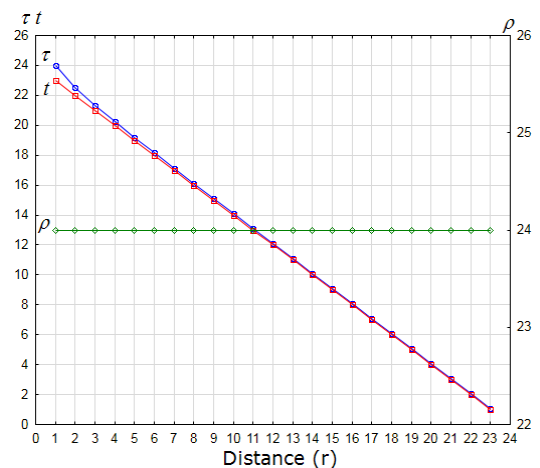
Figure 3: Time and distance from the center of the gravity, Case-2 (linear model): f(r) = 1/r and g(r) = r

For the distorted time and space, two models are set: Case-1 (non-linear model) and Case-2 (linear model) as shown in Table

1, which assign the functions of $f(r)$ of the equation (31) and $g(r)$ of the equation (32).

Table 1: Two models for simulating the distorted time-space

	Case-1 (Non-linear model)	Case-2 (Linear model)
$f(r)$	$log \cdot r$	$1/r$
$g(r)$	e^r	r

Then the distributions of r and t are set to simulate the description of the reference [3], "Any signal, even a light signal, would take an infinite time to cross the boundary of a black hole". In our research the term, a black hole, is replaced by the strong gravity that distorts time and space. However it is not possible to set the infinite value for the numeric simulation of this research that uses a personal computer. Therefore we used 24 discrete finite values as the surrogate that simulates "It takes more time to travel closer to the center of the strong gravity".

The time, t, and the distance, r, are set as shown in Figure 2 for Case-1 and Figure 3 for Case-2. For simulating the waves, we assumed that θ would become smaller in far distance from the object as shown in Figure 4, because the spin of the waves should stabilize the flow as the waves move forward from the center of the gravity; then the angle θ should be focused and smaller in the far distance.

For simulating the rotation of the object, two modes of the frequencies are set: ϕ_1 (rotation 1) and ϕ_2 (rotation 2) also as shown in Figure 4. The frequency ϕ_2 is higher than ϕ_1; therefore, rotation 2 represents the faster rotation of the object than rotation 1. With these settings, sin θ, cos θ, cot θ, sin ϕ_1, sin ϕ_2, cos ϕ_1 and cos ϕ_2 change their values as shown in Figure 5.

The value of the stress energy tensor is set as 1; therefore, this simulation doesn't calculate absolute values of the intensities, but relative intensities. The calculated coefficients have dimensions. For example, the coefficient of the first term of the equation (39), $16/\{9(\rho - \tau)^2\}$, has the dimension of the inverse of $(\rho - \tau)^2$ that is the squared dimension of the difference between the distorted distance, ρ, and the distorted time, τ.
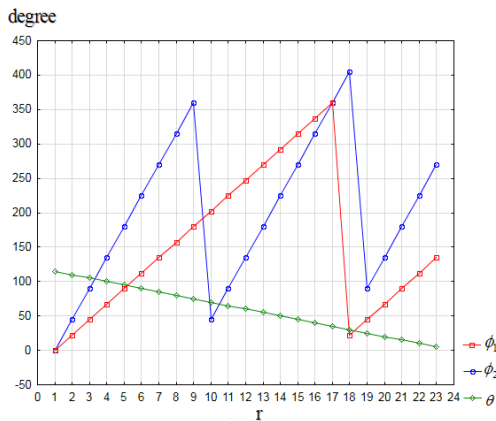
Figure 4: Angles, θ, ϕ_1 and ϕ_2 for the simulation
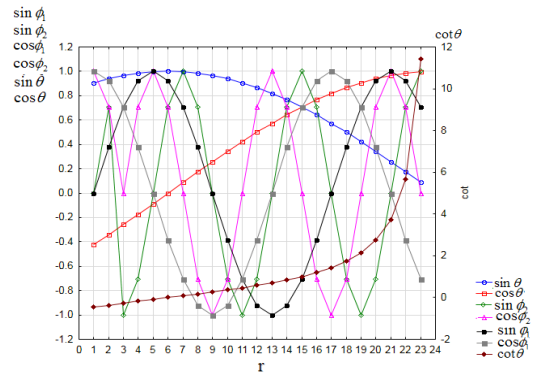
Figure 5: Trigonometric functions, sin θ, cos θ, cot θ, sin ϕ_1, sin ϕ_2, cos ϕ_1 and cos ϕ_2 for the simulation

3. Result

3.1. Energy intensity of the waves

The calculated energy intensities of the waves are shown in Table 2 for Case-1 (non-linear model) and Table 3 for Case-2 (linear model). The first and the third columns of these tables are the mathematical forms of the components of the waves, which are taken from the equations from (39) to (41), but each of the constant factors is simplified to unity. For example, $16/\{9(\rho - \tau)^2\}$ is simplified to $1/(\rho - \tau)^2$ because the factor, 16/9, is not necessary as it is to be included in the coefficient when it is calculated by the equation (49).

The calculated coefficients have either positive or negative values. For example, in Case-1 (non-linear model) the calculated coefficient of $1/(\rho - \tau)^2$ is $4.081 \cdot 10^{-6}$ with no rotation; and, it changes to $-1.465 \cdot 10^4$ with rotation 1, and then $4.178 \cdot 10^3$ with rotation 2. It suggests that the energy intensity of this component of the waves changes its sign from negative to positive when the object rotates faster from rotation 1 to rotation 2.

In order to interpret the calculated results, hereinafter we define that the positive flow holds positive (+) coefficient, and the negative flow holds negative (–) coefficient. For example in the above case of $1/(\rho - \tau)^2$, the wave is a negative flow with rotation 1; but, it changes to a positive flow with rotation 2.

Table 2: Results of the simulation of energy intensity of the waves, Case-1

Components of Waves (No rotation)	Coefficients (No rotation)	Components of Waves with the rotation	Coefficients (Rotation 1)	Coefficients (Rotation 2)
$1/(\rho - \tau)^2$ *	$4.081 \cdot 10^{-6}$	$cos\phi/(\rho - \tau)^2$	$-1.465 \cdot 10^4$	$4.178 \cdot 10^3$
$1/(\rho - \tau)^{11/3}$	$-1.892 \cdot 10^{-4}$	$cos\phi/(\rho - \tau)^{11/3}$	$1.682 \cdot 10^5$	$-2.311 \cdot 10^5$
$-1/(\rho - \tau)^{13/3}$	$-3.270 \cdot 10^{-4}$	$-cos\phi/(\rho - \tau)^{13/3}$	$2.836 \cdot 10^5$	$-4.028 \cdot 10^5$
$sin^2\theta$	2.000	$cos\phi sin^2\theta$	$8.397 \cdot 10^{-2}$	$3.875 \cdot 10^{-2}$
$cos^2\theta - sin^2\theta$	1.000	$cos\phi(cos^2\theta - sin^2\theta)$	1.436	$-6.610 \cdot 10^{-2}$
–	–	$1/(\rho - \tau)^2$ ⁂	$1.041 \cdot 10^4$	$3.825 \cdot 10^2$
$cot^2\theta$	$-4.358 \cdot 10^{-10}$	$cot^2\theta$ ⁂	$7.534 \cdot 10^{-2}$	$8.819 \cdot 10^{-2}$

Notes for Table 2 and 3.

*: This component belongs to both W_{11}, $(16/9(\rho - \tau)^2 + 8/3(\rho - \tau)^2$, of the equation (39) and W_{33}, $(128/9(\rho - \tau)^2$, of the equation (41).

✳: These components are on the axis of the rotation and belong to W_{33}; therefore, $\cos\phi$ is not multiplied.

Table 3: Results of the simulation of energy intensity of the waves, Case-2

Components of Waves (No rotation)	Coefficients (No rotation)	Components of Waves with the rotation	Coefficients (Rotation 1)	Coefficients (Rotation 2)
$1/(\rho-\tau)^2$ *	$-1.563\cdot10^{-7}$	$\cos\phi/(\rho-\tau)^2$	-17.34	-63.97
$1/(\rho-\tau)^{11/3}$	$2.206\cdot10^{-6}$	$\cos\phi/(\rho-\tau)^{11/3}$	$-1.313\cdot10^2$	$1.156\cdot10^3$
$-1/(\rho-\tau)^{13/3}$	$2.493\cdot10^{-6}$	$-\cos\phi/(\rho-\tau)^{13/3}$	$-1.460\cdot10^2$	$1.390\cdot10^3$
$\sin^2\theta$	2.000	$\cos\phi\sin^2\theta$	0.1066	$6.768\cdot10^{-2}$
$\cos^2\theta-\sin^2\theta$	1.000	$\cos\phi(\cos^2\theta-\sin^2\theta)$	0.4353	-0.5905
–	–	$1/(\rho-\tau)^2$ ✳	28.03	12.58
$\cot^2\theta$	$-4.306\cdot10^{-10}$	$\cot^2\theta$ ✳	$8.474\cdot10^{-2}$	$8.236\cdot10^{-3}$

The summations of the values of each of the positive coefficients and the negative coefficients are shown in Table 4 for both Case-1 and Case-2, and Figure 6 for Case-1 and Figure 7 for Case-2. In both cases, positive and negative flows of the waves are not visible in the figures when the object doesn't rotate; but, they become visible when the object rotates.

Table 4: Energy intensities of positive and negative flows

	Positive flow	Negative flow
No rotation (Case-1)	3.000	$-4.358\cdot10^{-10}$
Rotation 1 (Case-1)	$4.622\cdot10^5$	$-1.465\cdot10^4$
Rotation 2 (Case-1)	$4.560\cdot10^3$	$-6.339\cdot10^5$
No rotation (Case-2)	3.000	$-1.567\cdot10^{-7}$
Rotation 1 (Case-2)	$2.865\cdot10^1$	$-2.947\cdot10^2$
Rotation 2 (Case-2)	$2.558\cdot10^3$	$-6.456\cdot10^1$
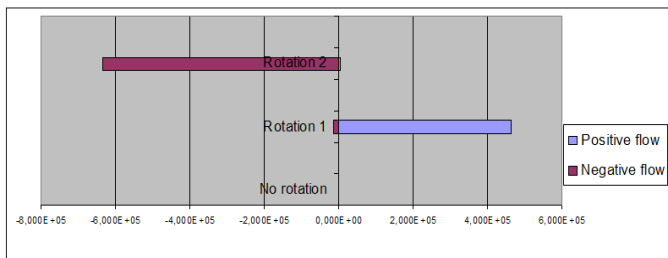
Figure 6: Energy intensity of positive and negative flows: (Case-1)
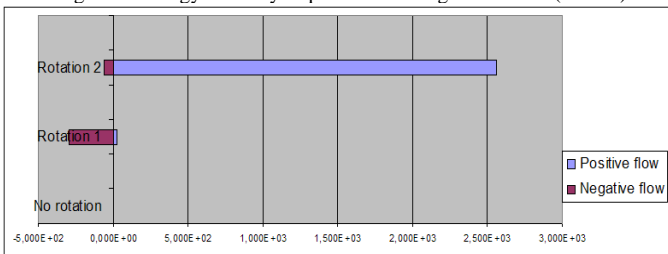
Figure 7: Energy intensity of positive and negative flows: (Case-2)

3.2. Spin angular momentum

Table 5 and Table 6 show the calculated relative intensity of the spin angular momentum of the waves, which are projected in the directions of $\sin\phi\cdot W_{22}$ and $-\sin\phi\cdot W_{11}$. (Note: $\sin\phi\cdot W_{22}$ and $-\sin\phi\cdot W_{11}$ are perpendicular to the rotation axis.)

The calculated coefficients of the spin angular momentum hold either positive or negative signs. For example, in Case-1 (non-linear model) the coefficient of $-\sin\phi/(\rho-\tau)^2$ is $-6.003\cdot10^3$ with the rotation 1, but it changes to $7.215\cdot10^3$ with the rotation 2. $-\sin\phi/(\rho-\tau)^2$ is one of the projected momenta of $1/(\rho-\tau)^2$ in the direction of $-\sin\phi\cdot W_{11}$, where $1/(\rho-\tau)^2$ represents the first and the second terms of the equation (39), $16/\{9(\rho-\tau)^2\}$ and $8/\{3(\rho-\tau)^2\}$.

Then we compare the changes of the spin angular momentum with the changes of the energy intensity. The energy intensity of $1/(\rho-\tau)^2$ in Table 2 for Case-1 (non-linear model) changes its sign from negative to positive when the rotation of the object becomes faster from rotation 1 to rotation 2. It suggests that this component is a negative flow of the waves when the object rotates with the frequency of ϕ_1, but when the object rotates faster with the frequency of ϕ_2, the waves become a positive flow; and, the spin angular momentum also changes its spinning direction from negative to positive as shown in Table 5.

Table 5: Relative intensity of spin angular momentum, Case-1

Components of the waves	Rotation 1	Rotation 2
$-\sin\phi/(\rho-\tau)^2$	$-6.003\cdot10^3$	$7.215\cdot10^3$
$-\sin\phi/(\rho-\tau)^{11/3}$	$1.400\cdot10^6$	$-1.734\cdot10^6$
$-\sin\phi/(\rho-\tau)^{13/3}$	$4.592\cdot10^6$	$-5.688\cdot10^6$
$-\sin\phi\sin^2\theta$	$-1.163\cdot10^{-2}$	$-1.234\cdot10^{-2}$
$\sin\phi(\cos^2\theta-\sin^2\theta)$	0.2898	-0.5123

Table 6: Relative intensity of spin angular momentum, Case-2

Components of the waves	Rotation 1	Rotation 2
$-\sin\phi/(\rho-\tau)^2$	-71.08	7.644
$-\sin\phi/(\rho-\tau)^{11/3}$	$5.946\cdot10^2$	$1.031\cdot10^2$
$-\sin\phi/(\rho-\tau)^{13/3}$	$6.159\cdot10^2$	$1.605\cdot10^2$
$-\sin\phi\sin^2\theta$	$-2.250\cdot10^{-4}$	$-1.270\cdot10^{-2}$
$\sin\phi(\cos^2\theta-\sin^2\theta)$	1.129	-0.5360

The summations of the values for positive and negative coefficients of the spin angular momenta are shown in Table 7 for Case-1 and Case-2, and Figure 8 for Case-1 and Figure 9 for Case-2. In both cases, the values change upon the selection of the rotation-mode of the object, rotation 1 or rotation 2. As shown in

the figures, the spin angular momentum changes its sign between rotation 1 and rotation 2 in Case-1 (non-linear model), while such a rule of changing sign is not so clear in Case-2 (linear model).

Table 7: Relative intensity of spin angular momentum

	Positive flow	Negative flow
Case-1 Rotation 1	$5.992 \cdot 10^6$	$-6.003 \cdot 10^3$
Case-1 Rotation 2	$7.215 \cdot 10^3$	$-7.421 \cdot 10^6$
Case-2 Rotation 1	$1.212 \cdot 10^3$	$-7.108 \cdot 10^1$
Case-2 Rotation 2	$2.713 \cdot 10^2$	$-5.486 \cdot 10^2$
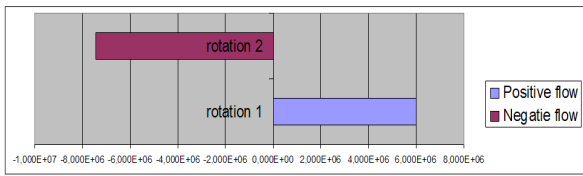
Figure 8: Relative intensity of spin angular momentum Case-1
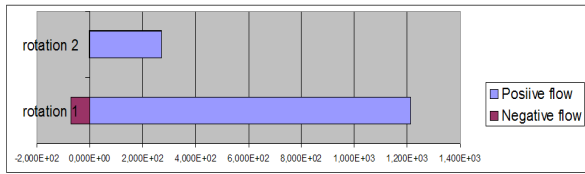
Figure 9: Relative intensity of spin angular momentum Case-2

3.3. Discussion

The results show that positive and negative flows of the waves appear if the object rotates, and some of the components that indicate positive and negative flows have opposite spinning directions. The results also show that the rotation-mode (rotation 1 or rotation 2) of the object relates to the energy intensity of the emitted waves, as well as their spinning directions.

Table 8 and Table 9 for Case-1 (non-linear model) and Table 10 and Table 11 for Case-2 (linear model) summarize the changes of the signs of the energy intensity and the spin angular momentum.

In Case-1 (non-linear model), except $sin^2\theta$, these components change their energy intensities either from positive to negative or from negative to positive, and simultaneously change their signs of the spin angular momenta either from positive to negative or from negative to positive. For example, the component of $1/(\rho-\tau)^2$ is a negative flow with rotation 1, but it is a positive flow with rotation 2; while its spinning direction changes from negative to positive when the rotation-mode changes from rotation 1 to rotation 2.

The coefficients of $1/(\rho-\tau)^{11/3}$, $-1/(\rho-\tau)^{13/3}$ and $cos^2\theta-sin^2\theta$ show the same rule of changing sign, but in the opposite directions of the flow and the spin of $1/(\rho-\tau)^2$.

This finding in Case-1 (non-linear model) is consistent with Figure 10 of our previous research [5] upon the analysis in rectilinear coordinates, in which the negative flow has the clockwise spin, while the positive flow has the anti-clockwise spin.

The exception, $sin^2\theta$, relates to $cos\phi sin^2\theta$ of Table 2 for the energy intensity and to $sin\phi sin^2\theta$ of Table 5 for the spin angular momentum. The absolute values of these calculated coefficients are smaller than those of the other components; and, the flow and the spin don't change their signs. Therefore this component is not considered significant.

In Case-2 (linear model) shown in Table 10 and Table 11, the rule of changing sign as observed in Case-1 applies only to $cos^2\theta-sin^2\theta$.

Table 8: The sign of the coefficient, energy intensity, Case-1

	No rotation	Rotation 1	Rotation 2
$1/(\rho-\tau)^2$	+	−	+
$1/(\rho-\tau)^{11/3}$	−	+	−
$-1/(\rho-\tau)^{13/3}$	−	+	−
$sin^2\theta$	+	+	+
$cos^2\theta-sin^2\theta$	+	+	−

Note for Tables 8, 9, 10 and 11: The component of $cot^2\theta$ doesn't appear in these tables because it is on the rotation axis, and not considered. $1/(\rho-\tau)^2$ is a simplified form of $16/9(\rho-\tau)^2$ and $8/3(\rho-\tau)^2$; $-1/(\rho-\tau)^{11/3}$ is a simplified form of $-64m/27\mu(\rho-\tau)^{11/3}$; $sin^2\theta$ is a simplified form of $8\mu^3 sin^2\theta/18m$ of the equation (39); and $cos^2\theta-sin^2\theta$ is a simplified form of $8cos^2\theta-8sin^2\theta$ of the equation (40). All these components are multiplied by $cos\phi$ for calculating the energy intensity and by $-sin\phi$ and $sin\phi$ for calculating the spin angular momentum.

Table 9: The sign of the coefficient, Spin angular momentum, Case-1

	Rotation 1	Rotation 2
$-sin\phi/(\rho-\tau)^2$	−	+
$-sin\phi/(\rho-\tau)^{11/3}$	+	−
$-sin\phi/(\rho-\tau)^{13/3}$	+	−
$-sin\phi sin^2\theta$	−	−
$sin\phi(cos^2\theta-sin^2\theta)$	+	−

Table 10: The sign of the coefficient, energy intensity, Case-2

	No rotation	Rotation 1	Rotation 2
$1/(\rho-\tau)^2$	−	−	−

$1/(\rho-\tau)^{11/3}$	+	−	+
$-1/(\rho-\tau)^{13/3}$	+	−	+
$sin^2\theta$	+	+	+
$cos^2\theta-sin^2\theta$	+	+	−

Table 11: The sign of the coefficient, Spin angular momentum, Case-2

	Rotation 1	Rotation 2
$-sin\phi/(\rho-\tau)^2$	−	+
$-sin\phi/(\rho-\tau)^{11/3}$	+	+
$-sin\phi/(\rho-\tau)^{13/3}$	+	+
$-sin\phi sin^2\theta$	−	−
$sin\phi(cos^2\theta-sin^2\theta)$	+	−

This particular finding on the spin angular momentum is consistent with the previous research [5]. However, with the linear model, most of the components of the waves don't follow the same rule of changing sign.

A parallel image of the source of the anti-gravity is a rotating black hole, and the same rule may apply. Then the positive flow of the waves may be gravitational waves and the negative flow may be the gravitational waves that flow in opposite directions. If those gravitational waves are observed from a rotating black hole and if they indicate the same rule of the energy intensity and the angular momentum as shown in this paper, it will certify the theory.

On the other hand, the theory of the time-space distortion should be verified with a compact nuclear fusion reactor for developing the flying craft. At the next step of the research, it is necessary to calculate the probability of proton-capture or proton-charge-exchange with the time-space distortion model that is used for the simulation reported in this paper.

References

[1] Y. Matsuki, P.I. Bidyuk, "Theory and Simulation of Artificial Antigravity," 2020 IEEE 2nd International Conference on System Analysis Intelligent Computing, 2020, doi: 10.1109/SAIC51296.2020.9239195.
[2] Y. Matsuki, P.I. Bidyuk, "Numeric Simulation of Artificial Antigravity upon General Theory of Relativity," Advances in Science, Technology and Engineering Systems Journal, **6**(3), 45-53, 2021, doi: 10.25046/aj060307.
[3] P.A.M. Dirac, General Theory of Relativity, New York: Florida University, A Wiley Inter-science Publication, John Wiley & Sons, 1975.
[4] H. Goldstein, C.P. Poole, J.L. Safko, Classical Mechanics, 3rd Edition published by Pearson Education. Inc., 2002.
[5] Y. Matsuki, P.I. Bidyuk, "Analysis of Negative Flow of Gravitational Waves," System Research & Information Technology, N4, 7-18, 2019, doi: 10.20535/SRIT.2308.8893.2019.4.01.

Figure 10:   Geometrical relation between positive and negative flows in flat space (from [5])

4.   Conclusion and recommendation

The result indicates that the rotation of the object can produce both positive and negative energy intensities of the waves, and they are interpreted as positive and negative flows. Both positive and negative flows appear when the object rotates, although the selection of the input data, the non-linear model or the linear model, leads to differences of the energy intensities.

With the non-linear model, the calculated spin angular momentum of the waves changes its sign, + or −, when most of the components change the signs of their energy intensities either from positive to negative or from negative to positive, depending on the speed of the object's rotation. This finding means that most of the components of the waves change their spinning directions when the energy intensity of the flow of the waves changes its sign + or −, also depending on the speed of the object's rotation.

Acoustic Scene Classifier Based on Gaussian Mixture Model in the Concept Drift Situation

Ibnu Daqiqil Id[*], Masanobu Abe, Sunao Hara

Graduate School of Interdisciplinary Science and Engineering in Health Systems, Okayama University, Okayama, 700-0001, Japan

A R T I C L E I N F O

A B S T R A C T

The data distribution used in model training is assumed to be similar with that when the model is applied. However, in some applications, data distributions may change over time. This situation is called the concept drift, which might decrease the model performance because the model is trained and evaluated in different distributions. To solve this problem for scene audio classification, this study proposes the kernel density drift detection (KD3) algorithm to detect the concept drift and the combine–merge Gaussian mixture model (CMGMM) algorithm to adapt to the concept drift. The strength of the CMGMM algorithm is its ability to perform adaptation and continuously learn from stream data with a local replacement strategy that enables it to preserve previously learned knowledge and avoid catastrophic forgetting. KD3 plays an essential role in detecting the concept drift and supplying adaptation data to the CMGMM. Their performance is evaluated for four types of concept drift with three systematically generated scenarios. The CMGMM is evaluated with and without the concept drift detector. In summary, the combination of the CMGMM and KD3 outperforms two of four other combination methods and shows its best performance at a recurring concept drift.

1. Introduction

Human–computer interaction through audition requires devices to recognize the environment using acoustic sound analysis. One of the primary research topics in this area is acoustic scene classification (ASC), which attempts to classify digital audio signals into mutually exclusive scene categories. ASC is an important area of study covering various applications, including smart homes, context-aware audio services, security surveillance, mobile robot navigation, and wildlife monitoring in natural habitats. Machine audition applications have a high potential to lead to more innovative context-aware services.

We intend to develop an ASC system for environmental or scene audio in specific locations (i.e., beach, shop, bus station, and airport) with different acoustic characteristics. The scene audio contains an ensemble of background and foreground sounds. One of the most important aspects of the audio scene in real life is the concept drift [1], whose data distribution might evolve or change in the future. For example, the foreground event sounds in a bus station, such as an ambulance siren, wind noise, and rustling sounds, might change because of the physical environment, human activities, or nature [2]. These changes cause the acoustic data

distributions to change, potentially causing a lower performance in the trained model [3].

The simplest solutions for handling the abovementioned problems to maintain the model performance are periodic retraining and redeployment of the model. Nevertheless, these solutions can be time consuming and costly. Moreover, the decision for the frequency of retraining and redeployment is a difficult task. Another promising approach is to use an evolving or incremental learning method [4], [5], where the model is updated when a new subset of data arrives [6]. Each iteration is considered as an incremental step toward revisiting the current model.

In this study, we propose a combine–merge Gaussian mixture model (CMGMM) and kernel density drift detection (KD3) to solve the concept drift problem [7]. The CMGMM is an algorithm based on the Gaussian mixture model (GMM) that adapts to the concept drift by adding or modifying its components to accommodate the emerging concept drift. The algorithm's advantages are adaptation and continuous learning from stream data with a local replacement strategy to preserve previously learned knowledge and avoid catastrophic forgetting.

In [7], we compared the CMGMM to the incremental GMM (IGMM) [8] and KD3 to adaptive windowing (ADWIN) [9] and HDMM [10] in two approaches, namely the active and passive

*Corresponding Author: Ibnu Daqiqil Id, Okayama, Japan, daqiqil@s.okayama-u.ac.jp

approaches. In the active approach, the concept drift is detected using a specific algorithm, then adapts the model. In the passive approach, the model is continuously adapted at a specific fixed interval. The result is that the combination of CMGMM and KD3 outperforms other combinations in two of three evaluation scenarios.

The work described herein extends and improves that of previous publications [7,11] in the following respects:

- The algorithm has been modified to use component pruning to overcome the overfitting problem and support the Scikit-Multiflow Framework [12].
- The KD3 hyperparameter is optimized, and the algorithm is evaluated using prequential evaluation for better results and online performance monitoring in several concept drift types and scenarios.

The rest of this paper is organized as follows: Section 2 presents the related work of this research and the fundamental equations used in our proposed solutions; Section 3 describes the proposed CMGMM and KD3; Section 4 outlines the experimental setup; Section 5 discusses the experimental results; and finally, Section 6 provides our conclusions.

2. Related Work

This section provides a brief overview of the related work on the concept drift and the fundamental theory used in the proposed method.

2.1. Concept Drift in Audio Scenes

A concept is defined as a set of object instances [13]. Probabilistically, a concept is defined using prior class probabilities $p(y)$ and class conditional probabilities $p(X|y)$ [4]. $p(y)$ and $p(X|y)$ determine the joint distribution $p(X, y)$ [3]; hence, a concept is defined as the joint probability distribution of a set of input features X and the corresponding label y in dataset $\mathfrak{D}$. In this paper, X is an acoustic scene sound defined as a mixture of specific event sounds $(\hat{x})$ perceived and defined by humans [14], and y is the label of X. X has numerous types of acoustic event sounds and background noises $\hat{x}$ that often overlap with each other. In other words, the relationship between of X and $\hat{x}$ determines $p(X, y)$, $X \in (\hat{x}_1, \hat{x}_2, \hat{x}_3, ..\hat{x}_i,)$, where i denotes the number of $\hat{x}$ in X.

In the future, the relationship of $\hat{x}$ in X might change, which then changes the relationship of $p(X, y)$. For example, another event sound might appear, or some existing event sounds may disappear. This situation is called the concept drift, which is expressed as follows:

$$\exists X: p_{w_0}(X, y) \neq p_{w_n}(X, y). \tag{1}$$

Eq. (1) and Figure 1 describe concept drift as the change in the joint probability distribution between two-time windows, w_0 and w_n. Models built on previous data at w_0 might not be suitable for predicting new incoming data at w_n. This change may be caused by a change not only in the number of $\hat{x}$, but also in the underlying data distribution of $\hat{x}$. These changes require model adaptation because the model's error may no longer be acceptable with the new data distribution [15].
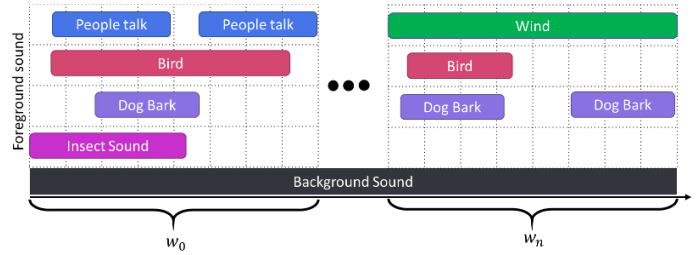
Figure 1: Illustration of the concept drifts in an acoustic scene audio at a park

The change in the incoming data at w_n depends on a variety of different internal or external influences (e.g., event sounds that exist in a park depending on the season). The initial data recorded in the winter may only consist of people talking, bird calls, and dogs barking. However, the event sounds change in the summer, and new event sounds, such as insect and wind sounds, emerge.

2.2. Concept Drift Detection Methods

Several methods have been proposed to detect concept drifts from a data stream. This study focuses on window-based methods that use fixed windows as a reference for summarizing previous information. This approach has more accurate results than other more straightforward methods, such as cumulative sum [16]. However, the computational time and space used are higher [17]. This approach usually utilizes statistical tests or mathematical inequalities to compute the change in data. Some of the state-of-the-art methods used in this paper is presented below:

- ADWIN is a sliding window-based concept drift detection algorithm. The size of ADWIN windows SW might change depending on the instance in the distribution. Its size increases when the instance in the stream continues in the same distribution. SW shrinks when distribution changes occur [9]. ADWIN detects concept drifts when the averages between these windows are higher than a given threshold.

- The HDDM is a concept drift detection algorithm based on fixed windows and probability inequalities [10]. The author proposes two types of HDDM, namely HDDMA and HDDMW. The HDDMA uses moving averages, whereas the HDDMW uses weighted moving averages to detect the concept drift. The HDDMA is suitable for detecting the abrupt concept drift, whereas the HDDMW is suitable for detecting the gradual concept drift.

- KSWIN [18] is a window-based concept drift detection method that utilizes the Kolmogorov–Smirnov statistic test (KS-Test) to compare the distances of two distributions. This test is a non-parametric test that does not require any assumptions about the underlying data distribution.

Each method has its optimal hyperparameters, which differ based on the datasets used and the type of drift in those datasets.

2.3. Merging Gaussian Mixture Component

A component of a Gaussian distribution is represented by (w, μ, P) and $\{(w_1, \mu_1, P_1), (w_2, \mu_2, P_2), ... (w_n, \mu_n, P_n)\}$ to denote a mixture of n Gaussian components, where w, μ, and P are the weight or prior probability, distribution means, and covariance matrix, respectively. This mixture must satisfy $w_1 + w_2 + \cdots + w_n = 1$ and has the pdf defined in Eq. (2).

$$f(X) = \sum_{i=1}^{n} w_i N(x; \mu_i, P_i), \qquad (2)$$

where,

$$N(x; \mu_i, P_i) = \frac{1}{\sqrt{(2\pi)^d \det P_i}} e^{\left[-\frac{1}{2}(x-\mu_i)^T P_i^{-1}(x-\mu_i)\right]}. \qquad (3)$$

Suppose we wish to merge two Gaussian components $\{(w_i, \mu_i, P_i), (w_j, \mu_j, P_j)\}$, where $w_i + w_j \leq 1$, and approximate the result as a single Gaussian. The new Gaussian candidate $(w_{ij}, \mu_{ij}, P_{ij})$ must preserve the zeroth-, first-, and second-order moments of the original Gaussian. The moment-preserving merge is shown in Eqs. (4)–(6).

$$w_{ij} = w_i + w_j \qquad (4)$$

$$\mu_{ij} = \frac{w_i}{w_i + w_j} \mu_i + \frac{w_j}{w_i + w_j} \mu_j \qquad (5)$$

$$P_{ij} = \frac{w_i}{w_i + w_j} P_i + \frac{w_j}{w_i + w_j} P_j \qquad (6)$$
$$+ \frac{w_i w_j}{(w_i + w_j)^2} (\mu_i - \mu_j)(\mu_i - \mu_j)^T$$

2.4. Kullback–Leibler (KL) Dissimilarity

Kullback–Leibler (KL) discrimination, known as KL divergence or relative entropy, is a tool for measuring the discrepancy between two probability distributions. The KL discrimination between $f(x)$, a probability distribution for random variables X and $g(x)$, another probability distribution is the expected value of the log-likelihood ratio. The KL divergence of $f(x)$ and $g(x)$ is defined in Eq. (7), where $\Re^d$ is the sample space of the random variable X.

$$d_{KL}(f, g) = \int_{\Re^d} f(x) \log \frac{f(x)}{g(x)} dx. \qquad (7)$$

Based on Eq. (7), KL is not a perfect distance metric because it is asymmetric $d_{KL}(f, g) \neq d_{KL}(g, f)$ and does not satisfy the triangle inequality, $d_{KL}(f, g) + d_{KL}(g, h) \geq d_{KL}(f, h)$. However, we can use it as metric distance because the KL discrimination of two probability distributions is larger than zero, $d_{KL}(f, g) \geq 0$, and the discrimination of the identical two probability distributions is zero, $d_{KL}(f, f) = 0$.

Accordingly, we apply the KL dissimilarity by computing the Kullback–Leibler discrimination upper bound of the post-merge mixture with respect to the pre-merge mixture. In the case of the Gaussian mixture, where $f(x) = N(w_1, \mu_1, P_1), g(x) = N(w_2, \mu_2, P_2)$ and $w_1 + w_2 < 1$, the KL dissimilarity between $f(x)$ and $g(x)$ is shown in Eq. (8).

$$d_{KL}(\{(w_1, \mu_1, P_1)\}, \{(w_2, \mu_2, P_2)\}). \qquad (8)$$
$$= \frac{1}{2} \Big[tr(P_2^{-1}[P_1 - P_2 + (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T]) + \log \frac{\det(P_2)}{\det(P_1)} \Big]$$

Please refer to [19] for more details about the KL dissimilarity of the Gaussian distributions.

3. Proposed Method

This study extends the combine–merge Gaussian mixture model (CMGMM) [7] to classify audio scenes in concept drift situations. The CMGMM was developed based on the GMM algorithm that can incrementally adapt to the new identified component. This algorithm can add new components as new concepts and update existing components as a response to the change of the current existing concept in the current data. The CMGMM implementation is available in our public repository[1], with the algorithm pipeline shown in Figure 2.
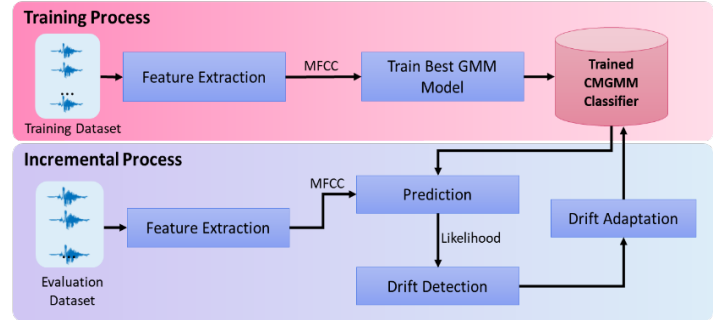
Figure 2: Combine–merge Gaussian mixture model (CMGMM) general workflow

In the training process, we extract the feature of the scene audio from the training dataset D_0 and train an optimal model $M_{optimal}$. We use the Expectation maximization (EM) [20] algorithm to train the model and the Bayesian information criterion (BIC) [21] to select the best model.

In the incremental process, the $M_{optimal}$ performance is observed through the prediction likelihood. When KD3 detects a significant likelihood change, the model activates the concept drift adaptation process. The concept drift adaptation process then begins by creating a local model $M_{drifted}$ from the new coming data. $M_{drifted}$ represents the new concepts or concept updates in the incoming data. Finally, we combine the $M_{optimal}$ and $M_{drifted}$ components to include any new concepts from $M_{drifted}$ that may not exist in the $M_{optimal}$ at the initial training and merge similar components to update the existing component in $M_{optimal}$.

The CMGMM pipeline process is detailed in the subsections that follow.

3.1. Feature Extraction

Feature extraction is the first step of both the training and incremental processes. In this research, we use normalized Mel-frequency cepstral coefficients (MFCCs) that represent the short-term power spectrum of audio in the frequency domain of the Mel scale. MFCCs are commonly used as features in audio processing and speech recognition. The first step is pre-emphasis for enhancing the quantity of energy in high frequencies. The next step is windowing the signal and computing the fast Fourier transformation to transform the sample from the time domain to the frequency domain. Subsequently, the frequencies are wrapped

---

[1]https://github.com/ibnudaqiqil/CMGMM

on a Mel scale, and the inverse DCT is applied [22]. Finally, each of the MFCCs is normalized using mean and variance normalization based on Eq. (9):

$$MFCC_{norm} = \frac{(MFCC - \mu)}{S}, \qquad (9)$$

where, μ and S denote the mean and the standard deviation of the training samples, respectively.

3.2. Model Training

The training process is intended to build a set of models from the training dataset D_0 containing training data $x = \{x_1, x_2, ..., x_n\}$, where x_n denotes the MFCC vector. The models are trained Q times using the EM algorithm. For each training cycle, a different number of components K ranging from K_{min} to K_{max} is used, where $Q = K_{max} - K_{min}$. Consequently, a set of models $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, ., \mathcal{M}_Q\}$ is obtained based on the different numbers of components.

The next step is model selection using the BIC. In [23], the BIC value of a model $\mathcal{M}_K$ trained over the dataset X with K components, BIC(X, M_K), is defined as follows:

$$BIC(X, M_k) \equiv -2 \log L(X, M_K) + v \log N , \quad (10)$$

where, L denotes the model likelihood; v denotes the degree of freedom of the model parameters; and N denotes the number of training data points. The model with the lowest BIC value is selected because it maximizes the log-likelihood [6]. Algorithm 1 presents the steps of the learning process.

Algorithm 1: Training the Optimal Model

Input: Initial Dataset D$_{init}$, Minimum Component Number
K$_{min}$, and Maximum Component Number Kmax

Result: Best GMM Model

BIC$_{best}$ = ∞

for K$_{min}$ to K$_{max}$ **do**

    M$_{candidate}$= EMTrain(D$_{init}$, k)

    BIC$_{candidate}$ = ComputeBIC (D$_{init}$, M$_{candidate}$)

    if BIC$_{candidate}$ < BIC$_{best}$ **then**

      | M$_{best =}$ M$_{candidate}$

    end

end

return M$_{best}$

3.3. Concept Drift Detection

We propose Kernel Density Drift Detection (KD3) to detect the concept drift. KD3 is a window-based algorithm for concept drift detection. It works based on estimating the window density using the Kernel Density Estimation (KDE) or the Parzen's window [24]. The KDE is a non-parametric probability density estimator that automatically estimates the shape of the data density without assuming the underlying distribution. The concept drift can be detected by comparing the probability functions between these windows. The greater the variation between the windows, the more evidence obtained for the concept drifts. Aside from detecting concept drifts, KD3 also collects data for adaptation (D$_{drift}$) by identifying a warning zone when data begin to show indications of concept drift.

KD3 requires three hyperparameters, namely α, β, and h, which denote the margins for detecting the concept drift and

accumulating the density distance and the window length, respectively. α is used to determine the threshold of the density variation in the concept drift, while β is employed to determine the threshold of the density variation in the warning zone. Therefore, α must be greater than β. KD3 accepts a set of likelihood windows z_c as input. z_c is the current likelihood window that contains a sequence of log-likelihood ℓ from the model prediction, $z_c = \{\ell_1, \ell_2, \ell_3, ..., \ell_h\}$.

First, this algorithm aims to estimate the density (f_{kde}) of the current z_c and previous z_{c-1} windows. Let ℓ_n be the latest generated ℓ. Let z_c contain the h-latest ℓ from ℓ_n, $z_c \in [\ell_{n-h}, \ell_n]$, and let z_{c-1} contain the h-latest ℓ from ℓ_{n-h}, $z_{c-1} \in [\ell_{n-2h}, \ell_{n-h}]$. To detect a concept drift, the distance $\dot{d}_t$ between f_{kde} of z_c and z_{c-1} is computed using Eq. (11) within the bounds of $b1$ and $b2$. The bounds are computed based on the maximum and minimum values of the joined ℓ of z_c and z_{c-1}.

$$\dot{d}_t = \frac{1}{2} \int_{b1}^{b2} |f_{kde}(z_c) - f_{kde}(z_{c-1})| \, dz, \text{where} \qquad (11)$$

$$z_c \in [\ell_{n-h}, \ell_n], \; z_{c-1} \in [\ell_{n-2h}, \ell_{n-h}],$$

$$b1 = \min(\ell_{n-2h}, \ell_n), b2 = \max(\ell_{n-2h}, \ell_n).$$

Finally, the algorithm compares $\dot{d}_t$ to α and β.hen $\dot{d}_t > \beta$. Suppose that the accumulative distance is equal to or greater than α. In that case, the algorithm sends the collected data to the model for adaptation. Figure 3 and Algorithm 2 illustrate the detailed KD3 process.
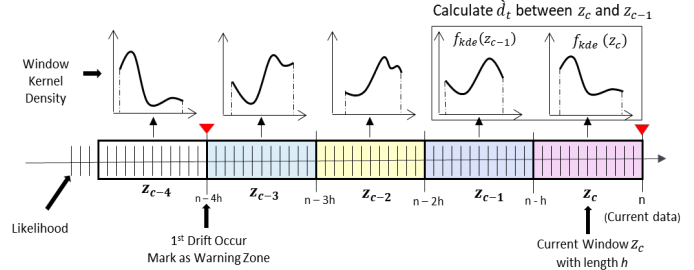
Figure 3: Illustration of the Kernel density drift detection (KD3) concept

---

Algorithm 2: Detecting the Concept Drift

Input: Set of t likelihood ℓ, drift margin α, warning margin β (α
    > β), window length h,

Result: Drift Concept Signal, Drifted Dataset (D$_{drift}$)

Window$_1$ = $\ell[c - h : c]$;

Window$_2$ = $\ell[c - 2h : c - h]$;

B$_{min}$, B$_{max}$ = CalculateWindowBound($\ell[c - 2h : c]$);

KDE$_1$ = EstimateKDE(Window$_1$)

KDE$_2$ = EstimateKDE(Window$_2$)

diff = distance(KDE$_1$, KDE$_2$, B$_{min}$, B$_{max}$)

if (diff ≥ α) **then**

    resetWarningZoneData()

    return true, $[c - h : c]$

end

if (diff ≥ β) **then**

    accumulativeWarning += diff

    if (accumulativeWarning ≥ α) **then**

      | resetWarningZoneData();

```
        return true, [c − h: c]
    end
    return false, [c − h: c]
end
return false, null
```

3.4. Model Adaptation

Model adaptation aims to revise the current model upon newly incoming data that might contain new concepts or concept changes. The result of this adaptation is an adapted weighted mixture component that respects the original mixture.

The model adaptation method starts by training a new model M_{drift} from data drifts D_{drift} using Algorithm 1, and then combining the existing model $\mathcal{M}$. Consequently, the newly adapted model $\mathcal{M}$ accommodates the new concept represented by the components in $\mathcal{M}_{drift}$. The next step is to calculate the pairwise distance between the components in $\mathcal{M}$ using KL discrimination. The KL discrimination formula (Eq. (8)) enables us to set an upper bound on the discrimination of the mixture before and after the merging process. According to this formula, components with low weights means close to their variances, and similar covariance matrices are selected for merging. When two components are merged, the moment-preserving merging method [25] is used to preserve the mean and the covariance of the overall mixture (Eqs. (4)–(6)). Figure 4 illustrates the CMGM adaptation process.
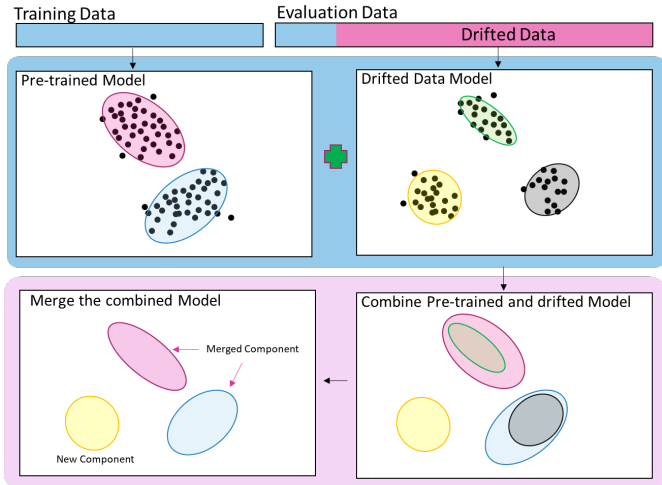
Figure 4: Illustration of the combine–merge Gaussian mixture model (CMGMM) adaptation process

As a result, the reduction process generates a set of merged models $\mathcal{M}_{merge}$. To select the best $\mathcal{M}_{merge}$ model, the accumulative BIC is computed by combining sampling data from $\mathcal{M}_{curr}$. D_{drift} then computes the BIC value using Eq. (10). The smaller the value of the accumulative BIC, the better the newly adapted model.

Based on [7] and [11], the CMGMM tends to increase the number of components because it combines and merges them. This mechanism leads to an overfitting problem because the adaptation frequency increases due to the sensitive KD3 hyperparameter.

To maintain the compactness of the CMGMM and avoid overfitting, we design a strategy to merge statistically equivalent components into one component, then prune the inactive components. The inactive components are identified by the

proximity of the ratio of w^2 and P^2 of the merged component to zero. In practice, components with w that are very close to zero are ignored by the model, whereas those with a large covariance tend to overlap with other components. Algorithm 3 presents in detail the steps of the proposed CMGMM-based method.

Algorithm 3: Model adaptation

Input: Current Model M, Drifted Dataset D_{drift}
Result: Adapted Model
M_{drift} = findBestGMM(D_{drift})
$M_{combine}$ = CombineGMMComponent(M_{drift}, M)
distanceMatrix = KLDissimalarity($M_{combine}$)
ds = D_{drift} + M.generateData()
$nComp_{min}$ = M.number_component
$nComp_{max}$ = $M_{combine}$.number_component
BIC_{best} = ∞
for targetComponent = $nComp_{min}$ **to** $nComp_{max}$
    M_{merge} = mergeComponent(target, distanceMatrix)
    if useComponentPrune **then**
        ComponentPrune(M_{merge})
    End
    $BIC_{candidate}$ = ComputeBIC (M_{merge}, ds)
    if $BIC_{candidate}$ < BIC_{best} **then**
        M_{best} = M_{merge}
    End
end
return M_{best}

---

4. Experiment

This section provides information about the datasets and experimental setup used in this study to train, optimize, and evaluate the proposed method.

4.1. Datasets

We used three types of datasets in this experiment, that is, training, optimization, and evaluation. The training dataset consisted of audio signals extracted with a 10-seconds window from 15 scenes in the TUT Acoustic Scenes 2017 [26] and TAU Urban Acoustic Scenes 2019 datasets [26]. The scenes were home, airport, beach, office, cafe, grocery store, bus, tram, metro, city center, residential area, street pedestrian, and shopping mall.

To simulate the concept drift in the datasets, the optimization and evaluation datasets were generated by overlay new additional event sounds from and UrbanSound8K datasets [27] and the BBC Sound Class Library [28]. When the sounds were added, the numbers of additions (1–10), positions in the time axis (0–9000 ms), and loudness (−20,0) of the sounds were randomly changed at random. In total, the dataset had 46 new event classes and 371 added event sounds.

We generated four concept drift types with three scenarios. The four types of concept drift are as follows:

- abrupt concept drift (AB), where ongoing concepts are replaced with new concepts at a particular time;

- gradual concept drift (GR), where new concepts are added to an ongoing concept at a particular time;

- recurring concept drift type 1 (R1), where an ongoing concept is replaced at a particular time with concepts that previously appeared; and

- recurring concept drift type 2 (R2), where concepts that previously appeared are added to an ongoing concept at a particular time.

Figure 5 illustrates the scenario generation. The three scenarios were designed to have different data distribution complexities. The scenario details are described below:

- Scenario 1 (Sc1): A unique event sounds from a specific event sounds is repeatedly introduced with a random number of times, gain, and timing. For example, in the airport scene, unique sounds representing the airplane sound, crowd background, and construction site are overlaid with a random number of times (1–10), position (0–9000 ms), and loudness (−20,0).

- Scenario 2 (Sc2): Several event sounds are randomly selected from a set of the same sound labels in Sc1 and added using the same rule as Sc1.

- Scenario3 (Sc3): This scenario differs from Sc1 and Sc2 in that event sounds coexist among scenes. For example, a set of rain sounds exists in other scenes (e.g., beach, city center, and forest paths). The methods of selection and addition are the same as those in Sc2.
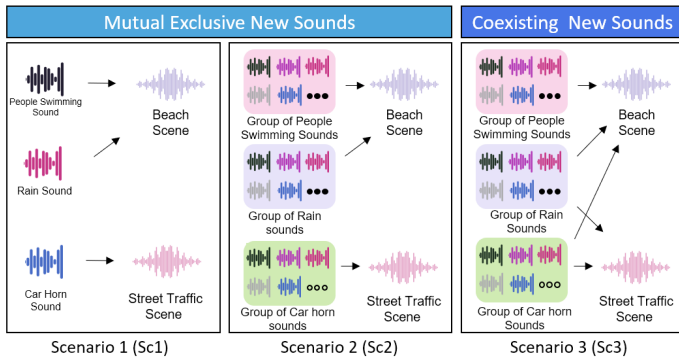
Figure 5: Concept drift scenario

Table 1 shows a list of event sounds that appear at scene types in every concept drift scenario. The mutually exclusive event sounds appear in all scenarios, but coexisting sounds only appear for Sc3.

Table 1: Setting of the novel sounds in scene audio for Sc1, Sc2, and Sc3

Scene	Mutually Exclusive Sounds in Sc1, Sc2, and Sc3	Additional Coexisting Sounds in Sc3
Airport	Helicopter, crowd, and construction site	Airplane, footsteps, and children playing
Beach	People swimming, footsteps on the sand, and rain	Teenage crowd, dog, and birds
Bus	Car horn, engine, and city car	Kitchenware, phone ringing, children playing, and teenage crowd
Café /restaurant	Washing machine, food mixer, and kitchenware	Phone ringing, children playing, and teenage crowd
City center	Sound of bird, ambulance, and wind	Footsteps, phone ringing, and children playing
Grocery store	Footsteps, children playing, and shopping cart	Vacuum cleaner, phone ringing, and footstep
Home	Frying, door, and vacuum cleaner	Clock and phone ringing
Metro station	Siren, road car, and thunder	Footsteps, crowd, and wind
Office	Typing, phone ringing, and sneeze	Broom, camera, and footsteps
Public square	People running, music, and airplane	Birds, rain, and teenagers talking
Residential area	Wind, camera, and cat	Birds, sneeze, and clock
Shopping mall	Clock, camera, and teenage crowd	Children playing, phone ringing, dog
Street pedestrian	Dog, bicycle, and bird	Footsteps and children playing
Street traffic	Motorcycle, horn, and train	Siren, airplane, and bell
Tram	Coughing, bell, and footsteps on the pavement	Teenage crowd and children playing

Finally, we have one training dataset, four optimization datasets, and 12 evaluation datasets in this experiment. Each training dataset contained 3,000 scene audio, while the optimization and evaluation dataset contained 15,000 scene audio. The datasets are available in our repository[2].

4.2. Experimental Setup

The CMGMM accuracy was evaluated under four concept drift types in three scenarios (i.e., Sc1, Sc2, and Sc3). The evaluations are performed using the two following approaches:

- Active CMGMM adaptation: In this approach, the CMGMM actively detects the concept drift using a certain method and only adapts the model when the concept drift is detected. In this study, we compared KD3 to ADWIN [9], HDDMA, HDDMW [10], and KSWIN [18].

- Passive CMGMM adaptation: In this approach, the CMGMM adapts as soon as a particular datum is received without requiring the explicit prior detection of the concept drift. Several adaptation cycle sizes were tested, that is, 25, 50, 100, 150, and 200.

5. Experiment Result

The experimental result of the proposed method are presented herein.

5.1. Hyperparameter Optimization

The first step in this experiment is the systematic optimization of the KD3 hyperparameter. We used the grid-search method using a combination of hyperparameters α from 0.1 to 0.001, β from 0.0001 to 0.000001, and ŉ from 45 to 300. We prepared a particular

dataset for the KD3 hyperparameter optimization in four types of concept drift.

In [11], we reported that hyperparameters β and $\hbar$ did not have a significant effect on accuracy. Therefore, during the initial step, we observed the performance change according to β and $\hbar$. Figure 6 shows the average accuracy in all concept drift types according to hyperparameters β and $\hbar$. In this experiment, the best β and $\hbar$ were set at 0.0001 and 45, respectively.
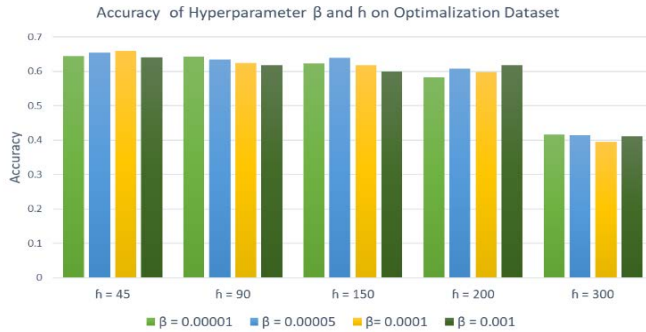
Figure 6: Result of hyperparameters β and $\hbar$ in four types of concept drift

Table 2 lists the experimental results of α in the optimization dataset in four types of concept drift. Based on this experiment, every concept drift type has its respective hyperparameter α according to the concept drift characteristics. AB and GR have similar characteristics. There are no repeating concepts in the future; hence, a more sensitive concept drift detector than R1 and R2 is required.

Table 2: KD3 hyperparameter optimization result

Concept Drift Types	Hyperparameter α (β = 0.001, $\hbar$ = 45)				
	α = 0.1	α = 0.05	α = 0.01	α = 0.005	α = 0.001
AB	0.6568	0.7113	0.7069	**0.7137**	0.7066
GR	0.6007	0.6173	**0.7050**	0.6922	0.7002
R1	**0.7332**	0.7232	0.7158	0.7054	0.6927
R2	**0.7268**	0.7133	0.7228	0.7090	0.6823
Overall	0.6793	0.6912	**0.7126**	0.7050	0.6950

In AB and GR, a sensitive hyperparameter α accelerated the update frequency. In the experimental result for these types of concept drifts, a high adaptation frequency reduced the loss received. However, a less-sensitive hyperparameter showed a better result in recurring concept drifts, where an old concept reappears in the future. A less-sensitive hyperparameter α provided the model with longer data compared to the sensitive hyperparameter.

We also selected the overall hyperparameter setting based on this experiment. The overall hyperparameter was selected from the best average performance of the hyperparameter optimization (α = 0.01, β = 0.001, and $\hbar$ = 45). We used this hyperparameter for further CMGMM and KD3 evaluations.

5.2. Active Combine–Merge Gaussian Mixture Model (CMGMM) Adaptation Result

Table 3 presents the experimental results of the active CMGMM adaptation. In general, the model performance without a concept drift detector is low in all concept drift types and

scenarios. The adaptations of the CMGMM on R1 and R2 showed better accuracy than those on AB and GR. On average, AB exhibited the lowest accuracy, while R2 showed the highest accuracy. This high accuracy on recurring was caused by the CMGMM being designed to preserve the old concept, even though the new concept is adapted in the model. Thus, the model can recognize the previously learned concept if it is repeated in the future.

The CMGMM experiment result depicted that KD3 outperformed other combinations in two of the four concept drift types in GR and R2. Meanwhile, ADWIN showed the best accuracy in AB. KSWIN demonstrated the best accuracy in R1, whereas HDDM was unsuitable for this case. Despite getting the highest overall accuracy score, the combination of the CMGMM and KD3 needed more frequent adaptations than ADWIN and KSWIN. In contrast, both HDDM-based methods showed worse performances compared to all others. HDDMA overdetected the concept drift in all concept drift types for more than 3000 times in GR.

The abovementioned results illustrated that the concept drift detector plays a vital role in the concept drift adaptation. The model performance decreased over time if the drift detector failed to detect or delay detecting or over detecting the concept drift.

Performance of the combine–merge Gaussian mixture model (CMGMM) and kernel density drift detection (KD3)

KD3 showed the best average accuracy of 0.6983 compared to ADWIN, KSWIN, and HDMM with 209 adaptations. This combination also showed its best results on R2 with 0.7321 accuracy, followed by R1 with 0.7373 accuracy, GR with 0.6999 accuracy, and AB with accuracy 0.6469. Furthermore, this combination was the most stable in all scenarios. The maximum performance decrements in AB, GR, R1, and R2 were 1.38%, 1.2%, 1.34%, and 0.94%, respectively.

Despite achieving a good performance in all concept drift types, the number of concept drifts detected in this combination was higher than ADWIN and KSWIN. The most significant number of adaptations occurred in AB. The disadvantage of a high number of adaptations is the higher computation time required to finish the task and possible overfitting. In this case, the higher numbers of adaptations in AB and GR are obtained because the concept constantly changes over time, and the learned concept becomes obsolete in the future; hence, the higher the adaptation, the better the performance.

Performance of the combine–merge Gaussian mixture model (CMGMM) and ADWIN

In general, the combination of CMGMM and ADWIN showed a good performance in every concept drift type, especially on the abrupt datasets, where this combination showed its best performance. The overall accuracy was 0.6371 with 83 times of adaptation. The overall accuracies of this combination in AB, GR, R1, and R2 were 0.6369, 0.6475, 0.6912, and 0.7169, respectively. Furthermore, this combination had the advantage of a small number of adaptations in all concept drift types. Hence, ADWIN showed an effective performance in using resources and had a reasonably good performance. This combination performed very well on Sc1, but showed a performance drop in Sc2 and Sc3. For

Table 3: Experiment result of the CMGMM with the concept drift detector

Concept Drift Detector		Accuracy				F1				Number of Concept Drift Detection			
		Sc1	Sc2	Sc3	Overall	Sc1	Sc2	Sc3	Overall	Sc1	Sc2	Sc3	Overall
AB	**ADWIN**	**0.6989**	**0.6525**	**0.6214**	**0.6369**	**0.713**	**0.6495**	**0.6353**	**0.6599**	83	89	85	85
	HDDM_A	0.4157	0.4476	0.4345	0.4326	0.4586	0.4804	0.477	0.472	3287	3350	3359	3332
	HDDM_W	0.4041	0.4455	0.4546	0.4347	0.4485	0.4902	0.5004	0.4797	489	413	409	437
	KD3*	0.6469	0.6359	0.6331	0.6386	0.6476	0.6467	0.6395	0.6446	207	236	220	221
	KSWIN	0.6611	0.6241	0.6345	0.6317	0.6722	0.6369	0.6411	0.6501	132	121	123	125
	Without Detector	0.4121	0.4054	0.4095	0.4090	0.4235	0.4125	0.4095	0.4151				
GR	ADWIN	0.6723	0.6341	0.6363	0.6475	0.6845	0.6506	0.6478	0.6609	92	81	83	85
	HDDM_A	0.4134	0.4463	0.4311	0.4302	0.4580	0.4946	0.4701	0.4742	3306	3308	3265	3293
	HDDM_W	0.4131	0.4497	0.4544	0.439	0.4524	0.487	0.4816	0.4736	0	409	401	270
	KD3*	**0.6999**	**0.6942**	**0.6879**	**0.694**	**0.7044**	**0.7004**	**0.6867**	**0.6971**	190	218	221	209
	KSWIN	0.6532	0.6241	0.618	0.6322	0.6631	0.6326	0.6204	0.6387	129	125	107	120
	Without Detector	0.3554	0.3524	0.3489	0.3522	0.3571	0.3542	0.3501	0.3538				
R1	ADWIN	0.7222	0.6948	0.6568	0.6912	0.7393	0.6981	0.6333	0.6902	78	83	87	82
	HDDM_A	0.4721	0.5204	0.4896	0.494	0.5206	0.5566	0.5392	0.5388	3154	3258	3201	3204
	HDDM_W	0.4847	0.4815	0.5068	0.491	0.5305	0.5215	0.5357	0.5292	0	662	647	436
	KD3*	0.7373	0.7334	0.7239	0.7315	0.7389	0.7341	0.7247	0.7325	208	201	204	204
	KSWIN	**0.7818**	**0.7351**	**0.7357**	**0.7508**	**0.7876**	**0.7404**	**0.7416**	**0.7565**	142	135	126	134
	Without Detector	0.4512	0.4458	0.4257	0.4409	0.4512	0.4458	0.4257	0.4409				
R2	ADWIN	0.7353	0.7066	0.7089	0.7169	0.7398	0.7131	0.7163	0.723	75	86	82	81
	HDDM_A	0.5919	0.6034	0.6259	0.607	0.6369	0.6398	0.6658	0.6475	3052	2957	3017	3008
	HDDM_W	0.5789	0.5996	0.5875	0.5886	0.6369	0.6398	0.6658	0.6475	759	604	543	635
	KD3*	**0.7321**	**0.7325**	**0.7227**	**0.7291**	**0.7352**	**0.7315**	**0.7221**	**0.7296**	195	207	205	202
	KSWIN	0.6914	0.6964	0.677	0.6882	0.6977	0.6976	0.6811	0.6921	156	143	141	146
	Without Detector	0.4254	0.4205	0.3985	0.4148	0.4279	0.4198	0.3968	0.4148				
Overall	ADWIN	**0.7072**	0.6720	0.6559	0.6731	**0.7191**	0.6778	0.6581	0.6835	82	84	84	83
	HDDM_A	0.4733	0.5044	0.4953	0.4910	0.5185	0.5428	0.5380	0.5331	3199	3218	3210	3209
	HDDM_W	0.4702	0.4941	0.5008	0.4883	0.5170	0.5346	0.5458	0.5325	312	522	500	444
	KD3*	0.7041	**0.6990**	**0.6919**	**0.6983**	0.7065	**0.7031**	**0.6932**	**0.7009**	200	215	212	209
	KSWIN	0.6969	0.6699	0.6663	0.6757	0.7051	0.6768	0.6710	0.6843	139	131	124	131
	Without Detector	0.4110	0.4060	0.3956	0.4042	0.4149	0.4080	0.3955	0.4061				

(*) Proposed method

example, the accuracies of AB, GR, R1, and R2 decreased in Sc3 by 4.64%, 3.82%, 2.74%, and 287%, respectively.

Performance of the combine–merge Gaussian mixture model (CMGMM) and Hoeffding's bounds-based method (HDDM)

In this experiment, both Hoeffding's inequality-based algorithms showed underperformance results for all concept drift types. Both algorithms were less effective in detecting the concept drift in this case. The overall accuracies of HDDMA in AB, GR, R1, and R2 were 0.4326, 0.4302, 0.494, and 0.607, respectively. The number of HDDMA adaptations exceeded 3000 times of adaptation. This high adaptation process was ineffective because the amount of trained data for each adaptation was too small. This condition led to an overfitting and decreased the model performance.

HDDM W also experienced the same problem. In some cases, HDDMA failed to detect the drift concepts, such as GR, R1, and R2 in Sc1. The overall accuracies of this HDDMW in AB, GR, R1, and R2 were 0.4347, 0.439, 0.491, and 0.5886, respectively.

Performance of the combine–merge Gaussian mixture model (CMGMM) and KSWIN

The combination of CMGMM and KSWIN showed the best performance in R1, with an overall accuracy of 0.750 with 134 adaptations. The accuracies of this combination in AB, GR, R1, and R2 were 0.6317, 0.6322, 0.7508, and 0.6882, respectively. On average, KSWIN required eight to nine adaptations per scene in all dataset types. This algorithm seems able to detect occurring changes in data and supports the concept drift handling process with good indicators at a given time.

5.3. Passive Combine–Merge Gaussian Mixture Model (CMGMM) Adaptation Result

Table 4 lists the experimental results of the passive CMGMM adaptation. The best performance in AB, GR, and R1 was obtained with a cycle size of 50, and that in R2 was obtained with a cycle size of 100. The best accuracies of AB, GR, R1, and R2 were 0.7152, 0.7139, 0.7323, and 0.7155, respectively.

Table 4: The experiment result of CMGMM without concept drift detector

Concept Drift Types	Cycle Size	Accuracy			
		Sc1	Sc2	Sc3	Average
AB	25	0.6290	0.6236	0.6256	0.6260
	50	**0.7122**	**0.7159**	**0.7177**	**0.7152**
	100	0.6285	0.5925	0.5955	0.6055
	150	0.6361	0.5776	0.5851	0.5996
	200	0.5580	0.5059	0.5119	0.5252
GR	25	0.6294	0.6232	0.6004	0.6176
	50	**0.7133**	**0.7169**	**0.7115**	**0.7139**
	100	0.6173	0.5887	0.6089	0.6049
	150	0.6371	0.5818	0.5797	0.5995
	200	0.5334	0.5094	0.5076	0.5168
R1	25	0.6186	0.5799	0.5608	0.5864
	50	**0.7235**	**0.7320**	**0.7416**	**0.7323**
	100	0.7211	0.7105	0.6988	0.7101
	150	0.7332	0.7086	0.7120	0.7179
	200	0.6639	0.7018	0.7146	0.6934
R2	25	0.5133	0.5444	0.5669	0.5415
	50	0.7396	0.6991	0.7011	0.7132
	100	**0.7431**	**0.7012**	**0.7023**	**0.7155**
	150	0.6865	0.6639	0.6803	0.6769
	200	0.6502	0.6011	0.6089	0.6200

Similar to active adaptation, R1 and R2 showed good performances compared to AB and GR, but better performances in passive adaptation. Although R1 exhibited the best adaptation at cycle size 50, it also showed good result at cycle sizes 100 and 150. If you consider the time and the computing resources used, then cycle sizes 100 and 150 are recommended.

In passive adaptation, the cycle size is vital in achieving a good performance. This cycle size determines the adequacy of the data for adaptation. If the cycle size is too short, the number of data adapted is small, leading to overfitting problems.

5.4. Suggestions and Limitations

The experiment results showed that the combination of CMGMM and KD3 has a higher number of adaptations compared to that of ADWIN and KSWIN due to the selection of KD3 hyperparameters that are sensitive to accommodating GR and AB. When the number of adaptations is increased, then in certain cases, such as GR and AB, the accuracy is improved, albeit with a higher computing cost. The advantage of KD3 compared to the other methods is that it could be applied in multi-dimensional probability distribution; hence, it is more flexible to apply in other models and cases.

In cases where the time or location of the concept drift can be predicted, the use of a passive adaptation strategy is more beneficial and has a lower computational cost than the active strategy. However, if the adaptation cycle is too far from the concept drift, then the model performance will decrease over time.

6. Conclusion

This paper presented KD3 hyperparameter optimization and the CMGMM performance evaluation in four types of concept drift. All concept drift types had optimum hyperparameter configurations. AB and GR showed similar patterns that required a smaller α value than the recurring concept drift. In this type of

concept drift, concepts continuously appear; hence, a sensitive drift detector is needed to update the model early or have a high-frequency model adaptation. However, in recurring concepts, the new concept may be repeated in the future; thus, a lower-frequency adaptation shows a better performance.

The evaluation results demonstrated that the proposed algorithms work well in detecting and adapting to four types of concept drift and three scenarios. Overall, the CMGMM works better in R1 and R2 than in AB and GR because it is designed to preserve old concepts to preserve previously learned knowledge. Component pruning is only performed on components with a minimal impact on the prediction.

In the active adaptation strategy, the proposed combined method of CMGMM and KD3 outperformed two of four other combination methods in GR and R1. These methods showed the most stable performance among Sc1, Sc2, and Sc3 in all concept drift types. Furthermore, ADWIN showed the best results on AB. This algorithm is efficient in computing resources, but less stable in more challenging scenarios, such as Sc2 and Sc3. KSWIN showed the best results in recurring drift and good performance for all concept drift types. HDDM overdetected or failed to detect the concept drift and was not suitable in this case.

In the passive adaptation strategy, a short or long adaptation cycle could reduce the performance, and a short cycle size makes the adaptation less effective. A short cycle could disrupt the model component because the model is trained with insufficient data, leading to an underfitting problem.

In AB and GR, the model requires a high-frequency adaptation to maintain the performance; therefore, sensitive hyperparameters or a short cycle size is required. Based on the experimental results, the passive method is more suitable in these concept drift types. However, the computational costs are higher than those of the active method. Furthermore, in recurring drift scenarios, a less-sensitive hyperparameter, or a moderate cycle size is needed. The active method is recommended for recurring drift types considering the computation cost and better accuracy in experiments.

As part of our future work, we plan to improve the proposed KD3 concept drift detection algorithm to achieve adaptive weight, reduce the kernel density computation, and optimize the number of data points considered when detecting the concept drift and decreasing the computation time.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] R. Elwell, R. Polikar, "Incremental Learning of Concept Drift in Nonstationary Environments," IEEE Transactions on Neural Networks, **22**(10), 1517–1531, 2011, doi:10.1109/TNN.2011.2160459.

[2] S. Ntalampiras, "Automatic analysis of audiostreams in the concept drift environment," in 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, 2016, doi:10.1109/MLSP.2016.7738905.

[3] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, A. Bouchachia, "A Survey on Concept Drift Adaptation," ACM Computing Surveys, **46**(4), 1–37, 2014, doi:10.1145/2523813.

[4] T.R. Hoens, R. Polikar, N. V. Chawla, "Learning from Streaming Data with Concept Drift and Imbalance: An Overview," Progress in Artificial Intelligence, **1**(1), 89–101, 2012, doi:10.1007/s13748-011-0008-0.

[5] I. Žliobaitė, "Learning under Concept Drift: an Overview," 1–36, 2010, doi:10.1002/sam.

[6] C. Chen, C. Wang, J. Hou, M. Qi, J. Dai, Y. Zhang, P. Zhang, "Improving Accuracy of Evolving GMM under GPGPU-Friendly Block-Evolutionary Pattern," International Journal of Pattern Recognition and Artificial Intelligence, **34**(3), 1–34, 2020, doi:10.1142/S0218001420500068.

[7] I.D. Id, M. Abe, S. Hara, "Concept Drift Adaptation for Acoustic Scene Classifier based on Gaussian Mixture Model," in IEEE Region 10 Annual International Conference, Proceedings/TENCON, IEEE, Osaka: 450–455, 2020, doi:10.1109/TENCON50793.2020.9293766.

[8] J.M. Acevedo-Valle, K. Trejo, C. Angulo, "Multivariate regression with incremental learning of Gaussian mixture models," Frontiers in Artificial Intelligence and Applications, **300**, 196–205, 2017, doi:10.3233/978-1-61499-806-8-196.

[9] A. Bifet, R. Gavaldà, "Learning from Time-changing Data with Adaptive Windowing," Proceedings of the 7th SIAM International Conference on Data Mining, (April), 443–448, 2007, doi:10.1137/1.9781611972771.42.

[10] I. Frías-Blanco, J. Del Campo-Ávila, G. Ramos-Jiménez, R. Morales-Bueno, A. Ortiz-Díaz, Y. Caballero-Mota, "Online and Non-Parametric Drift Detection Methods Based on Hoeffding's Bounds," IEEE Transactions on Knowledge and Data Engineering, **27**(3), 810–823, 2015, doi:10.1109/TKDE.2014.2345382.

[11] I.D. Id, M. Abe, S. Hara, "Evaluation of Concept Drift Adaptation for Acoustic Scene Classifier Based on Kernel Density Drift Detection and Combine Merge Gaussian Mixture Model," in 2021 Spring meeting of the Acoustical Society of Japan, 2021.

[12] J. Montiel, J. Read, A. Bifet, T. Abdessalem, "Scikit-multiflow: A Multi-output Streaming Framework," Journal of Machine Learning Research, **19**, 1–5, 2018, doi:10.5555/3291125.3309634.

[13] G.I. Webb, R. Hyde, H. Cao, H.L. Nguyen, F. Petitjean, "Characterizing Concept Drift," Data Mining and Knowledge Discovery, **30**(4), 964–994, 2016, doi:10.1007/s10618-015-0448-4.

[14] T. Zhang, J. Liang, B. Ding, "Acoustic scene classification using deep CNN with fine-resolution feature," Expert Systems with Applications, **143**(1), 113067, 2020, doi:10.1016/j.eswa.2019.113067.

[15] A. Tsymbal, "The Problem of Concept Drift: Definitions and Related Work," 2004.

[16] E.S. Page, "Continuous Inspection Schemes," Biometrika, **41**(1/2), 100, 1954, doi:10.2307/2333009.

[17] Y. Yuan, Z. Wang, W. Wang, "Unsupervised concept drift detection based on multi-scale slide windows," Ad Hoc Networks, **111**, 102325, 2021, doi:10.1016/j.adhoc.2020.102325.

[18] C. Raab, M. Heusinger, F.M. Schleif, "Reactive Soft Prototype Computing for Concept Drift Streams," Neurocomputing, 2020, doi:10.1016/j.neucom.2019.11.111.

[19] A.R. Runnalls, "Kullback-Leibler Approach to Gaussian Mixture Reduction," IEEE Transactions on Aerospace and Electronic Systems, **43**(3), 989–999, 2007, doi:10.1109/TAES.2007.4383588.

[20] A.P. Dempster, N.M. Laird, D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," Journal of the Royal Statistical Society. Series B (Methodological), **39**(1), 1–38, 1977.

[21] A.D.R. McQuarrie, C.-L. Tsai, Regression and Time Series Model Selection, WORLD SCIENTIFIC, 1998, doi:10.1142/3573.

[22] P. Pal Singh, "An Approach to Extract Feature using MFCC," IOSR Journal of Engineering, **4**(8), 21–25, 2014, doi:10.9790/3021-04812125.

[23] C. Fraley, "How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis," The Computer Journal, **41**(8), 578–588, 1998, doi:10.1093/comjnl/41.8.578.

[24] E. Parzen, "On Estimation of a Probability Density Function and Mode," The Annals of Mathematical Statistics, **33**(3), 1065–1076, 1962, doi:10.1214/aoms/1177704472.

[25] J.L. Williams, P.S. Maybeck, "Cost-function-based Gaussian mixture reduction for target tracking," Proceedings of the 6th International Conference on Information Fusion, **2**, 1047–1054, 2003, doi:10.1109/ICIF.2003.177354.

[26] A. Mesaros, T. Heittola, T. Virtanen, TUT Acoustic scenes 2017, Development dataset, 2017, doi:10.5281/ZENODO.400515.

[27] J. Salamon, C. Jacoby, J.P. Bello, "A Dataset and Taxonomy for Urban Sound Research," MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia, (November), 1041–1044, 2014, doi:10.1145/2647868.2655045.

[28] BBC, BBC Sound Effects, https://sound-effects.bbcrewind.co.uk/search, Mar. 2019.

Development of Miniaturized Monolithic Isolated Gate Driver

Hsuan-Yu Kuo, Jau-Jr Lin[*]

Department of Electrical Engineering, National Changhua University of Education, Changhua, 500 Taiwan, R.O.C.

A R T I C L E   I N F O

A B S T R A C T

Gate driver has been applied in many ways, exemplified by that, by using the DC-isolated and AC-pass characteristics of gate driver's primary and secondary sides, the problem of floating endpoint in semiconductor power switch can be solved. However, the conventional design of isolated gate driver provides circuit voltage blocking by optically coupled components. Due to the need for optoelectronic conversion, it requires III-VI semiconductor process and non-standard CMOS process, and the cost is always high. Therefore, in order to better solve the above mentioned problem, an electronic isolated gate driver is proposed. It employs an on-chip transformer to provide voltage isolation between the primary and secondary sides of the circuit, and converts the control signal in the circuit into a high-frequency modulated signal, which in the secondary side is then demodulated by the rectifier through the on-chip transformer to produce the original control signal. The miniaturized isolated gate driver proposed herein adopts TSMC T25HVG2 process and uses an on-chip transformer design in lieu of an optically coupled components. As the amplitude shift keying, on-chip transformer, full-wave quadruple rectifier and data buffer amplifier involved in this design are all integrated on the same chip, the integration can be improved. The size can be smaller than the generally separating electronic isolated gate driver, with the interference from external noise being reduced. In addition, the proposed gate driver generates large signals, in terms of chip layout, therefore, the circuit is put inside the on-chip transformer, which can further save area.

1. Introduction

Conventional galvanic isolated gate drivers use optically coupled components for a high degree of voltage isolation. Due to the need for optoelectronic conversion, it requires III-VI semiconductor process and non-standard CMOS process, and the cost is always high. In desirous of solving these problems, MOSFET electronic isolated gate driver has been studied as a new direction. It couples the energy to the secondary side of the transformer by the transformer's resonant coupling method in lieu of the optical coupling driver. Since the isolation is realized by optical and electrical energy conversion, many of the shortcomings in the traditional design can be improved, such as the larger size and higher process cost in the past.

This paper is an extension of work originally presented in the 2nd IEEE Eurasia Conference on IOT, Communication and Engineering 2020[1]. The proposed system concept in which is shown in Figure 1. The circuit is mainly composed of a VCO, a mixing circuit (switch), an on-chip transformer, a quadruple

rectifier and a data buffer amplifier. We input a PWM-controlled signal to the chip, and then the ~1 GHz oscillation signal and the PWM signal are modulated by ASK (amplitude shift keying) at the switch, and then the modulated signal is brought to the secondary side by the on-chip transformer; the quadruple rectifier then further amplifies the small signal and rectify the DC voltage of the process to provide enough driving voltage for the data buffer amplifier to generate the PWM signal, and that's the expected result.
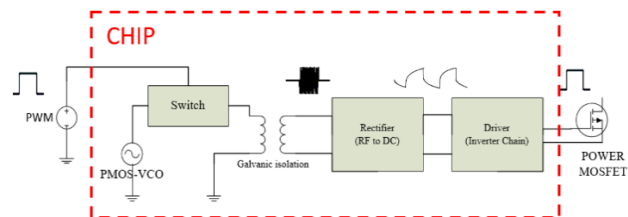
Figure 1: System concept of proposed isolated gate driver.

The circuit proposed is integrated on a single chip, so the size is much smaller than the separated design [2, 3]; besides, according

to the simulation, if a PWM control signal of 100- kHz 50% duty circle is given at the primary side's input end, we hope that after the voltage isolation, the output can remain the same. However, on account of the transmission delay caused by the rise time and fall time, the output PWM control signal will be slightly degraded as the simulation shows that the duty circle of the output PWM control signal is 48.4%, a difference of 1.6%; the difference between the rise time and fall time is only 0.2 μs. In the next section, we introduce the architecture of the chip, and discuss the detailed circuit design and the model of the on-load transformer in detail in the implementation section. Finally, we present the results of different process variations under Post-Simulation in Full-chip Simulation, all of which are in line with the expected specifications.

2. Chip architecture

The system block diagram and schematic diagram proposed in this article are shown in Figure 2 and Figure 3, respectively. Its primary side is composed of a current mirror with a current controller and a PMOS VCO. The input PWM signal and the 1GHz signal provided by the oscillator, after modulation, are transmitted to the secondary side through the on-chip transformer. The full-wave quadruple rectifier can further amplify and demodulate the modulated signal into a DC voltage. However, the potential and ripple of the rectified DC voltage is not suitable for being directly used as a gate drive signal, and the data buffer amplifier can improve this problem by further pulling the voltage to 5 V and greatly improving the ripple, so as to increase its driving force.
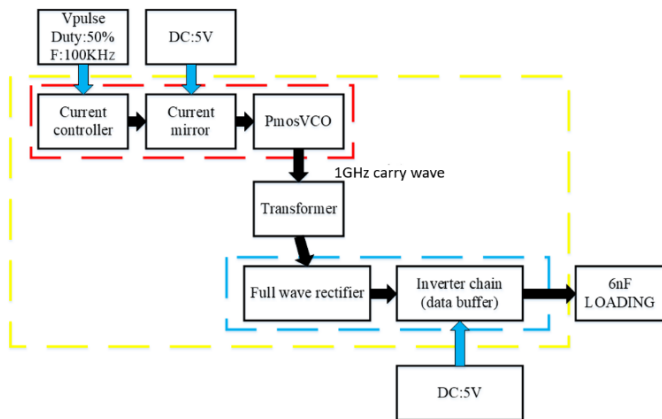
Figure 2: The system block diagram of the proposed design.

On and off of the switch can be controlled by putting a PWM control signal on the primary side. Once the switch is on, the current will increase. At this point, the large current causes the VCO to generate a larger amplitude oscillation signal. That is, as the current becomes larger or smaller, the PMOS VCO generates two signals with different amplitudes. The above phenomenon can be interpreted as that, the carrier wave generated by the PMOS VCO and the PWM signal are subjected to an ASK modulation action, and then the modulation signal is transmitted to the secondary side through the on-chip transformer. However, since this signal voltage level is not high enough to drive the data buffer amplifier, the full-wave rectifier (voltage multiplier) will increase the voltage level and rectify the RF signal to a DC voltage. Finally, the data buffer amplifier further enhances the driving capability. In

addition, the data buffer amplifier converts the ASK signal into a PWM gate signal to drive the subsequent semiconductor power switch.
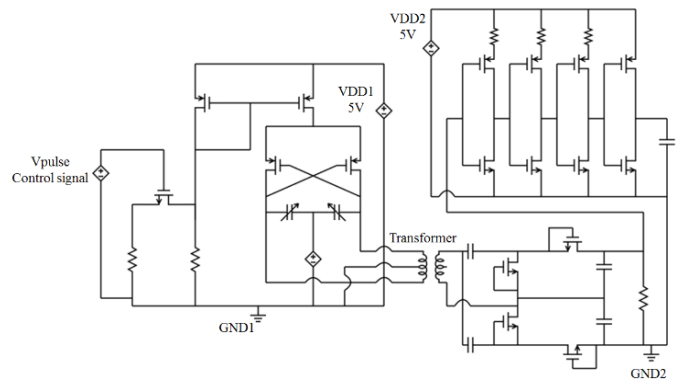
Figure 3: The schematic diagram of the proposed isolated gate driver.

There are two types of transformers for on-chip transformer design: stack transformer or tapped transformer [1, 4]. It provides galvanic isolation in the circuit proposed in this article. The stack transformer performs better in coupling, but its isolation is limited by process as the distance between the metal layers is fixed. In other words, although the coupling degree of the tapped transformer is less than that of the stack transformer, the distance between the two inductors is longer than that of the stack transformer. The transformer is more flexible. The transformer applied in this paper is a stack transformer which features higher coupling as mentioned above.
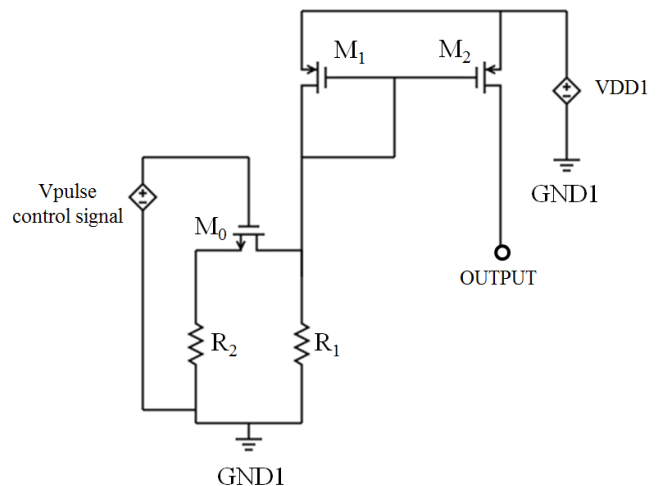
Figure 4: The schematic of current mirror and current controller

3. Circuit implementation

3.1. Current Mirror with Current Controller

The current controller, shown in Figure 4, consists of two resistors and a switch. The switch is controlled by a PWM control signal generated by a signal generator outside the chip. When the switch is on, the two resistors become parallel, which makes the impedance smaller and current higher (Figure 5). The lower part of Figure 5 is the input current of Figure 4 current mirror, and the upper part of Figure 5 is the output current after magnification of 3 times by the Figure 4 current mirror. This design serves the

purpose of performing ASK modulation of the PWM signal and the oscillation signal. Another reason for this design is that if the OOK modulation is achieved by simply turning off the oscillator instead of generating oscillation signals of different amplitudes, it takes more time for the system to enter the steady-state mode, which causes even longer transmission delay. Therefore, in order to achieve ASK modulation, the oscillation signal needs to be decreased so that it cannot drive the data buffer amplifier instead of directly closing the entire system.
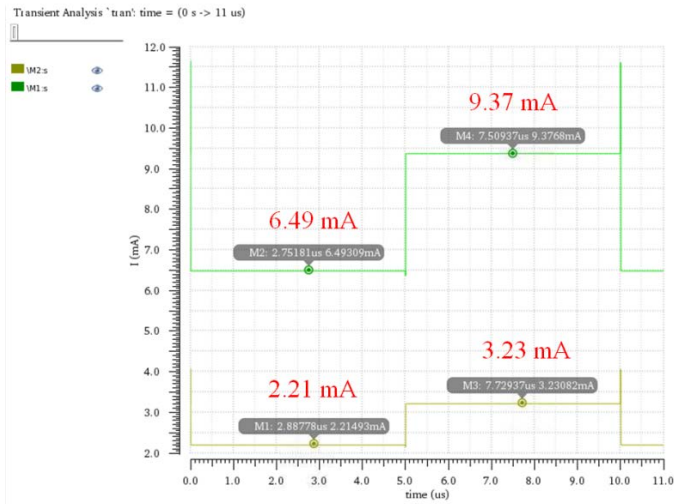
Figure 5: The output current of the current controller and the output current after going through the current mirror.

A current mirror is employed to increase the current, and the result is shown in Figure 5. The lower curve is the current value of current controller side and the upper curve is the current value that follows into VCO. When the voltage control signal is low, the current values of the current control and VCO are 2.21 mA and 6.49 mA, respectively. When the voltage control signal is high, the current values of the current control and VCO are 3.23 mA and 9.73 mA, respectively. The current mirror magnifies the current to a set magnification, such as (1). Since $L_1 = L_2$ and $W_2 = 3W_1$ in this design, the magnification should be 3 times the current in Figure 5, and the result is obtained. M is Magnification of current mirror.

$$M = (W_2 \cdot L_1) / (W_1 \cdot L_2) \qquad (1)$$

3.2. PMOS Voltage Control Oscillator

The main function of the VCO, shown in Figure 6, is to change the capacitance value by adjusting V_{BIAS} [4-6], thereby generating different oscillation frequencies. As this LC resonant circuit is composed of an inductance with fixed inductance value and a variable capacitance, this study can therefore increase the exhaustion region in the diode by enhancing the reverse bias voltage of the VCD. Therefore, when the reverse bias voltage becomes larger, the distance between the two metals in the VCD increases, the capacitance value decreases, and the resonance frequency of the LC circuit reduces. In addition, the PMOS VCO design proposed in this paper must achieve the ASK modulation effect with a current mirror with a current controller, that is, the PWM signal and the oscillation signal generated by the PMOS VCO need to be modulated.
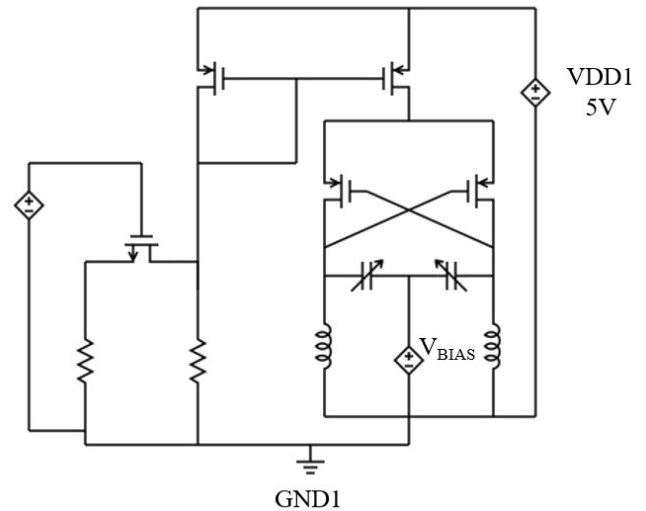
Figure 6: The circuit schematic of the PMOS VCO.
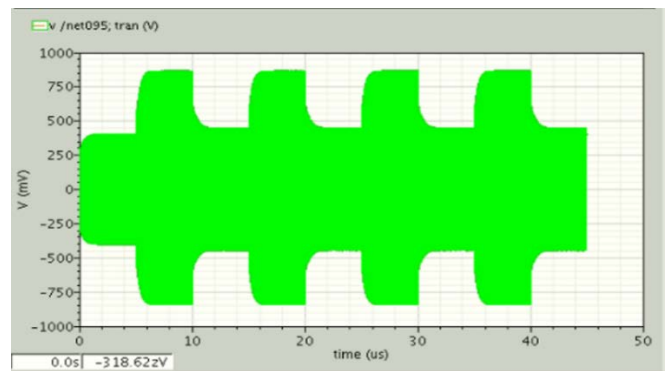
Figure 7: The output signal of the PMOS VCO with the current control voltage low and high.

As shown in Figure 7, the current controller and current mirror make the VCO generate low and high currents, which further makes the PMOS VCO generate two RF signals with different amplitudes. Generally, it takes longer to reach a stable state after the VCO oscillation stops. Therefore, it needs to make the small oscillation signal keep oscillating, and the signal must be small enough so as not to affect the components in the design proposed herein.

3.3. Stack transformer

The stack transformer [1, 4, 7], shown in Figure 8, is chosen because of its higher theoretical coupling, which means that, compared with the tapped transformer [8, 9], the stack transformer can transmit signal to the secondary side more effectively. However, the stack transformer still has shortcomings. The distance between the two inductors is limited by the process, that is to say, the distance between each layer is fixed, which means that the isolation is limited. A tapped transformer, however, uses a same layer (metal in the top), so the distance between the metals is more flexible despite the lower degree of coupling. The results are shown in Figure 9 and 10, indicating that the inductance value is about 3.57 nH, the quality factor is about 6.94, and the coupling coefficient is about 0.917 at 1 GHz.
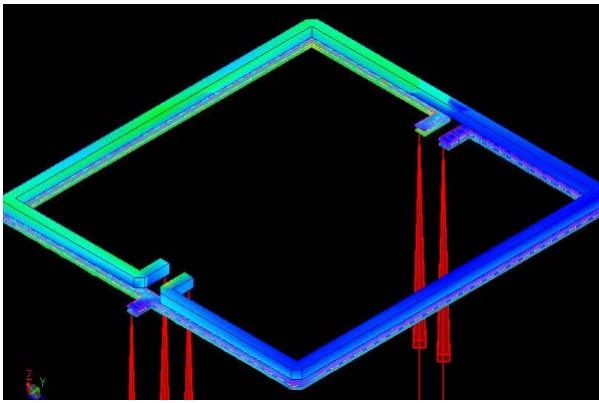
Figure 8: 3D graph of the stack transformer.

Figure 10: Coupling coefficient of the proposed transformer.
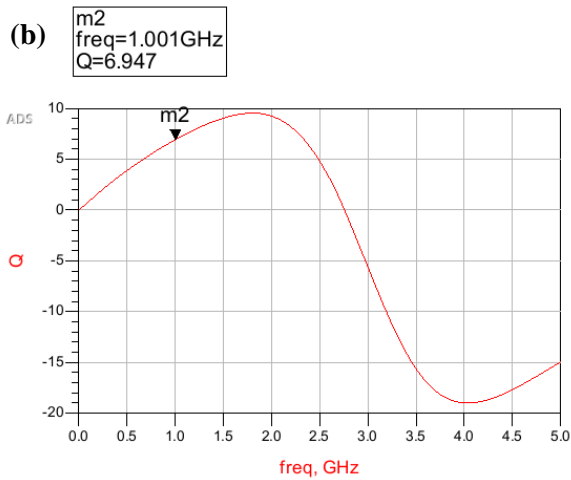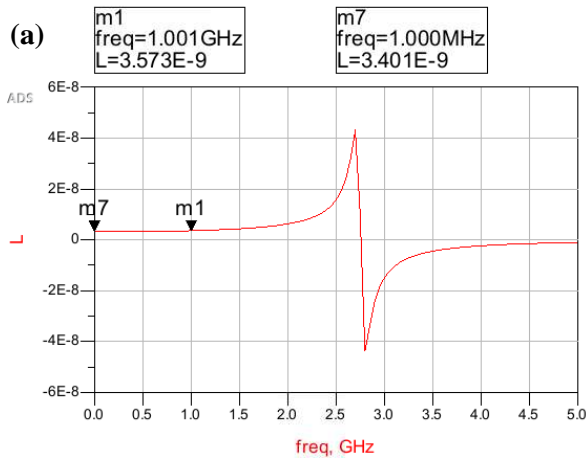
Figure 9: (a) The inductance of transformer and (b) Quality factor in the primary side.
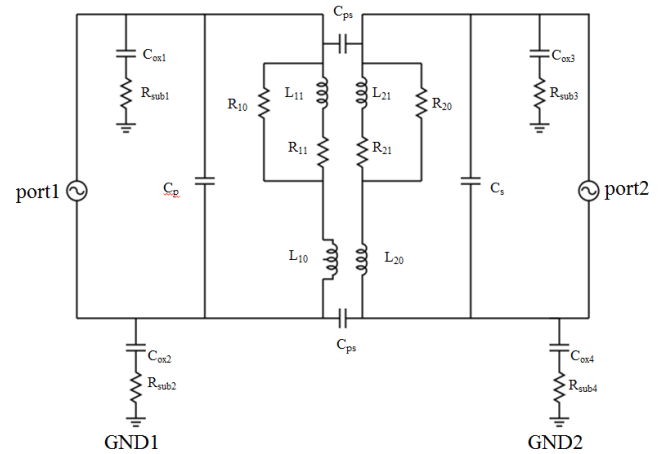
Figure 11: Modeling of the on-chip transformer.

Figure 11 shows the modeling of the on-chip transformer proposed in this paper [10]. R_{10} and R_{20} are the impedance of the transformer itself at low frequency. The values of L_{10} and L_{20} in Figure 11 are from the model of the ADS momentum transformer. L_{11}, L_{21}, R_{11}, R_{21} are the inductance and impedance generated by the proximity effect and skin effect [11,12] when the two on-chip inductors are at high frequencies. C_{ox} and R_{sub} are the parasitic resistance and capacitance between the metal layer and the substrate, and C_{ps} is the coupling capacitance between the two inductors.

In the circuit, L_{10} and L_{20} can be obtained by (2), and the result is shown in Figure 9. L_{10} and L_{20} are the inductance values in the circuit at low frequency, so we can get this data in Figure 9 (a). R_{11} and R_{21} are also the impedance of the circuit at low frequency, so its parameters can be obtained in smith and z_parameter by the same way. C_p and C_s are the parasitic capacitances in the inductance, which can be reversely obtained according to its natural vibration frequency, such as (3) (4). The natural vibration frequencies of the primary side and the secondary side are represented by $freq_p$ and $freq_s$ respectively [13].

$$L_{10} = imag(Z_{11}) / (2 \cdot pi \cdot freq) \qquad (2)$$

$$freq_p = 1 / [2 \cdot pi \cdot sqrt(L_{10} \cdot C_p)] \qquad (3)$$

$$freq_s = 1 / [2 \cdot pi \cdot sqrt(L_{10} \cdot C_p)] \qquad (4)$$

As the chip itself is a 1:1 symmetrical architecture, $R_{sub1} = R_{sub2} = R_{sub3} = R_{sub4}$, and $C_{ox1} = C_{ox2} = C_{ox3} = C_{ox4}$. we can see the comparison between the simulation and the optimized modeling from Figure12. In Figure 12, curves of S(1,1), S(2,2) and S(1,2) are from ADS momentum simulation in Figure 8, and curves of modeling2..S(1,1), modeling2..S(2,2) and modeling2..S(1,2) are

from optimized modeling in Figure 11. Coupling coefficient K and quality factor Q can be obtained from (5) (6). The difference between the results of ADS momentum simulation and optimized modeling simulation is shown in Table 1.

$$K = sqrt(imag(Z_{12}) \cdot imag(Z_{21})) / sqrt(imag(Z_{11}) \cdot imag(Z_{22})) \quad (5)$$

$$Q = imag(Z_{11}) / real(Z_{11}) \quad (6)$$

(a)

freq (1.000MHz to 5.000GHz)

(b)

freq (1.000MHz to 5.000GHz)
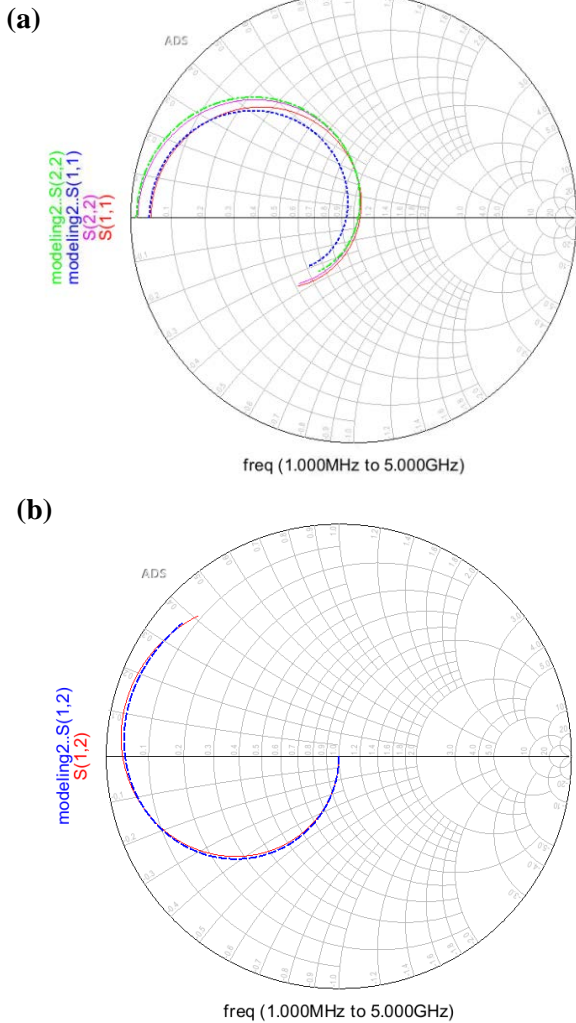
Figure 12: Comparison of the results between ADS momentum (Figure 8) and modeling (Figure 11) (a) S_{11} and S_{22} (b) S_{12} results

3.4. Full Wave Rectifier (Voltage Multiplier)

Since the RF signal received from the transformer output is low, the full-wave rectifier rectifies the RF signal to a DC voltage and multiplies the signal, making it has a greater ability to drive the data buffer amplifier. Figure 13 is the circuit schematic of the full-wave rectifier [14] proposed in this paper.

The traditional voltage multiplier is a diode voltage doubler in cascaded capacitor, where the voltage doubler uses voltage multipliers to connect N stages. If the threshold voltage is regarded as zero, it can theoretically reach N times of magnification. However, Zener diode threshold voltage and low threshold NMOS cannot be regarded as zero. In the formula (7), V_{IN} is the amplitude

of the RF oscillation signal, V_T is the threshold voltage, and N is the series of voltage multiplier in cascaded capacitor.

$$V_{OUT} = 2N \cdot (V_{IN} - V_T) \quad (7)$$

However, conventional diode multipliers come with the disadvantages of high output impedance, large DC output ripple, and low conversion efficiency. Therefore, in this study, the proposed design replaces the Zener diode with a low threshold voltage NMOS (Figure 13).

Table 1: Comparison of s parameter simulation and modeling simulation

	ADS momentum	Optimized Modeling
R_{20}	0.78 ohm	0.65 ohm
L_{20}	3.16 nH	3.51 nH
R_{10}	2.36 ohm	2.17 ohm
L_{10}	3.40 nH	3.33 nH
$freq_p$	2.75 GHz	3.05 GHz
$freq_s$	2.75 GHz	3.05 GHz
K	0.91	0.93
Q	6.94	8.43
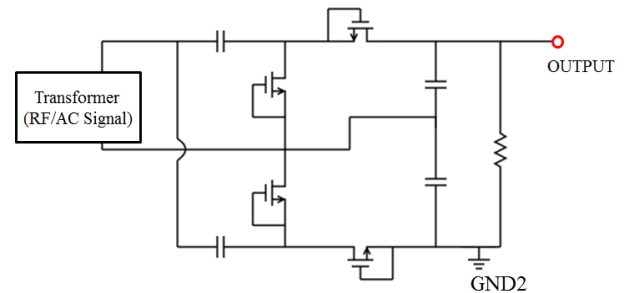
Figure 13: Proposed full-wave rectifier.

3.5. Data buffer

The data buffer amplifier consists of four inverters connected in series. The basis of the data buffer is to effectively enhance the driving capability of the gate signals [15]. In addition, it also reduces the transmission delay of the signal in inverters. Figure 14 shows the relationship between the delay time and the different stages of the inverter chain. N is the design time, f is the ratio of C_{in} to C_{out}, and tp is the transmission delay of signal through the data buffer amplifier. As N increases, tp can be reduced. This design serves the purpose of effectively reducing the ratio of electrical output capacitance to gate capacitance in each stage, and of gradually increasing the current size of each stage, so that it can shorten the rise and fall times by connecting a multi-stage inverter and thus reduce the transmission delay, which is better than the original small capacitor charging a large one. However, this stage connection has its limits as too many stages make the output of the signal generate excess delay, and the size is not easy to control. The advantage of the four-stage architecture adopted in this paper is that the phase of the output control signal is the same as that of

the primary side's PWM control signal, and the transmission delay can be minimized to avoid too many long stage connections.

Since the capacitance ratio of each stage in the four-stage architecture is three times larger than that of the previous stage, the gate capacitance (C_{in}) can be changed by application of CMOS technology, the width and length of MOSFET. In other words, since the inverter chain proposed in this paper consists of four inverters, the width of the next MOSFET stage needs to be modified so that it is three times larger than that of the previous MOSFET.

	N	f	tp
	1	64	65
	2	8	18
	3	4	15
	4	2.8	15.3
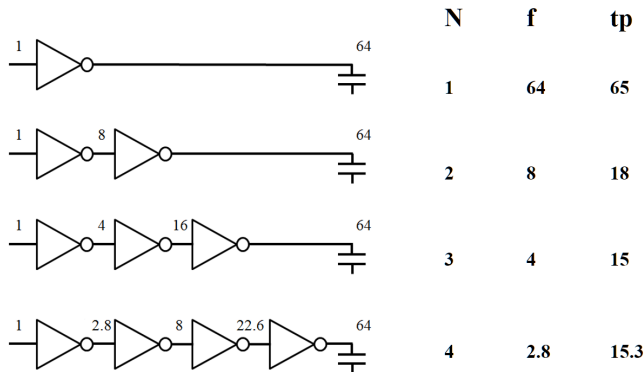
Figure 14: Signal delay time with different stage inverter.
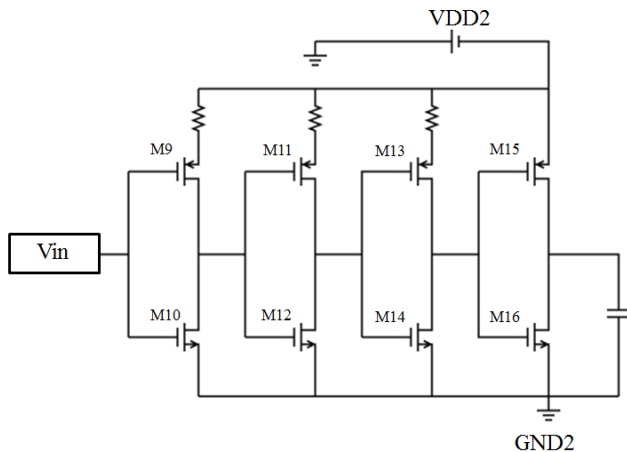
Figure 15: Proposed data buffer.

In the design, DC voltage cannot be used directly as a gate control signal for gate driver because of its large ripple and long rise and fall times. Therefore, the data buffer amplifier is an important element to improve the drawback by enhancing the drive capability. In the proposed design (Figure 15), the data buffer amplifier consists of four inverters and three resistors. Since the rectified DC voltage is not sufficient to drive the data buffer, the resistors are used to lower the threshold of the driving voltage. The four-stage data buffer can also amplify the input current by 64 magnitudes, which allows the circuit to generate enough current to drive the potential larger power MOSFET. In addition, the inverter chain eliminates ripples and also reduces fall and rise times.

4. All chip simulation

This monolithic isolated gate driver proposed in this paper is designed and manufactured with TSMC 250nm HV COMS technology and 5 V transistors. The on-chip transformer, with area 1.000 mm × 1.125 mm, provides galvanic isolation between the primary and secondary sides. The control signal sets a 100-kHz 50% duty cycle 5-V PWM signal for the current controller to switch the current value. The VCO generates 1-GHz oscillation signals of different amplitudes due to the different magnitudes of current from the current mirror and the current controller. The results are shown in Figure 16.
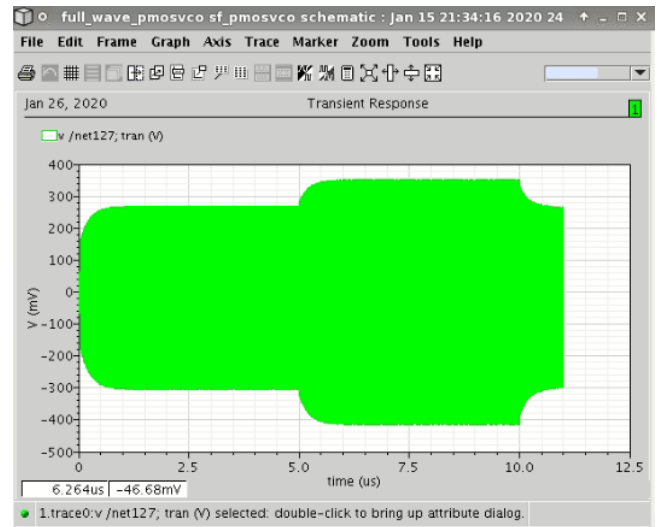
Figure 16: The oscillation signal of the PMOS VCO.
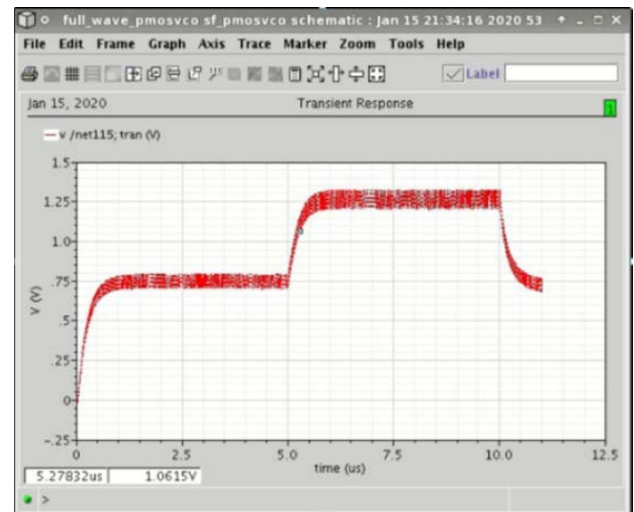
Figure 17: The output voltage of the full wave rectifier.

The signal generated by VCO is not high enough to drive the data buffer amplifier, so the full-wave rectifier amplifies the voltage and rectifies the RF signal to DC voltage, as shown in Figure 17. The driving threshold voltage of the data buffer amplifier is shown in Figure 18, and the threshold voltage of the four-stage chain is around 1V. When the driving voltage is higher than 1V, the rectified signal can trig the inverter chain. When the driving voltage is lower than 1V, it cannot trig. The results are shown in Figure 18(a). Figure 18(b) shows the result when

resistors are not added to the data buffer amplifier to reduce its threshold drive voltage, its original voltage is about 2.5 V. The threshold drive voltage that requires more voltage multipliers to be larger leads to extra energy loss and chip size increase due to the too many components.

In the final design, after the signal passes through the whole system, the duty cycle of the PWM signal with 50% duty cycle is kept at 48.4% and the rise and fall times are kept below 10% of the cycle (rise and fall times are about 0.2 µs, and the cycle time at 100 kHz is 10 µs). The results are shown in Figure 19.

Table 2: In case of different corners, the expected result and the simulation result of changing the control DC bias of the variable vessel

	Specification	Post_sim (ss)	Post_sim (tt)	Psot_sim (ff)
VCO output signal (off)	0.2 ~ 0.8	0.24 ~ 0.73	0.35 ~ 0.74	0.43 ~ 0.89
VCO output signal (on)	> 0.9	0.69 ~ 1.19	0.72 ~ 1.21	0.77 ~ 1.26
Multiplier output (off)	< 1 V	0 ~ 0.65	0.05 ~ 0.85	0.12 ~ 1.03
Multiplier output (on)	> 1V	0.53 ~ 1.91	0.78 ~ 2.12	0.85 ~ 2.43
Threshold voltage for data buffer	1 V	0.95 V	1.01 V	1.09 V
Rise time	< 1 µs	0.242 µs	0.206 µs	0.166 µs
Fall time	< 1 µs	0.239 µs	0.227 µs	0.200 µs

(a)

(b)

Figure 18: (a) The threshold of the data buffer is ~1V. (b) the threshold voltage in data buffer without resister is ~2.5 V

Primary side

secondary side

Figure 20: Chip layout

Figure 19: The output of the data buffer.
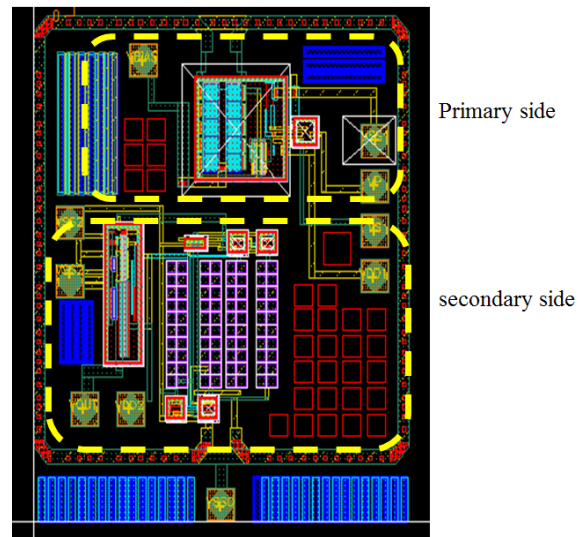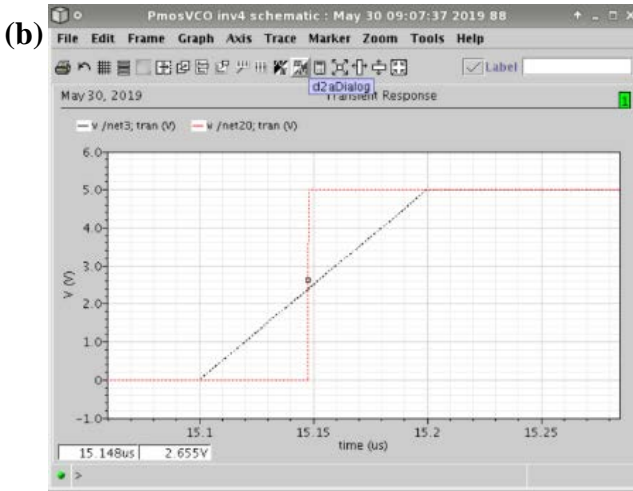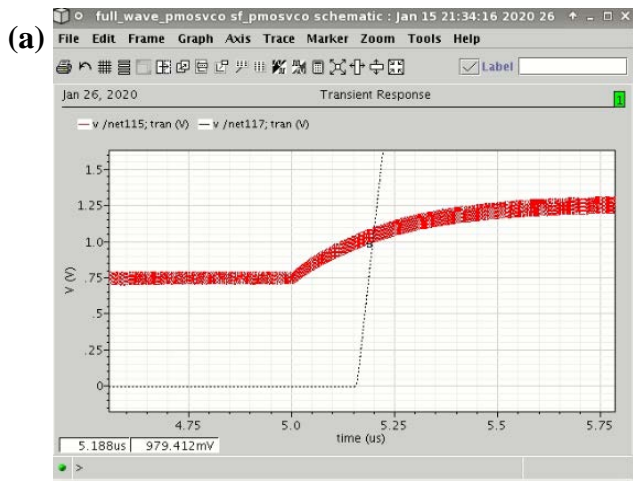
Table 2 shows the specifications and simulation results of the circuit proposed in this paper. The output voltages of the VCO and multiplier vary depending on the control voltage given to the varactor diode. It needs attention in the circuit, because in design, VCO must generate two different voltages, and after the two voltages are rectified by the voltage quadrupler, one of the rectified DC voltages needs to be able to drive the data buffer amplifier, while the other rectified DC voltage cannot drive the data buffer amplifier, so as to ensure that these two voltages can control the first-stage inverter of the data buffer amplifier, making the circuit generate PWM control signals. Assuming that this condition cannot be guaranteed to be established, it may cause the data buffer amplifier to become routinely open or closed, which leads to the failure to achieve our expected results. In addition, the varactor also makes it difficult for the VCO to start up as shown in Table 2. As to the output signal of VCO and the voltage

amplifier, the result also increases as the control DC bias of the varactor increases, which means that if insufficient or too much control DC bias is given, it may cause the first-stage structure of the inverter chain to be routinely open and closed, thus making the circuit lose its intended function. Figure 20 shows the layout of the chip to be delisted. This gate driver generates large signals, so the layout of this chip puts the circuit inside the on-chip transformer to save more area.

5. Conclusion

Traditional isolated gate drivers generally adopt optically coupled driver components to achieve high isolation, so they require III-VI semiconductor process and non-standard CMOS process, making the cost even higher. In order to solve this problem, electronic isolated gate drivers are gradually attracting more attention. On-chip transformers can reduce the energy loss caused by the optical coupling components, but a separate architecture is still used in many designs – devices with different functions are designed on different chips, which leads to other problems, such as larger size, and poor treatment of external noise. Therefore, the problem of high cost remains unsolved in such designs. The architecture proposed in this paper is to integrate the whole oscillator, transformer, full-wave multiplier and rectified DC signal processing system on one chip, so as to reach optimization in both integration and size.

On the primary side, the signal generator provides the chip with a 100 kHz PWM switching signal with a 50% duty cycle, which is then modulated by a ~1 GHz carrier wave generated by a VCO. After passing through a full-wave rectifier and a data buffer amplifier, it retains a 48.4% duty cycle and a 5-V output signal. All devices are integrated in a single chip. This will reduce the system dimension and enhance the system integration significantly.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] H.Y. Kuo, J.J.Lin, "Implementation of Miniaturized Monolithic Isolated Gate Driver," in 2nd IEEE Eurasia Conference on IOT, Communication and Engineering 2020, ECICE 2020, 129–132, 2020, doi:10.1109/ECICE50847.2020.9301940.

[2] S. Nagai, Y.Yamada, N.Negoro, H.Handa, M.Hiraiwa, N.Otsuka, D.Ueda, "A 3-phase AC-AC matrix converter GaN chipset with drive-by-microwave technology," IEEE Journal of the Electron Devices Society, **3**(1), 7–14, 2015, doi:10.1109/JEDS.2014.2364307.

[3] C. Liu, D.Mandal, Z.Yao, M.Sun, J.Todsen, B.Johnson, S.Kiaei, B.Bakkaloglu, "A 50-V Isolation, 100-MHz, 50-mW Single-Chip Junction Isolated DC-DC Converter with Self-Tuned Maximum Power Transfer Frequency," IEEE Transactions on Circuits and Systems II: Express Briefs,

66(6), 1003–1007, 2019, doi:10.1109/TCSII.2018.2869074.

[4] J. Lin, K.C.Lin, "A monolithic isolated gate driver with on-chip transformer and a Voltage Level Shifter," in 2017 IEEE 3rd International Future Energy Electronics Conference and ECCE Asia, IFEEC - ECCE Asia 2017, 1256–1259, 2017, doi:10.1109/IFEEC.2017.7992223.

[5] M. DaTsai, Y.H.Cho, H.Wang, "A 5-GHz low phase noise differential colpitts CMOS VCO," IEEE Microwave and Wireless Components Letters, **15**(5), 327–329, 2005, doi:10.1109/LMWC.2005.847696.

[6] X. Wang, B.Bakkaloglu, "Systematic design of supply regulated LC-tank voltage-controlled oscillators," IEEE Transactions on Circuits and Systems I: Regular Papers, **55**(7), 1834–1844, 2008, doi:10.1109/TCSI.2008.918004.

[7] A. Ghadiri, K.Moez, "Bandwidth enhancement of on-chip transformers using negative capacitance," IEEE Transactions on Circuits and Systems II: Express Briefs, **59**(10), 648–652, 2012, doi:10.1109/TCSII.2012.2213360.

[8] H.M. Hsu, C.W.Tseng, K.Y.Chan, "Characterization of on-chip transformer using microwave technique," IEEE Transactions on Electron Devices, **55**(3), 833–837, 2008, doi:10.1109/TED.2007.914841.

[9] R. Wu, N.Liao, X.Fang, S.Member, "A Silicon-Embedded Transformer for High-Efficiency, High-Isolation, and Low-Frequency On-Chip Power Transfer," **62**(1), 220–223, 2015. doi:10.1109/TC.2015.2517635.

[10] Q. El-Gharniti, E.Kerhervé, J.B.Bégueret, "Modeling and characterization of on-chip transformers for silicon RFIC," IEEE Transactions on Microwave Theory and Techniques, **55**(4), 607–615, 2007, doi:10.1109/TMTT.2007.893647.

[11] J. Kang, L.Sun, J.Wen, M.Zhao, "An equivalent lumped-circuit model for on-chip symmetric intertwined transformer," in ASICON 2009 - Proceedings 2009 8th IEEE International Conference on ASIC, IEEE: 674–677, 2009, doi:10.1109/ASICON.2009.5351336.

[12] J. Liu, L.Sun, Z.Yu, "An accurate compact model for on-chip vertically coiled transformers," IEEE Electron Device Letters, **34**(4), 484–486, 2013, doi:10.1109/LED.2013.2247559.

[13] H. Liao, C.Wang, "A New Approach to Parameter Extraction for On-Chip Symmetric Transformers," in 2009 IEEE International Symposium on Radio-Frequency Integration Technology, RFIT 2009, IEEE: 257–260, 2009.

[14] S. Iqbal, "A hybrid symmetrical voltage multiplier," IEEE Transactions on Power Electronics, **29**(1), 6–12, 2014, doi:10.1109/TPEL.2013.2251474.

[15] X.J. Wang, T.Kwasniewski, "Propagation Delay-Based Expression of Power Supply-Induced Jitter Sensitivity for CMOS Buffer Chain," IEEE Transactions on Electromagnetic Compatibility, **58**(2), 627–630, 2016, doi:10.1109/TEMC.2016.2517635.

ASTES

iDRP Framework: An Intelligent Malware Exploration Framework for Big Data and Internet of Things (IoT) Ecosystem

Osaretin Eboya[*], Julia Binti Juremi

Asia Pacific University (APU), Faculty of Computing, Engineering, and Technology (FCET), Kuala Lumpur 57000, Malaysia

A B S T R A C T

The Internet of Things (IoT) is at a face paced growth in the advanced Industrial Revolution (IR) 4.0 in the modern digital world. Considering the current network security challenges and sophistication of attacks in the heavily computerized and interconnected systems, such as an IoT ecosystem, the need for an innovative, robust, intelligent and adaptive malware attacks and threats security solution is becoming predominant in the current cyberspace. An integrated and scalable IoT malware detection framework called iDRP framework with deep learning method was proposed as a solution to current IoT malware attacks that are largely obfuscated. The novel framework utilized systematic pre-processing and post-processing techniques and methods on the BoTNetIoT malware datasets that contains both benign and malicious IoT traffic data infected by modern day IoT attacks such as Mirai and Gafgyt etc. IoT malware variants in an IoT ecosystem. The raw IoT malware binaries were converted to image files (Gray-scaled) and computed statistically with synthesised sparsed and differential evolutionary hidden feature structures techniques, which were cyclically trained, tested, and cross-validated to establish empirical anomalies with precision in the detection, recognizing, and prediction of malware anomalies in a modern IoT ecosystem. Preliminary experiments were conducted with standardized image binary files such as the MNIST (2-D), and NORB (3-D) datasets as sound scientific exploratory experiments with profound results. The comparative results of the performance of our integrated techniques and methods on the BoTNetIoT IoT malware datasets achieved a 99.98% accuracy, 99.99% ROC/AUC, 99.95% precision, and 99.93 recall rate etc. utilizing the integrated iDRP framework mechanisms for effectively detecting IoT malware in an IoT ecosystem.

1. Introduction

This paper is an extension of the work originally presented in *IEEE 8th R10 Humanitarian Technology Conference (R10-HTC) 2020* [1]. The Internet of Things (IoT) has in recent times emerged as ubiquitous technology to everyday lives especially with the advent of the Industrial Revolution (IR) 4.0 in the digital world. The IoT technologies comprises of smart devices and objects interconnected in a heavily computerized network environment that constitutes the backbone of modern innovative critical infrastructures that supports fast-paced technological driven world. The IoT technologies has been disruptive in many industries such as: (1) IoT in Healthcare, (2) IoT in Manufacturing, (3) IoT in Transportation and Mobility, (4) IoT in Buildings, (5) IoT in Cities, (6) IoT in Agriculture, (7) IoT in Energy, (8) IoT in Retail and Marketing, (9) IoT in Logistics and Supply Chain, and (10)

IoT in Industries across various sectors of the economy. Similarly, IoT technologies has brought diverse approaches in solving mundane tasks in these sectors by leveraging and amalgamating innovative and intelligent automation systems. Additionally, the IoT technologies at the present time has been able to help connect multiple devices and objects in an expansive interconnected network area while collecting high treasure trove of data – *big data*. The big data collected in an IoT ecosystem has become paramount to making business decisions, analytics, and gaining valuable insights in multidimensional industries in the modern era. Furthermore, this big data that are continuously generated in an a heavily interconnected environment such as the IoT ecosystem has inadvertently set off more attractions and incentives to the cyber squatters and cyber criminals in a digital world – *Cyber hacking*. The security of IoT ecosystem is ever more crucial and concerning to safeguarding critical infrastructures in the smart environments against prolific attacks and state-of-the-art attacks – *Cyber*

intrusions. For example, in recent time there have been a number of cyberattacks that caused outages and losses such as the (https://www.msn.com/en-my/news/world/dc-police-suffer-massive-info-leak-after-ransomware-attack/ar-BB1gHTQl) attack and (https://www.msn.com/en-my/news/world/european-hackers-given-us-5-million-for-key-to-reopen-us-pipeline/ar-BB1gIq7K) outage coupled with the public health (https://www.msn.com/en-my/news/world/ireland-tests-cyber-attack-data-fix/ar-AAKe3Qd) compromise disruptions on critical infrastructures that serves an entire community in the modern society. This is one of the most significant attacks in recent time with massive consequences. The cyberattacks on critical infrastructures evidently indicates how vulnerable and compromising the present-day heavily computerized network systems can be, and consequently disruptive and detrimental to everyday lives in a digital world. Moreover, this clearly indicates that most *modern* highly interconnected critical infrastructures serving every day processes in the most developed world can be highly susceptible and vulnerable to being hacked with a myriad of modern malware attacks such as Ransomware attacks and '*Zero Day*' attacks. Theoretically, in the modern digital world, any device or object that can connect to a network system or grid can be hacked.

Considering the general consensus in the cybersecurity space that humans are usually the weakest link in the chain of command in the cybersecurity and network security effort, it is effectively safe to deduce that the current static, dynamic, and automated malware defense approaches against cyberattacks and cyber threats in an IoT ecosystem in a digital world is essentially a futile effort. Particularly in the bountiful cyberattacks and cyber threats warfare in the modern cyber space clearly indicates the fragility and exposure of current interconnected critical infrastructures and its related network systems. This has essentially led to the theory that *detection* is a better approach to combating cyberattacks and cyber threats in a smart ecosystem of massively interconnected computer network systems – *malware detection* [1]. Malwares are typically spread over an interconnected computer network system otherwise known as the internet, which is the core backbone of an IoT ecosystem. The lack of global unified principles and security of IoT protocols present major security and counter-intelligence conundrum on a daily basis in the IoT cyberworld. This inadvertently exposes various valuable resources and assets in an organization to the cybercriminals. The IoT ecosystem can be attacked, infected, and breached by different types of malicious software commonly known as malwares with categorized malware families such as: (a) Ransomwares, (b) Spywares, (c) Scareware, (d) Adware, (e) Viruses, (f) Worms, (g) Rootkits, (h) Botnets, (i) Trojan Horses that all threatens the *integrity*, *privacy*, and *security* of big data in a heavily computerized computer network system like an IoT ecosystem – *tampering of valuable assets; big dataset*. One of the major problems [1] with these malware families and their nefarious activities combined with innovatory technologies is that they have become even harder to detect with traditional and automated malware detection tools and techniques in a heavily computerized network system – *polymorphic malwares*. If and when potent malwares go undetected in a network system such as an IoT ecosystem, this can cause massive damages and disruptions to critical computer network infrastructures [1], loss of personal information, and potentially loss of livelihood and lives in a

modern technological-reliant society – *obfuscation and evasion of malware detection*.

We proposed a novel IoT Security framework known as the *intelligent*, *detection*, *recognition*, and *prediction* (iDRP) framework [1] to address these computer network security challenges and problems in the modern digital world. The proposed iDRP framework is an innovative, intelligent, robust and adaptive framework with ultramodern approach to solving malware anomalies in an IoT ecosystem. In the unique approach proposed, a synergy of the subsets of Artificial Intelligence (Ai); Machine Learning (ML) and Deep Learning (DL) will be synthesized with modularity and scalability implementation for a lifelong training and learning capabilities. Eventually, the goal of the research work is to effectively detect, recognize, and predict anomalies in an IoT ecosystem innovatively using a subset of Artificial Neural Network (ANN) with focus on Multilayer Perceptron (MLP); *hierarchical* DL techniques i.e., Deep Neural Network (DNN) to solving the wide range of polymorphic malware anomalies problems in an IoT ecosystem.

2. Related Work

The drawback of heavily depending on the storage of big data set in one or more interconnected storage medium is that adversaries now have the perceived knowledge and understanding of where to target in their quest in wreaking havoc on present-day interconnected network systems. In the event that proper anomaly analysis mechanism for countering the attacks on the treasure trove of big data is not put in place, then, it means that the assailants will in most cases get away with their attacks either undetected or worst-case scenario, reside on the host system while plunging the gigantic and diverse data in the system [1]. The lack of highly-developed, efficient, robust security apparatus with well-established intelligent processes that can accurately measure and predict future occurrences of such attacks could be disastrous to the specific individual, organization, or government alike. The general IoT Security challenges range from Authentication, Access Control, Privacy, Confidentiality, Integrity, Policy Enforcement, Trust, Secure Middleware, Mobile Security [2],[3] [4], and Network Security. The complexity of the interactivity between the gigantic number of smart devices and objects together with smart network systems give rise to a major security gap in the IoT ecosystem. This poses big security risks due to the incessant and sophisticated cyber-attacks in the modern internet age – *IR 4.0*. A common denominator amongst all the aforementioned IoT Security challenges is the Distributed Denial of Service (DDoS) attack. The DDoS attacks occur on major global establishments on a regular basis [5]-[7]. Several peculiar IoT Security constraints such as integrity, availability, authentication, authorization, and privacy have hampered the efforts to address the growing IoT Security challenges posed by malicious actors.

The malicious attacks on large corporations such as Fortune 500 enterprises have clearly shown that the security of internet-connected devices and objects is paramount considering the fact that the particular DDoS Malware attack called the "Mirai Botnet" crippled even the network servers of these largest technology corporations. Some of these large corporations are situated in the United States and major European countries [8], (https://www.theguardian.com/technology/2016/oct/26/ddos-

attack-dyn-mirai-botnet), and (https://www.wired.com/story/reaper-iot-botnet-infected-million-networks/). This type of nefarious malware attacks contributes to the urgent needs for a viable solution to the lingering IoT Security challenges [6], [9], [10] with the ever-growing IoT inter-connected network environment. According to [8], [11], [12]. One of the security challenges of smart devices and objects today is the inability to *detect* when there is an intrusion in the IoT network system. This is largely due to the fact that most smart devices lack the capability to log and report when such malicious attack occurs in the IoT ecosystem. IoT Security needs to be improved and revolutionized to prevent, identify, and neutralize the malicious network traffic threats on the Billions of IoT connected servers. The exponential ramifications of the risks and costs that the malicious attacks on the IoT infrastructures pose [13] to businesses, governments, and individuals who rely on IoT services such as Healthcare, electricity, industrial production, social amenities, communications, and logistics is still unknown. Nevertheless, the researchers discovered that, the two major challenges that an IoT ecosystem currently face are; Low-end Embedded Devices [6], [8], [14] with insecure Operating System (OS), and Mirai-Styled Malware Attacks [10] otherwise known as Mirai Botnet, which is a modern and sophisticated variation of a Distributed Denial-of-Service (DDoS) attack on the current precarious IoT environments and its smart devices and objects.

2.1. IoT Attacks

IoT applications are prone to attacks from different hostile actors. The multidiscipline industries that adopt IoT as a technology have their own variation of attacks on their IoT applications and infrastructures. The concept of sensors connecting to each other and transmitting big data in a two-way traffic in an IoT ecosystem typically creates vulnerabilities [13], [15]-[17] in the IoT network system. In the Smart City ecosystem in Ukraine that utilizes Smart energy grid system to serve the city, 30 of such Smart energy were attacked [8], [16], and (https://www.thenational.ae/uae/smart-cities-open-door-to-cyber-attacks-say-security-experts-1.672214), and 80,000 residents were without power supply for several hours due to the adversarial malicious attack on the Smart Grid system. In the United States, 156 emergency sirens designed to alert residents of the city of Dallas [8], [18], and (https://www.nytimes.com/2017/04/08/us/dallas-emergency-sirens-hacking.html) in the event of critical emergencies such as earthquakes, floods, and nuclear attacks were set off because of an attack [2] on their Smart City system. The impact of such false alarm set off in such a malicious manner can be consequential in the end. There is a high likelihood of the residents of the city to ignore a true alarm warning in the event of an actual emergency, which could potentially lead to loss of lives and livelihood [2]. Similarly, 2,200 residents were without water supply (https://www.wired.com/2011/11/hackers-destroy-water-pump/) when hackers attacked and destroyed the Smart Water Pump Supply system in the municipality of Springfield, Illinois, United States.

The malicious attacks on IoT applications are not limited to Smart Cities and critical infrastructures alone, but also Smart Automobile compute system where attackers targeted a popular Smart Car maker and hijacked its core automation compute system that powers their connected cars for their financial gains. The

financial gains, also known as "*Cryptojacking*" [13], [18], [19] where hackers use illegally acquired computational resources to mine the cryptocurrency shows that hackers will launch attacks on IoT applications for monetary reasons. In Smart Healthcare system [13], attackers have successfully hacked [6], [8] a heart monitoring electronics health (e-health) infrastructure that intelligently monitors and analyses patient's heart rate and helps prevent cardiac arrest and heart attacks [20]. Furthermore, the security of a Smart Home system that helps parents monitor their infants was compromised when attackers [6], [8] were able to obtain the login credentials of the Smart Home system and about 700 Webcam feeds of babies were posted on the internet [13], [20] in a malicious attempt by the hackers. The security breaches ultimately led to privacy [20] on the victims of the IoT attacks. All these incessant IoT attacks [6], [13] indicate that IoT devices and objects are more vulnerable and susceptible [19], [21] to various hostile adversaries who are constantly looking for weaknesses [16], [18] in the IoT infrastructures and IoT software applications

2.2. IoT Malwares

Malicious software and worms are very devastating to the IoT applications' ecosystem. Malware such as worms pose greater threats to the security of IoT applications. Some of the known malware worms are Mirai Botnets [22], Ransomware [20], [23], Over The Air (OTA) Worm [21],[24], and Struxnex Worm. According to the case study and analysis conducted by [18], discovered that Mirai Botnet can be directed towards the disruption of the availability of the targeted IoT application and the IoT compute resources in the IoT environment. In addition, [8],[23] classified the Ransomware as primarily Crypto Ransomware, and Locker Ransomware, which encrypts and locks critical files, respectively in an IoT ecosystem have increased by a staggering 670% and 350% between 2015 and 2017 respectively and continue to grow exponentially nowadays. The implications of such malicious software are severe and could disrupt and harm crucial IoT infrastructure and potentially cripple an entire IoT ecosystem as evident [13],[19] in the coordinated attacks by hackers on the key connected internal applications in the city of Atlanta, United States.

The OTA Worm exploits vulnerabilities in the IoT smart devices and objects and spreads in the IoT environment by compromising the integrity, privacy, and security of the interconnected devices and objects therein. The OTA Worm embeds itself into the IoT ecosystem thereby causing a breakdown of the IoT ecosystem referred to as an "epileptic seizure" [6], [16], [24] of the specific IoT smart devices and objects at a close range of 350 meters to the building with the help of a drone or Unmanned Arial Vehicle (UAV) equipped with the malicious worm. Moreover, the mode of launching the OTA Worm attack is also carried out with mobility such as physically getting close or driving cars to a close proximity of the IoT subject. Even though the OTA Drone Worm [6], [16], [24] was successfully launched on a major Smart home appliances manufacturer's Smart bulb product line, it was, however, performed in a controlled environment and the concerned manufacturer of the defected IoT Smart Light Bulb was notified, and the appropriate fix was implemented, which now reduces the possible attack range to 1 meter [6], [16], [24]. However, the risk of exploitation by enemies is present. This shows that malicious adversaries are becoming more sophisticated

in their approach towards unleashing malicious worm attacks with technological advancement. Furthermore, the Struxnex Worm, which is a variant and obfuscation of malicious worm was directed at Smart Industry such as the Uranium plant [6] connected to an IoT infrastructure to intelligently monitor and analyse the increment and decrement of the acceleration of the centrifuge and enrich the Uranium plant as the need arise autonomously. The main goal of the Struxnex worm is to infect and break down the smart devices and objects in a Smart Industrial ecosystem [18] such as a Uranium plant.

2.3. IoT Existing Security Challenges and Gaps

The core existing gap in the IoT Security paradigm is IoT security, IoT privacy [3], [4], [17], [19]. However, this research has gone further with the IoT security and IoT integrity dimension. The importance of anomaly detection in an IoT ecosystem cannot be overemphasized in various IoT industries such as Smart Healthcare, Smart Education, Smart Agriculture, and Smart manufacturing, and Smart Buildings and Smart Cities [13], [25]-[28]. In the IoT Security dimension, we explored and discovered that research has been done on the Classification and Visualization of IoT security [20], [21], [29] in which the researchers performed experiments to visualize and compare different classes of malware in a similarity matrix table based on the images and features of some malware samples such as Worms and Trojans Horses. Their research work can be improved further to increase the accuracy of the detected similarities of the various malware samples and increase the performance of the workload on big data. Moreover, the experiment was conducted using 'mechanical' compute method, which means that their method could possibly detect existing malware at a significant slow pace, coupled with the fact that, their classification and visualization method cannot intelligently predict future occurrences [13], [19], [20] of such IoT malware anomalies autonomously.

3. Methodology

We present the schematic illustration process of the work proposed iDRP framework, including the realization of the universal approximations of Multilayer Perceptron (MLP) with multiple hidden layers by exploiting the capabilities of H-ELM and other related techniques [30]-[33], which have the preeminent underexplored potentials for accelerated speed, rapid feature learning, and improved classification performance [30], [31], [34], [35].
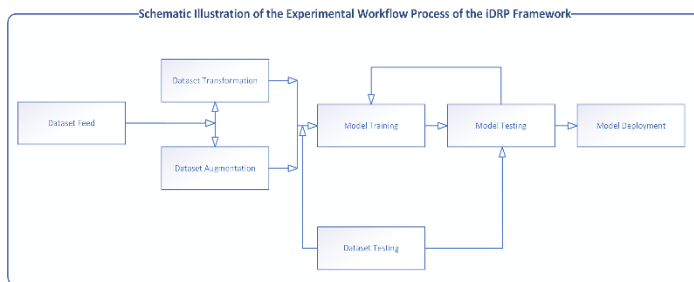
Figure 1: Schematic Illustration of the Experimental Workflow Process of the iDRP Framework

Figure 1 Shows the schematic illustration of the setup of the experimental workflow process of the iDRP framework. The big

dataset will be fed into the system, when then gets transformed by various MLP techniques such as augmentation, normalization and classification techniques, and subsequently prepared for the algorithm. Feature extractions will be carried out by augmenting and fine-tuning the big dataset for the generalization and capability of the proposed iDRP framework.

The experimental implementation is to demonstrate the proof-of-concept of the proposed iDRP framework using the generalized MLP techniques for demonstrating how a non-traditional malware detection mechanism [32], [33] can be applied to systematically, dynamically, adaptively, and intelligently detect, recognize, and predict malware anomalies in the generated big data of an IoT ecosystem through converted image file, such as a log file, and binary dataset to create synthesized Machine Learning (ML) and Deep Learning (DL) models. Based upon the premise that attackers are getting more sophisticated with the advancement of technology every day, and considering that attackers have been developing malwares that can dynamically modify their signatures while in the network environment, which approximately makes it more difficult to detect. This fundamentally relegates the notion of applying the traditional and automated Signature-Based Analysis method to solving a potential attacks and threats in an IoT ecosystem. Furthermore, the Heuristic-Based Analysis method, which is otherwise known as Behavioural-Based Analysis technique applied by several Anti-Virus and Anti-Malware vendors on computer network systems has not been sufficient either. Hence, there are various gaps [19],[20],[36] in the field of study to develop a novel approach to solving the inherent and existential security threats in computer network systems, especially in the fast-growing IoT ecosystems that has now become ubiquitous to everyday life. The hypothesis is that nowadays, contaminated and malicious programs in an IoT ecosystem can easily masquerade themselves as benign software applications once in the network space. Therefore, a dynamic, robust, and intelligent Pattern-based model approach to discerning and distinguishing the genuine and malicious program in a system as counter-intelligence is paramount as the building block to proactively combating an unwanted and illegitimate program from executing and consequently causing some irreversible damages and potentially loss of lives in an IoT ecosystem with its critical infrastructures.

To explore the binary classification and analysis techniques, we will be using the standardized Modified National Institute of Standards and Technology (MNIST) dataset (http://yann.lecun.com/exdb/mnist/), which has been demonstrated to have the capability to be embedded with malicious codes (virus-MNIST) (https://arxiv.org/abs/2103.00602) adaptively, NORB (https://cs.nyu.edu/~yann/research/norb/) to perform the experiment by encoding the dataset, which will serve as a standard benchmark for comparative training, testing, evaluation, and analysis of the experiment.
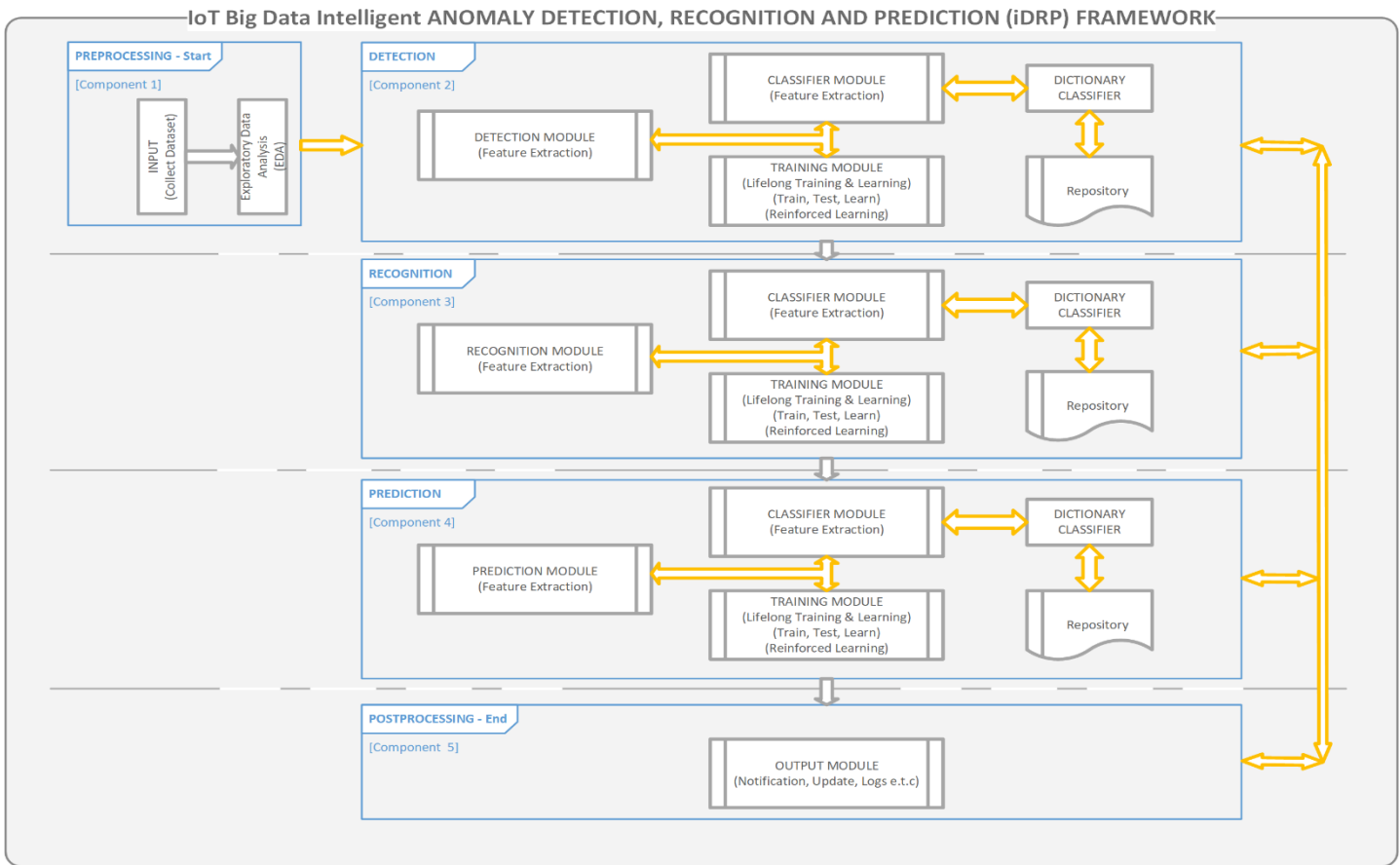
Figure 2: Proposed iDRP framework for Anomaly (Malware) Identification in IoT ecosystem – Overview [1]

4. The Proposed iDRP Framework Architecture

This section provides the analysis of the proposed Intelligent anomalies detection, recognition, and prediction (DRP) also known as iDRP-framework as shown in Figure 2. The conceptualization of the proposed iDRP-framework is based on the relationship between feature extraction, feature classification, precision in the identification of anomalies, and predictions of anomalies in the generated dataset of IoT ecosystem. Table 1 and 2 displays the synthesis of the list of the development techniques comparatively in tandem with the corresponding frameworks' techniques to be synthesized and applied for the development and validation of the proposed intelligent framework. The information gathered in the literatures of this research study has clearly shown that there still exist gaps in effectively tackling the myriad of anomaly detections, identifications, and recognition in complex data sets generated in an IoT ecosystem.

The limitation in the forecasting of anomalies (malwares) in the IoT ecosystem forms the spine of this research study and consequently the creation of an Intelligent framework called iDRP-framework to address the issues of anomalies (malwares) identification in an IoT ecosystem. In particular, the ability to accurately Detect, Recognize, and Predict (DRP) anomalies (malwares) from logging big datasets generated in an IoT ecosystem cannot be overemphasized.

Evidently, the extraction of miniscule features of the big datasets to assist in the '*fine-grained classification*' of features of complex datasets has been demonstrated by other researchers

[30],[36] to help in the precision of anomaly identification in an IoT ecosystem.
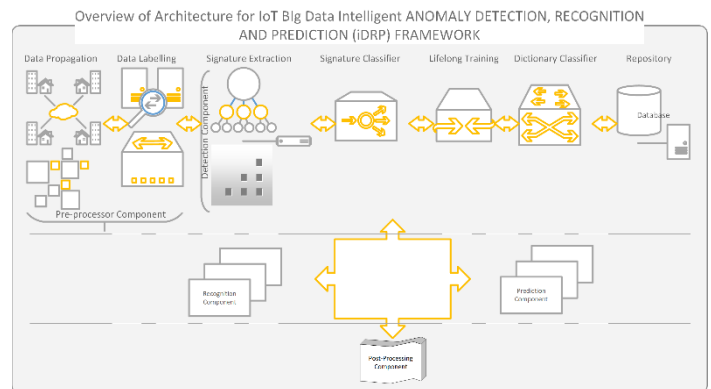
Figure 3: Proposed iDRP framework for Anomaly (Malware) Identification in IoT ecosystem – Architecture Breakdown

Figure 3 illustrates the overview of the proposed iDRP framework architecture with the "*Data Augmentation*" segment consisting of data propagation, data labelling, feature extraction, and feature representation etc. The proposed iDRP framework architecture is flexible, robust, and adaptive in its design to intelligently and effectively extract mapped minuscule information in the stacked layers of the proposed iDRP framework for detecting, recognizing, and predicting malware binaries in an IoT ecosystem.

Table 1: Synthesized Components of Development and Validation of the Proposed iDRP Framework I [1]

No	Features	Comparative Features of Synthesised Frameworks I						
		H-ELM	SAPIM	Smart Weather	Flight Monitoring	eTRIKS	Data Lifecycle	DRIPROM
1	Pre-processor	✓	✗	✓	✗	✓	✓	✓
2	Detection	✓	✓	✓	✓	✓	✓	✓
3	Recognition	✓	✗	✗	✗	✗	✓	✓
4	Prediction	✗	✓	✗	✗	✗	✗	✗
5	Post-processor	✗	✗	✗	✗	✗	✗	✗
6	Lifelong Learning	✗	✗	✗	✗	✗	✗	✗
7	Lifelong Training	✗	✗	✗	✗	✗	✓	✗
8	Cyclic Feed	✗	✗	✗	✗	✗	✓	✗

Key: H-ELM (Tang, Deng and Huang, 2016), SAPIM (Wang et al., 2018), Smart Weather (Onal et al., 2017), Flight Monitoring (Li, Ming and Li, 2017), eTRIKS (Oehmichen et al., 2017), Data Lifecycle (Arass, Tikito and Souissi, 2018), DRIPROM (Cuzzocrea and Damiani, 2018).

Table 2: Synthesized Components of Development and Validation of the Proposed iDRP Framework II

No	Features	Comparative Features of Synthesised Frameworks II					
		IoT-HarPSecA Framework	3-Way IoT Framework	IoT-Flock Framework	FIFAC Framework	BiDeL Framework	Combat Intelligence Framework
1	Pre-processor	✓	✗	✗	✗	✓	✓
2	Detection	✓	✓	✓	✓	✓	✓
3	Recognition	✓	✗	✗	✗	✗	✓
4	Prediction	✗	✓	✗	✗		
5	Post-processor	✗	✗	✗	✗	✗	✗
6	Lifelong Learning	✗	✗	✗	✗	✓	✗
7	Lifelong Training	✗	✗	✗	✗	✗	✓
8	Cyclic Feed	✗	✗	✓	✓	✗	✓

Key: IoT-HarPSecA Framework (Samaila et al., 2020), 3-Way IoT Security Framework (Zeeshan, Reed and Siddiqui, 2019), IoT-Flock Framework (Ghazanfar et al., 2020), FIFAC Framework (Awadelkarim Mohamed and Abdallah M. Hamad, 2020), BiDeL Framework (Otoo-Arthur and van Zyl, 2020), Combat Intelligence Information Monitoring Framework (Jin, Xing and Wang, 2020).

Table 1 and Table 2 lists the synthesis of the development techniques comparatively in tandem with the corresponding frameworks' techniques to be applied for the development and validation of the proposed iDRP framework.

4.1. Overview of Proposed iDRP Framework

Notably, Figure 1 shows the diagram of the proposed iDRP framework with extraction mechanism as a pre-processor and cyclical post-processor apparatus of logged complex dataset in an IoT ecosystem. These form the backbone of the novelty in the proposed iDRP framework. Significantly, from what has been learnt from the literatures of the existing knowledge domain that was conducted through the in-depth investigations, the aforementioned gaps in the anomalies (malwares) identification in an IoT ecosystem will be effectively tackled with the proposed technique for the novel iDRP framework.

4.2. Components of Proposed iDRP Framework

- *Preprocessing* – The pre-processing component, which is one of the major contributing steps in the framework to effectively address the aforementioned identified problems in the anomaly identification in IoT ecosystem. This component 1 involves the incorporation of Pre-Processor of a complex dataset by using ML technique like the enhanced traditional Data Augmentation techniques – blend of Enhanced Data Augmentation techniques and Propagations.

- *Detection* - The signatures extracted in component 1 will be fed forward to the component 2, the detection component for effective classification of signatures and labels i.e., '*fine-*

grained classification' mechanism for feature learning to precisely detect Malware anomalies in an IoT ecosystem. Likewise, the Learning and Training of anomalies signatures experiments will be conducted extensively combined with Dictionary Classifiers and the Database of both the new and historical malwares in the detection component.

- *Recognition* - A similar procedure will be performed in the component 3, the recognition component, with the primary aim of precisely recognizing each specific Malware in the logged complex data of IoT ecosystem.

- *Prediction* - Correspondingly, in the component 4, Prediction component, the forecast of anomalies will be rationalized and realized by synthesizing the components 2, 3 and 4 for accurate predictions of the occurrence of anomalies in an IoT ecosystem.

- *Postprocessing* - In component 5, Post-Processing, the assessed anomalies will be fed back into the entire system in a cyclic manner. This '*feed-forward*' and '*feed-backward*' propagation methods that were applied and implemented in the model of the framework will essentially ensure that the framework is well-adjusted, robust, intelligent, and adaptable adequately to address the growing threats and risks in the IoT security paradigm.

4.3. Preliminary Experiment for Proposed iDRP Framework

The preliminarily experiments were conducted using MNIST dataset, a standardized dataset, which is comprised of small 28x28 pixels of randomized hand written number images with corresponding annotated numbers as labels. The MNIST dataset presents a better experimental starting point for developing the MLP for the recognition of patterns in the converted log file image with minimal pre-processing and formatting overhead cost for the ML and DL models. The standardized MNIST dataset used in this experiment contains 34,300 training set samples and 14,700 test images of 28x28 = 784 pixels for each image, which is a subset of original 60,000 training set samples of the standard benchmark MNIST dataset.

4.4. Conceptual Implementation of Proposed Framework

The Data Augmentation techniques and propagation techniques [30],[31],[37], which helped with the '*fine-grained extraction and classification*' of signatures of complex datasets were applied to the model implementation of the proposed novel iDRP framework. The applied research approach for the development of the iDRP framework is implemented using Python Programming language (https://www.python.org/), which is an open-source language with rich and extensive libraries that is easily accessible to the general masses. Keras (https://keras.io/), a high-level library that is built upon a low-level library such as TensorFlow (https://www.tensorflow.org/) is used for implementing the required Neural Network (NN) with simplistic interface that is built on the Python platform as a native Python library. TensorFlow serves as the tensor manipulation library to Keras, as a bridge low-level Application Programming Interface (API) for the robustness of implementing minimalistic and concise NN models. Significantly, Keras has several extensible modules that are suitable and adaptable to advanced scientific research work, which has formed the basis for the informed decision to choose the combination in implementing the complex NN models

especially for the stacking of the various layers; hierarchical models for this experiment.

4.5. Neural Network (NN) of Framework

The NN has been able to provide access and manipulation to the fed benchmarked dataset dynamically while exploring the binary features of the converted logged image files with different categories in the pre-processing of the dataset, while the necessary adjustments were made accordingly. Importantly, different layered NN combinations have been explored while calling upon the Functional model, '*tf.keras*' to provide the building blocks and backend support for the configurable NNs [29] for the training, testing, and prediction in the application. The novelty of the approach in the development of the proposed iDRP framework is to use dynamic Pattern-Based model approach to intelligently and accurately detect, recognize, and predict anomaly in an IoT ecosystem.

4.6. H-ELM Derivative Formula

A formular for the H-ELM was derived from [30],[31] as a result of the amalgamation of the Radial Bias Function (RBF) nodes with the defined Activation function [30] for mapping indiscriminate features of the converted logged datasets of an IoT ecosystem. The primary aim is to precisely approximate the continuous targets with mobility at the evolving RBF nodes adaptively [30],[31],[37]. This ensures the irregular initialization of the hidden nodes H in a given set of big data in an IoT ecosystem. This is expected to significantly reduce the training errors while increasing the precision in the accuracy of the output with better performance and speed.

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \dots & h_L(x_1) \\ \vdots & \vdots & \vdots \\ h_1(x_N) & \vdots & h_L(x_N) \end{bmatrix} \quad (1)$$

where H is the randomized hidden layer output matrix, which denotes the hidden nodes and output weights of the Neural Network, while $x_1 \dots x_N$ is the number of training dataset.

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \dots & t_{1m} \\ \vdots & \vdots & \vdots \\ t_{N1} & \dots & t_{Nm} \end{bmatrix} \quad (2)$$

where T represents the training dataset of the target matrix.

In order to optimize the performance of the ELM feature mapping, the following formula is applied:

$$\beta = H^T \left(\frac{1}{\lambda} + HH^T \right)^{-1} T \quad (3)$$

The resultant output of the ELM feature mapping will be

$$f(x) = h(x)\beta = h(x) H^T \left(\frac{1}{\lambda} + HH^T \right)^{-1} T \quad (4)$$

Alternatively, the equation is condensed to

$$\beta = \left(\frac{1}{\lambda} + HH^T \right)^{-1} H^T T \quad (5)$$

The derived output function of the ELM feature mapping is

$$f(x) = h(x)\beta = h(x) \left(\frac{1}{\lambda} + HH^T \right)^{-1} H^T T \quad (6)$$

In the hidden nodes H of the training dataset, hidden information can now be exploited and extracted from the hidden layers h to transform T the output weights β. This will ensure minimal fine-tuning of the parameters or weights of the hidden layers with stable minimal overhead cost to achieve maximum output. This randomized approximation and differential evolution technique has been proven [30]-[32] to be effective, which ultimately secures the increment in both the feature learning and classification.

5. The Need to Convert IoT Malware Binaries to Images

The technological advancements in present-day cyberattacks has made the activities of advanced attackers more complex to detect as a result of emerging obfuscation techniques [27], [38], [39] and interactions [40] with stealth variations carried out on IoT ecosystems.

In modern times, the emerging polymorphic malware attacks in the IoT ecosystems have been a major concern [1] due to complex obfuscation code structures that are mostly time based [41], [42]. These IoT malware signatures attacks that are predominantly multivariate [39], [40], [42]-[44] are updated sequentially on a minute-by-minute or hour-by-hour basis by the attackers, thereby inundating and *silencing* any potential alert system, which may cause massive vulnerabilities for exploitations in an IoT ecosystem. The current widespread detection and mitigation mechanisms for these emerging polymorphic IoT malware attacks that are largely obfuscated intricately can be problematic and resource intensive to both the traditional and automated malware detection solutions such as the signature based (e.g., large database), and automated based techniques (insufficient information) etc., adopted by major cybersecurity vendors, practitioners, and researchers in the cross-discipline cybersecurity industries. The domain experts and analysts may write different rules manually or automatedly to detect complex obfuscated malware scripts in an IoT ecosystem as a possible approach to solving the problem, the caveat to such an approach in determining the benignity or maliciousness of such malware binaries in an IoT ecosystem is that it would be near impossible to efficiently and accurately determine either a clean or infected malware script in the IoT ecosystem – *subtle obfuscated IoT malware classification problem*. It certainly would be time and resource intensive to adequately write thousands and/or millions of rules for virtually all current and new malware obfuscation variants in the contemporary IoT ecosystem.

Evidently, the aforementioned complex IoT malware detection approaches are no longer suitable for current growing subtle and complex obfuscated cybersecurity paradigm for five (5) major reasons: (1) the distinct fragments (sections) of converted IoT malware image file can be visualized [27], [38]; (2) the distinct textural patterns of the converted IoT malware image files can be classified (and reclassified), and analyzed using image processing [27], [45] classification techniques commonly applied in computer vision [30], [31]; (3) the advanced attackers typically modify or reuse codes to generate new variants of complex obfuscated IoT malware scripts [38]-[40]; (4) the vector structures of converted IoT malwares image files of the same family are similar [27], which means vectorized representation (i.e., faster speed) of the IoT malware image files will be sufficient to be fed directly as

input into the similar ML-DL algorithms [30], [31] built for image and pattern recognitions [27], [45]-[47]; (5) the captured IoT malware dataset once converted to image files will be immutable (nondestructive and tamper proof), so there is no workaround its state by the attackers. Therefore, an amalgamation and syntheses of the modern ML-DL techniques and methods designed utilizing converted immutable IoT malware binaries to image files is both a logical and pragmatic solution to the current rapidly growing complex IoT malware problems that would scale adaptively.

5.1. The Benefits of Converting IoT Malware Binaries to Images

One of the biggest advantages of converting IoT malware binaries to Gray-scale image files is the fact [27] that the IoT image maintains the structures and patterns with rich information – *structural integrity*. The immutability of the IoT image files guarantees that converted complex IoT malware scripts will always be reliable to work with (ensuring a total control of the collected IoT malware dataset in the environment without the risk of compromise externally) by leveraging the ever-growing advancement of modern technologies such as Artificial Intelligence. Moreover, the structural integrity of the stately converted IoT malware image files facilitate the high-level precision and accuracy of detection of minuscule byte codes of the clean or infected codes with corresponding Cross-Validation techniques of modern ML-DL algorithms – *efficient detection, recognition, and prediction of anomalies in an IoT ecosystem*. Another advantage of converting IoT malware binaries to image file is the speed (time) of processing and computing the thousands and millions of both the benign and malicious IoT malware image file vectors using ML-DL techniques – *computational time*.

5.2. Visualization of IoT Malware Image for Exploitations

The textural and structural anatomy of IoT malware image files reveals the intricacies and patterns of the minuscule information that are sufficient for synthesized spatial and differential exploitations and analysis [27], [30], [31] of the tiniest byte codes (exploiting hidden structural and textural information by utilizing advanced ML-DL techniques and methods i.e., '*fine-grained*' layer-by-layer mapping and representations) of the features of converted complex IoT malware image files with high-level precision and accuracy in an IoT ecosystem – effectively distinguishing the benignity and maliciousness within a few bytes of codes as shown in Figures 4 and 5.
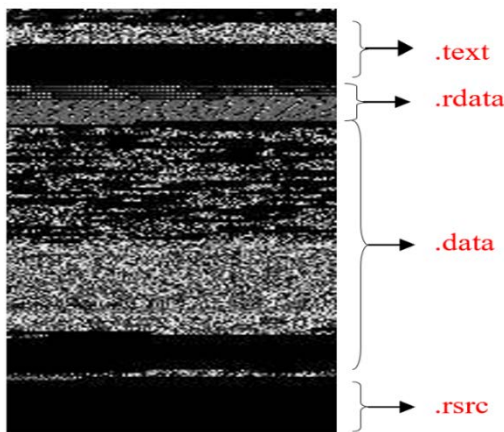
Figure 4: Visual structure of IoT Malware image file

The Figure 4 above shows the visual representation of the sections of a converted IoT malware image file with different segment textures (https://code.google.com/p/pefile/). The above section representations (*.text* = '*fine grained*' section with unique validity of data on the Operating System (OS) level e.g., Windows, Linux, *.rdata* and *.data* = the uninitialized and initialized scripts are the header and section table of the file that distinctively separates the unlaunched and launched codes respectively, *.rsrc* = the resource section that contains mostly padded zeros. These structures and patterns are important because the slightest of modifications (changes) in the variations of IoT malware binary codes can be easily detected stately – often within a few bytes.
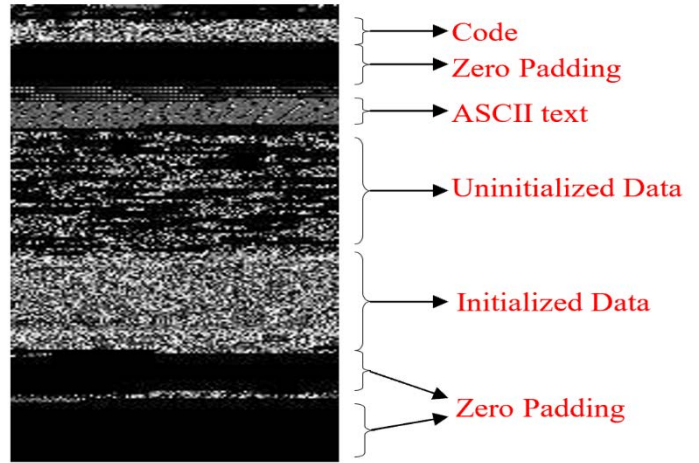
Figure 5: Visual structure of IoT Malware image file (Rich Information)

The Figure 5 above shows the rich visual representation of the section of an IoT malware image file with extensive details for exploitations and analysis. The segmented textures and patterns are essential for the pattern recognition of both benign and malicious IoT malware binaries converted to image files in an IoT ecosystem. The various segments of the IoT malware binary file structure are easily isolated for the synthesis of spatial and evolutionary feature mapping (i.e., spatial and evolutionary data matrix), extractions, representations, classifications, and Cross-Validation of the hidden minuscule features (i.e., the IoT malware signatures in the *header*, *sections 1*, *2*, and *section+n*) of the converted IoT malware image files by utilizing ML-DL techniques and methods with high-level precision and speed of detection of complex obfuscated anomalies in an IoT ecosystem.

5.3. Overview of Converting IoT Malware Binaries to Images

The conversion of typical binary files is based on standardized OS binary file format structures, which is analogous to debugging and homologous to reverse engineering of malware binary programs structurally. This can be simply performed using various *python* libraries such as *pefile* (https://code.google.com/p/pefile/) for fundamental exploration, understanding, and analysis of the layout features of a binary program. However, for a more in-depth exploration, exploitation, and analysis of rich hidden features, a systematic approach has to be applied to gain valuable insight into the various layers and architectures of the IoT malware binary file to image file. These procedures for converting IoT malware binaries to image files consists of the following five (5) fundamental steps: (1) Sort Dataset (i.e., variants, variant name,
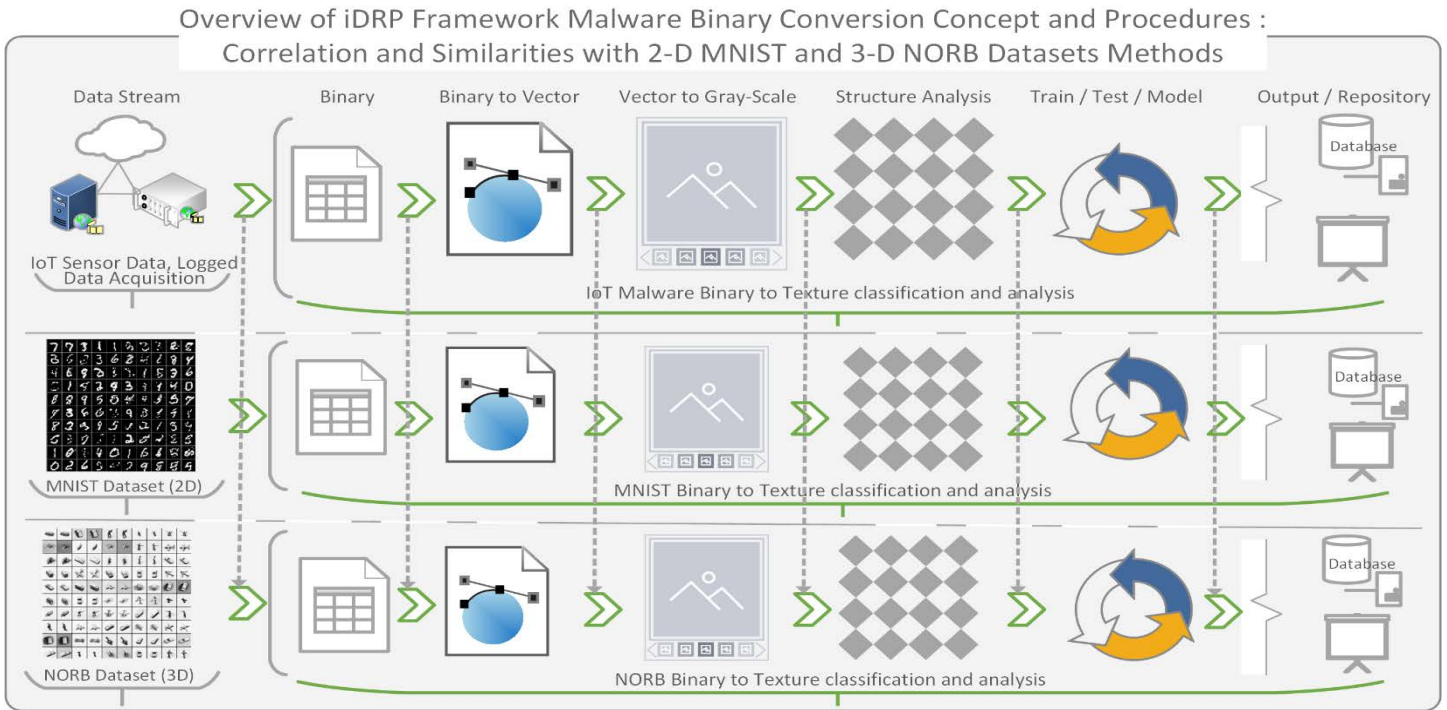
Figure 6: Comparison of similarities between the classification and analysis procedures for IoT malware images and MNIST and NORB datasets

no. of occurrence) retrieved by attributes; (2), Import Libraries (e.g., os, *numpy*, *scipy*, and *array*) to access and extrapolate the IoT malware binaries; (3) Augment IoT Dataset (i.e., *reshape*, *resize*, and *scale* et al) the dataset; (4) Standardize the IoT dataset; (5) Save the converted IoT malware images in a repository. Thereafter, feature selections techniques for consistent feature vectors (i.e., color range, intensity, pattern structure) can be performed on the converted IoT malware image files and processed as input feature classifications – *supervised learning*.

6. Experimental Implementation of iDRP Framework

The approach for the experiment of the proposed novel iDRP framework [1] involves the cleaning, manipulation, and extraction of IoT malware binaries to be preprocessed, and visualized as IoT malware image file for the evolutionary exploitation of hidden minuscule segment features for both supervised and unsupervised structural features representations and classifications with high-level precision to detect, recognize, and predict anomalies in a heavily computerized network system such an IoT ecosystem.

This experimental implementation of the proposed novel iDRP Framework is a proof-of-concept for the research study. The experiments were conducted with several distinct structural different datasets (BoTNet-IoT [43], MNIST, virus-MNIST, and NORB datasets etc.), yet similar conceptual procedures and methods (e.g., Figure 6) in the multiclassification and analysis to extrapolate meaningful hidden information for sound scientific evaluations and validation quantitatively. The Figures 7, 8, 9, 10 represent and demonstrate the correlations between the textual standardization methods of the IoT malware images, MNIST images, and NORB images datasets. This is to enable the systematic implementation and evaluation of the performance of the proposed novel iDRP framework techniques and methods (e.g., Figures 11, 12, 13, 14) for optimal malware binaries' classification and analysis in the present-day IoT ecosystem. The datasets used

to perform the experiments were converted to 8-bit gray-scale 2-D array with a channel range between 0 to 255 for consistencies in the datasets. Essentially, this process generates a vectorized 8-bit binaries, which are then converted to the gray-scale 2-D image files for structural texture uniformity for extracting and detecting padded hidden information in a logged IoT malware file.

(a) MNIST Samples A (2-D)   (b) MNIST Samples B (2-D)   (c) MNIST Distinctive Textures (2-D)
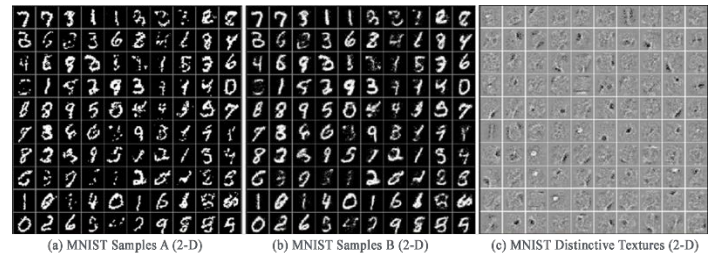
Figure 7: MNIST (2-D) images to Gray-scale images (0 to 255) standard channel

(a) NORB Samples A (3-D)   (b) NORB Samples B (3-D)   (c) NORB Distinctive Textures (3-D)
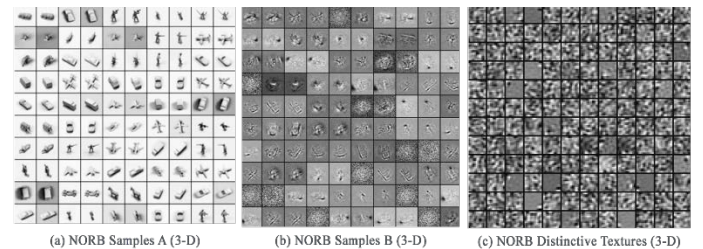
Figure 8: NORB (3-D) images to Gray-scale images (0 to 255) standard channel

The 3-D NORB image files, which has uneven generic object representations such as animals, vehicles, and nature etc. provide similar abundant natural environment and characteristics for representing, visualizing, and detecting certain distinct patterns of structurally complex textures of IoT malware variants that conceals hidden features that modern polymorphic and new malware engines in an IoT environment exhibits. This hypothesis was helpful for the scientific empirical and statistical observations

for the classification and analysis for the detection of IoT malware anomalies with high level precision and performance in the experiment for uniquely solving contemporary IoT malware obfuscation problems by attackers [1].

We performed the experiments on the premise and fact [48] that malware of the same family has similar characteristics and textures, while being distinct from other malware families in nature. Comparatively [48], for visual analysis of IoT malware binaries, the binary fragments of polymorphic and new malware variants are visually similar and consistently fragmented alike by the same token in terms of their image dissections and layouts as with their primitive nature irrespective of the malware variants or obfuscations and their specialized binary complex code structures.

(a) IoT Malware obfuscated payload (Trojan)   (a) IoT Malware unpacked

(b) IoT Malware obfuscated payload (Backdoor)   (b) IoT Malware unpacked
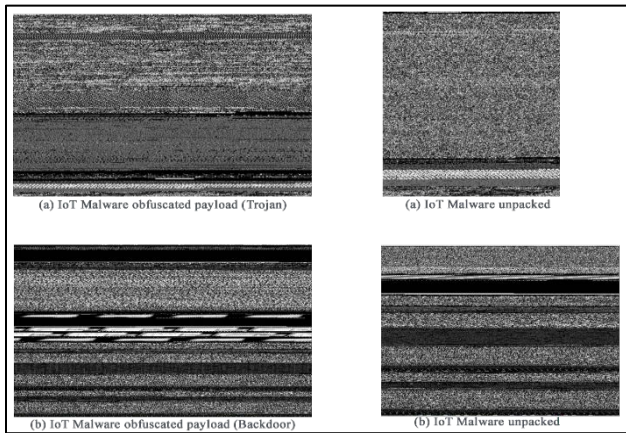
Figure 9: Captured IoT malware images with binaries I

Therefore, we approached the problem by proposing an iDRP framework techniques and methods that leverages both the H-ELM [30] and E-HELN [31] techniques and methods for universal approximation using differential evolution for optimized generation of complex image features to precisely extrapolate hidden features of IoT malware binary images files for expedient training and learning in the DNN. The advantage of these synthesized techniques and methods is that they both provide an optimized classification and characteristics learning and training coupled with minuscule and hidden feature representation modules for accurately *detecting*, *recognizing*, and *predicting* IoT malware binaries speedily.

(c) IoT Malware obfuscated payload (Worm)   (c) IoT Malware unpacked

(d) IoT Malware obfuscated payload (Virus)   (d) IoT Malware unpacked
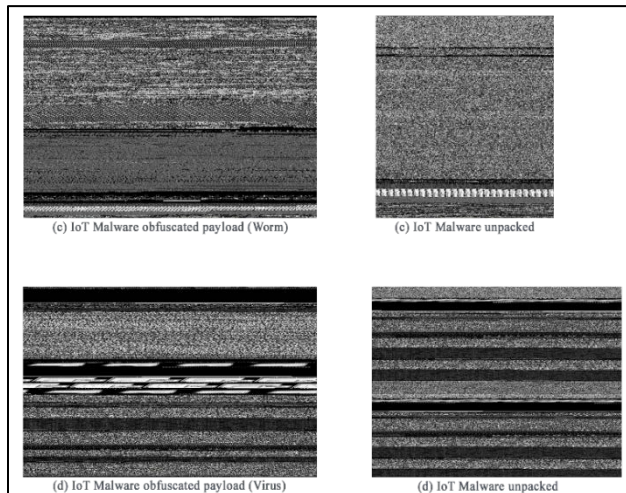
Figure 10: Captured IoT malware images with binaries II

The generative model techniques and methods that were applied cyclically in the iDRP framework to provide better discriminative capabilities for analyzing and evaluating the captured IoT malware binary information and statistics with high level precision and performance.

Overall, we explored and computed (gray-scaled etc.,) the binaries of the engineered BoTNetIoT malware dataset [43] exploits (Botnet attacks e.g., Mirai, Gafygt and their variants, and benign traffic etc. on IoT objects, devices, and environs), split them into 70% (training) and 30% (testing) ratio processed systematically and Cross-validated with a combination of both authentic benign and infected IoT malware traffic dataset with different strains [43], similar techniques and methods as with the statistical exploratory MNIST and NORB datasets experiments performed. The converted IoT malware gray-scaled images were normalized (min-max normalization, anti-aliasing techniques etc.), flattened, reshaped into 1-D and 2-D pixel and arrays for consistency and standardization. These were performed with the structured techniques and methods by applying the evolutionary adversarial training and testing techniques and methods for classification and Cross-validation to accurately estimate the comparative capabilities and skills (e.g., accuracy, precision, recall, and F1-score etc.) of the implemented techniques and methods in the proposed and implemented experiments. These were conducted for effective classification estimation and validation – *pre-processing of data*. These matched aggregated distribution of the train-test split techniques and methods together with the resampling methods, which are both semi-supervised and unsupervised classification clustered enabled the optimized resampling of the utilized datasets quantitatively. Specifically, with the categorical distributions mapped against data distributions of the IoT malware binaries and structures, empirical and statistical analysis of the results were scientifically established.

Evidently, captured IoT malware binary classification and analysis using the proposed integrated novel iDRP framework techniques and methods contributed provides treasure trove of hidden information exploited with both 2-D and 3-D images synthesized based on their distinct structural compositions and textures unpacked in comparison to static, dynamic, and automated analysis techniques and methods that are prone to present-day complex and unpacked code obfuscation and time consuming.

6.1. Pre-Processing Component Technique Overview

The data Pre-processing component in the proposed iDRP framework involved the augmentation, transformation, and classification of the raw dataset that are fed and parsed into the network. Feature extraction and mapping of the sparse data feed for layer representation of random features were performed to fully exploit the minuscule information [1] in the converted logged image files for universal approximation (i.e., MD5 Hashes and malware binaries) of the mapped datasets in the IoT ecosystem.

Figure 11 shows the pre-processing component of the proposed iDRP framework and its integration into the overall architecture. The pre-processor 'feed-forward' the systematically processed dataset (i.e., 'fine-grained' gray-scaled converted logged file) directly into the detection component of the proposed iDRP framework to ensure the universal approximation and classification of the precision and accuracy of detection of

anomaly is optimized for high-performance with detailed description for exploits – preparation of dataset for standardization.
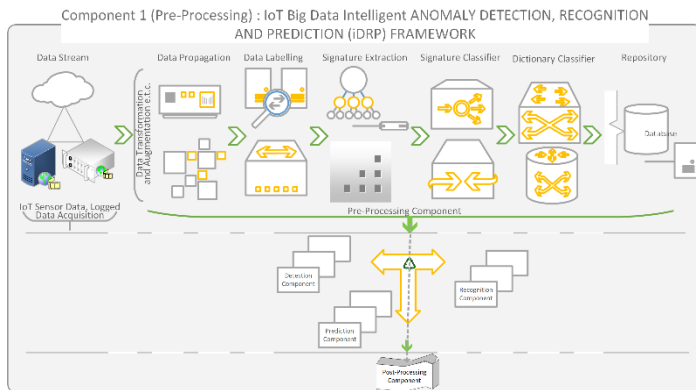
Figure 11: Pre-processing Component of iDRP Framework Overview

The systematic data manipulation techniques such as the systematic extrapolation of feature representations to form the vectorized structures applied in the experiment substantially helped in the distinction of the effective assignment of randomized and differentiated hidden layer weightages synthesized. The feature immutability techniques to stately capture complex code implemented in effect ensures that the feature extractions and data augmentations performed are compact and dynamic to materially eliminate unnecessary redundancies in the generated data. These scalable techniques intrinsically set the approximation numbers of the hidden nodes in the converted logged file of the original input dataset to achieve compact features for exploitations. Uniquely, these innovative and intelligent techniques implemented radically extrapolates rich hidden features for universal approximation competence on both the raw datasets and converted logged file to ensure that the delay in the binary classification analysis is effectively mitigated for optimal outputs in the proposed iDRP framework. Expediently, the compact features extracted and delivered to the next 'following' components iteratively, which is the detection components in the proposed iDRP framework.

6.2. Detection Component Technique Overview

The detection component involves the identification, determination, and classification of converted logged file dataset, whether such dataset is benign or infected. Considering this is a binary classification problem, several techniques [1] were deployed to achieve the expected results by conducting several systematic experimentations to discover the best suitable techniques. In particular, the systematic cross-validation splits, and weightage allocation techniques were beneficial for the classification precision and accuracy in the experiment of the detection component of the proposed iDRP framework.

Figure 12 shows the detection component of the proposed iDRP framework that is strategically the responsible component for determining whether the pre-processed data feed in the network is benign or infected with optimized classification precision and accuracy with maintained high-performance in an IoT ecosystem. Mainly, the mapped converted logged files are systematically rescaled and gray-scaled before they are dynamically fed 'feed forward' into the DNN of the IoT ecosystem.
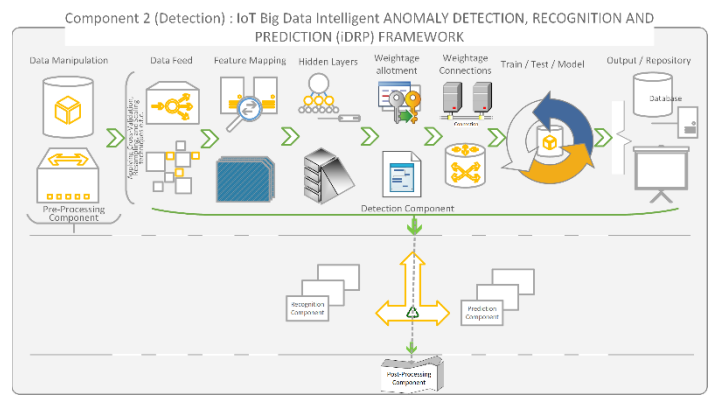
Figure 12: Detection Component of iDRP Framework Overview

The primary focus of the detection component is to intelligently identify the presence of polymorphic malware anomalies [1] with high level of accuracy in the converted logged dataset, which creates the unique separation of concern in its architecture – design principle. The detection component is robustly and innovatively designed in its structure to encapsulate the interaction between the dynamic and static states of the dataset in the converted logged file of an IoT ecosystem. By and large, the detection components feed and receives immutable mapped hidden information dynamically to exploit with other modular components in the proposed iDRP framework. The 'Loose Coupling' and yet cohesion of the detection component [1] essentially promotes the dynamic and robust interaction of the various components directly or indirectly connected to it in the proposed iDRP framework while explicitly performing the primary function of detecting anomalies (randomly and differential mapping hidden features) in the network system.

6.3. Recognition Component Technique Overview

The recognition component is inherently responsible for identifying what class of anomaly the detected malware family is associated with or lack thereof [1]. The recognition component explicitly curates the type of anomalies present if found in the network by precisely defining the exact variants or obfuscation of the IoT malware anomalies for launching attacks such as zero-day attacks. This unique technique involves the heuristic classification of variants of IoT malware anomalies in the network system [1], which are universally approximated in the random and differential feature mapped converted logged file that are rescaled and represented as gray-scale images in an IoT ecosystem.
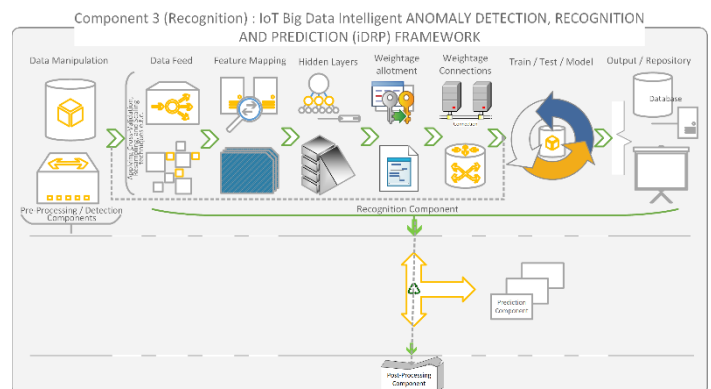
Figure 13: Recognition Component of iDRP Framework Overview

Figure 13 shows the recognition component of the proposed iDRP framework. The recognition component in the architecture of the proposed iDRP framework facilitates the structural distinction of the universally approximated mapped converted logged file in the DNN of the IoT ecosystem. Thereby determining and identifying the particular class of malware anomalies that each existential or inherent malware family belongs to while intelligently discovering new variants or obfuscation of such IoT malware anomalies in the process from the repository that could potentially result in a zero-day attack in an IoT ecosystem. This unique technique provides the capability for the proposed iDRP framework to be able to adaptively and robustly perform the required intelligent classification precision and accuracy [1] of recognizing specific types of anomalies in IoT ecosystem with high performance. The systematic integration of data from both the Pre-Processor and detection components ensures that there is both shared characteristics and cohesion amongst the interconnected components of the proposed iDRP framework. Constructively, the arbitrary boundaries established between the multi-directional connections that are 'Loosely coupled' essentially eliminates constraints in the shared information such as hidden layer features to and from 'fro' the recognition components to the preceding and forwarding components virtually in the proposed iDRP framework – architecture intelligence.

6.4. Prediction Component Technique Overview

The prediction component is practically responsible for predominantly predicting the occurrence of captured malware anomalies in the IoT ecosystem. The systematic elimination of unnecessary '*checkpointing*', '*overlearning*', and '*overfitting*' etc. techniques applied in the implementation of the prediction component effectively helped in the classification precision and accuracy of the predicted malware binaries innovatively in the experimental implementation of the proposed iDRP framework.
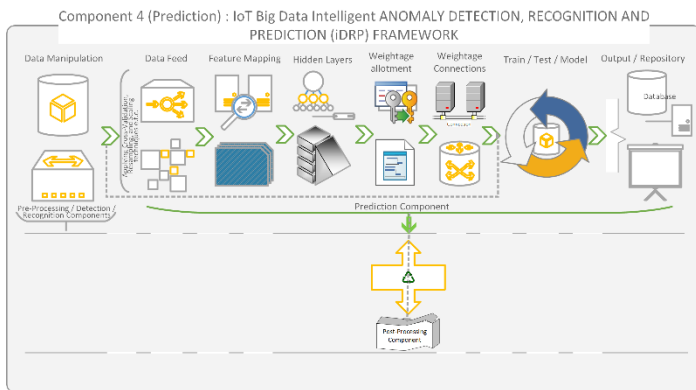
Figure 14: Prediction Component of iDRP Framework Overview

Figure 14 shows the prediction component of the proposed iDRP framework. One of the main features of the prediction components is that it can be configured to predict specialized family of malware binaries that are captured in the IoT ecosystem. Likewise, the systematic assignment of numerical score in the prediction component implementation ensures that the '*learning curve*' of the prediction classification precision and accuracy is properly monitored and dynamically optimized for high-performance. Crucially, the loss or error in the prediction component is systematically calibrated to ensure that the network

is neither '*under-fitting*' nor '*under-learning*' to achieve maximum probabilistic outcomes in detection, recognition and prediction of malware binaries in an IoT ecosystem.

The prediction component employs the network weights to accurately predict the occurrence of malware binaries in the network system. The cyclical and infinite interconnection of the prediction component [1] to other components with accessibility to the repositories in the network ensures that the continuous learning of rapidly evolving malware families is adaptively and intelligently tackled in the growing threats and attacks on the technologically driven IoT ecosystem. In the prediction component, rigorous systematic estimate of the capability of the DNN model were performed exclusively. The prediction component of the proposed iDRP framework is highly effective predominantly due to the capability to systematically stabilize the universal approximation of the mapped hidden features that are fully exploited in the network – intelligent classification prediction technique.

6.5. Post-Processing Component Technique Overview

The Post-Processing component ensures the cyclical and infinite enhancement of the information distributed in the entire proposed iDRP framework [1]. The optimization techniques implemented guarantees that the loss i.e., 'validation loss' in the network model will always be minimized while fostering the continuous and lifelong learning in the network. The substantially reintroduction of the refined information dynamically and innovatively into the network forms the unsupervised building block technique in the overall implementation of the proposed iDRP framework. Broadly, the 'autoencoder' combined with the 'weightage reloading' concepts [1] are integral aspects of the unique features of the Post-Processing component in effectively maximizing the classification precision and accuracy support for the other interconnected components in the proposed iDRP framework.
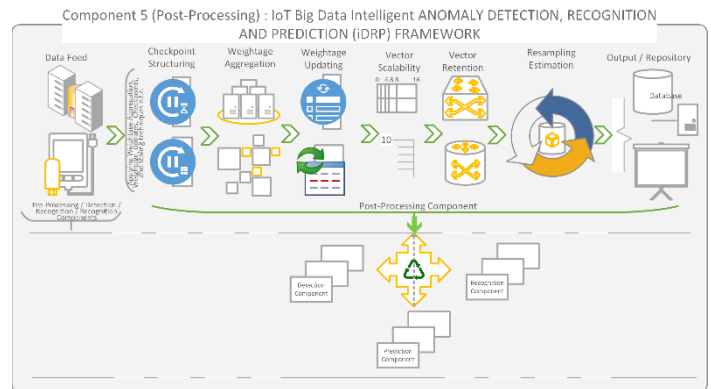
Figure 15: Post-Processing Component of iDRP Framework Overview

Figure 15 shows the Post-Processing component of the proposed iDRP framework. The systematic implementation techniques implemented in the Post-Processing component provide the comparable performance for the convergence of both the training and testing models for stochastic consistencies in the network while storing the models for exploration capabilities in the IoT ecosystem [1]. The Post-Processing component adequately helped stabilize the performance behaviour in the operations of the
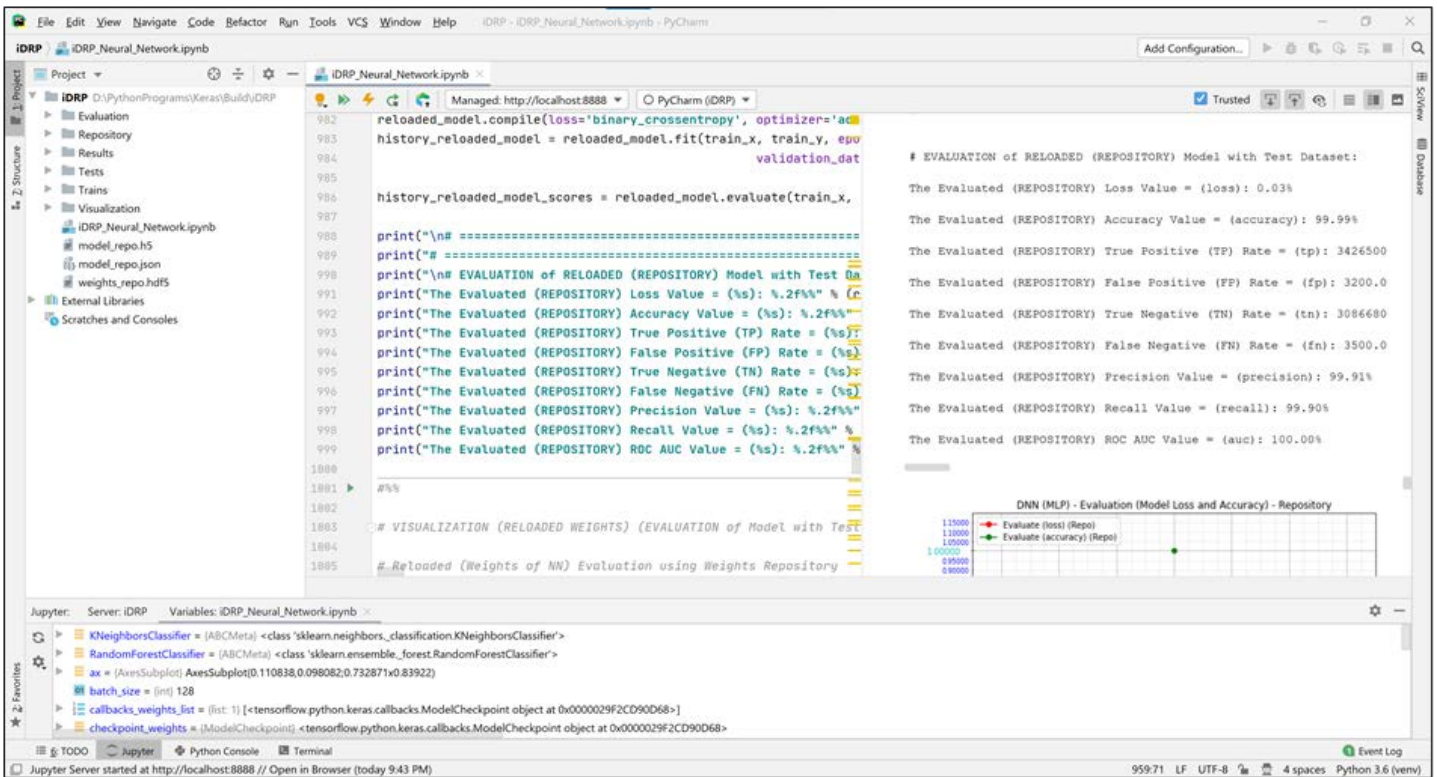
Figure 16: Results of implemented experiment of iDRP Framework with the corresponding snippet of metrics for evaluation

overall proposed iDRP framework via the *checkpoint structuring, weightage aggregation, weightage updating,* and *vector scaling and retention* techniques etc. With minimal 'hyper-parameterization', 'regularization', and 'resampling' techniques etc., the robust and innovative Separation of Concerns technique effectively helped to reduce redundancies in the implementation of the network for optimal competence and high-performance in detecting, recognizing, and predicting malware binaries and anomalies in a massively computerized network system such as an IoT ecosystem.

Specifically, the Post-Processing component is pivotal in the construction of the defined weights of the hidden features exploited for the lifelong learning of the overall proposed iDRP framework. With the Post-Processing component in the architecture of the proposed iDRP framework, minimal 'fine-tuning' is required while the other interconnected components can fully explore compact features in the universally approximated hidden features in the network [1]. Essentially, these specialized innovative and dynamic techniques implemented in the experiment further help to minimize the training, testing and learning time and overhead cost for both the existing and new datasets in the network [1]. The historical information aggregation technique in the network of the proposed iDRP framework are used to systematically discover and generate visualization and analysis of the classification precision and accuracy together with its overall high competency and performance.

6.6. Summary of Components Techniques

The generalized Artificial Neural Network (ANN) applied in the experimental implementation of this proposed iDRP framework is systematic and innovative Multilayer Perceptron (MLP) with Deep Neural Network (DNN) with integrated modular components to expedite the cross-validation methods for universal approximation and classification towards the effective learning, training, and testing capabilities for optimal outcome. The amalgamations of the unique innovative and intelligent random and differential feature extraction, augmentation, transformation, scaling, universal approximation, compactness, mapping, auto encoding, weightage saving, and weightage reloading, vector scaling and retention, cross-validation, resampling etc., techniques were successfully designed and implemented in the interconnected components of proposed iDRP framework to achieve core optimal results. The specialized immutable randomized and differentiated projections techniques implemented in the experiment formed the core of the feature representations of the complex malware binaries for classification precision, accuracy and analysis in the implementation of the proposed iDRP framework – *binary classification problem with complex structure system.*

The '*checkpoints*' of the network structures were set up to save model weights architecture intelligently and automatically whenever there is any improvement attained in the model classification system. The best part of some of the amalgamated specialized techniques implemented is that if there is no improvement in the weight classification of the model, then, the last highest improved weights will be applied in the model. Thereby maximizing the capability to accurately detect, recognize, and predict captured and generated malware binaries anomalies in an IoT ecosystem. The singularity of each of the components in the proposed iDRP framework ensured an easy to manage and maintain network. This generalized Separation of Concern technique ensures that each component in the proposed iDRP framework remains consistent and true to its primary function in the architecture.

The dynamic and innovative abstraction and segregation of the different components the implemented proposed iDRP framework makes it adaptable to the growing threats and attacks in the heavily computerized systems and smart technologies such as an IoT ecosystem. This fundamentally delineates specialized responsibilities amongst the various components in the network with flexible boundaries. Essentially, malware binaries data are analyzed and shared in the interconnected components while at its core separated in execution to perform specified operations in the proposed iDRP framework. The strategy deployed in the experimental implementation of the proposed iDRP framework has demonstrated the intelligent capabilities to determine the slightest of improvements in the network model weightage's structure and only save the best weights and architecture model together for future usage. This simply means that the overall system has been designed to continually learn, adapt, improve and perform better with time in combating the ever-growing malware threats such as the prevalent polymorphic IoT malware threats and attacks in the IoT ecosystem.

The overall procedures involve (a). pre-processing of dataset, (b). loading of dataset, (c). augmentation of dataset (d). creating of model architecture (e). compiling of model (f). fitting of model (g). evaluation of model (h). saving of model weights (i). loading of saved model weights (j). evaluation of saved model (k). post-processing of outcome. These approaches were conducted in multiple hidden layers Neural Network 'feed-forward' propagation and cyclic method using Deep Learning (DL) models to precisely, efficiently and optimally and intelligently detect, recognize, and predict (iDRP) anomalies in an IoT ecosystem. Cumulatively, several ANN techniques were leveraged innovatively and intelligently to significantly improve the classification precision, accuracy, and performance in the implementation of the proposed iDRP framework. Notably, the dynamic and innovative extensibility and maintainability of the proposed iDRP framework has made it highly adaptable and adoptable in effectively safeguarding and protecting critical infrastructures in multi-industries that depends on heavily computerized interconnected network systems and smart technologies such as an IoT ecosystem in the digital space.

7. Results, Observations and Discussions

A 100% accuracy rate, 100% ROC/AUC rate together with a precision rate of 99.94%, and recall rate of 99.90%, and merely a loss rate of 0.02% were achieved in the preliminary implementation of the iDRP framework experiment. Please see Figure 16 screenshot for the output of results and metrics. These results were attained based on the syntheses of the various techniques and procedures applied and discussed in the previous sections.

In like manner, a 99.98% accuracy rate, 99.99% ROC/AUC ROC/AUC rate together with a precision rate of 99.95%, and recall rate of 99.93 %, and a mere loss rate of 0.03% were achieved in the actual implementation of the integrated iDRP framework experiment using real IoT malware dataset – BoTNetIoT dataset.

The metrics used in evaluating the experimental implementation of the proposed iDRP framework are: Accuracy rate, Precision rate, Recall rate, Receiver Operator Characteristic (ROC) and Area Under the Curve (AUC) rate, Loss rate, True

Positive (TP) rate, False Positive (FP) rate, True Negative (TN) rate, False Negative (FN), and F1-Score respectively.
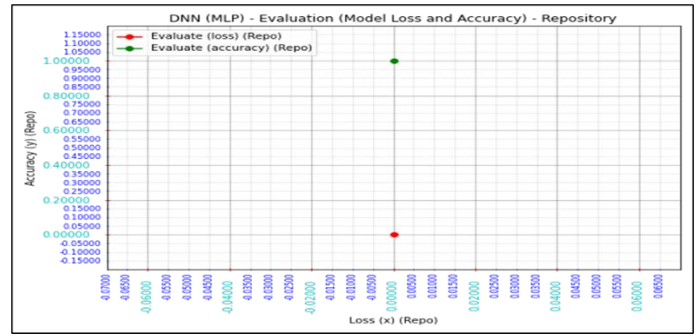
Figure 17: Graph of Accuracy and Loss Results of implemented experiment of proposed iDRP Framework

Figure 17 shows the result of the evaluated models of the experimental implementation together with the corresponding rates and scores. The 100% accuracy value, 100% ROC/AUC value with a precision value of 99.90%, and crucially with a much lower False Positive (FP) rate. that were attained in the experimental implementation of the proposed iDRP framework shows that there is a correlation in the accuracy, ROC/AUC, and precision rate in the Deep Learning (DL) classifiers in tandem with the hypothesis of the research study. It further proves that the experimental implementation of the proposed iDRP framework in this research study is highly innovative, applicable, useful, relevant and efficient; accurate, precise, robust, dynamic, adaptable, and scalable in detecting, recognizing, and predicting anomalies in the form of polymorphic malwares in an IoT ecosystem.

The visualization of the results of the models evaluated were represented in plotted Deep Neural Network graphs and charts to gain better insights and understanding of the complex binary classification matrix problems and solutions in the experimental implementation of the proposed iDRP framework. Figure 10 displays the optimal efficiency and accuracy correspondingly with insignificant loss rates in relations to the performance of the evaluated model for detecting, recognizing, and predicting the anomalies in an IoT ecosystem. Minimal training was required.

Logically, it is understandably that 50 epochs would likely fare better than a 5 epochs of model training and testing of datasets. At the same time, with a direct comparison of the regular 50 epochs with the 50 epochs' '*Checkpointing*' Deep Neural Network model technique of intelligently and systematically saving and reloading the best performance weightage in the repository of the overall system, a 100% accuracy rate and a 100% ROC/AUC rate together with a 99.94% precision rate and 99.90% recall rate of detection, recognition, and prediction of the anomalies in an IoT ecosystem was achieved in the experiment.

7.1. Cross-Validation and Evaluation Performance of Predictor Model Results

The results of the various output attained in the experiment on the proposed iDRP framework network has been reported in Table 3. The output of the different techniques towards the innovative synthesis, augmentations permutations, adjustments, and tuning of parameters in the Neural Networks together with various optimization techniques in the experiment has been extensively

investigated and compared to show the selected outcome in the project.

Table 3: Comparison of the Experimental Results

Cross-Validation Classification Parameter Tuning Metrics of iDRP (MLP) Model using *Keras* Library with Tensorflow Backend

Dataset (MNIST)	NN Layers	Epoch (Iterations)	Optimizer	Learning Rate	TPR	FPR	Precision	Recall	MAE	F-Score	Accuracy (%)	Training Time (s)
✓	5	5	ADAM	0.02	0.924	0.924	0.924	0.924	0.200	0.947	92.43	1.00
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
✓	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
✓	50	50	ADAM	0.02	0.996	0.996	0.996	0.996	0.159	0.979	99.60	10.00
✓	50	50	ADAM	0.01	0.999	0.998	0.999	0.998	0.010	0.998	100.00	0.03

Legend: MLP=**Multilayer Perceptron**, NN=**Neural Network**, Learning Rate=**Dropout Rate**, TPR=**True Positive Rate**, FPR=**False Positive Rate**, MAE=**Mean Absolute Error**, F-Score=**Measure of Test Accuracy**

Table 3 displays the comparison in the various parameter tuning and optimizations performed in the experiments together with their corresponding metrics and results. It is evident that the continuous tuning process has significant impact on the overall output, especially with the classification precision and accuracy of the predictors. Likewise, the optimization parameter ADAM plays a vital role in minimizing loss function while increasing the speed of the convergence of the Neurons in the system. The dynamic and cyclic hyperparameter tuning approach in the Neural Network construct of this experiment will ultimately facilitate an incremental and better outcome.

7.2. Comparison of Generalized Artificial Neural Network (ANN) Results

The selection of the appropriate NN model for the implementation of the proposed iDRP framework involved the systematic experimentations of various streamlined and focused ANNs such as Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Multilayer Perceptron (MLP), and Hybrid Network (synergy of RNN and CNN) to effectively determine the best possible NN model for the binary classification problem of this nature.

Table 4: Comparison of Artificial Neural Network (ANN) - RNN, CNN, MLP, and Hybrid Models Results

No.	Metric / ANN	RNN	CNN	MLP	Hybrid (RNN + CNN)
1	Detection	98.53%	99.71%	**100.00%**	99.83%
2	Recognition	98.24%	99.64%	**100.00%**	99.84%
3	Prediction	97.32%	99.52%	**100.00%**	98.91%

Table 4 shows the comparison between the various NNs in determining the chosen NN for the implementation of the proposed iDRP framework. Evidently, MLP performed best for the binary classification of precision and accuracy in the experiment – binary classification problem.

The MLP NN mainly performed better with model flexibility in terms of training, testing, and development. The optimal capability for an MLP NN to be abstracted in its architecture design and robustly extract fine-grained features with universal approximation made it a perfect candidate for the implementation NN model of the proposed iDRP framework. Chiefly, the MLP NN allows the Separation of Concerns in its architecture by creating boundaries (methods, and objects) for the several interconnected components in the proposed iDRP framework while providing the

layers and tiers the independent capability for each component to maintain their core operations dynamically.

7.3. Comparison of Classification Techniques on Standardized Datasets Results – Preliminary Experiment

The amalgamations of innovative MLP techniques applied in the implementation of the proposed iDRP framework network demonstrated to be advantageous in the specialized classification precision and accuracy on accurately detecting, recognizing, and predicting malware binaries in the standardized dataset in an IoT ecosystem.

Table 5 shows the comparison of the different classification results of experiments on standardized datasets. The synthesized innovative techniques in the proposed iDRP framework network technique that have been successfully implemented have proven to be more efficient and competent in tackling polymorphic IoT malware attacks and their variants or obfuscation that are typically responsible for zero-day attacks in a heavily computerized network system and smart technologies such as the IoT ecosystem. The learning capabilities and accuracy of a specific predictor problem of this nature was significantly amplified through the innovative classification precision and accuracy techniques that were proposed and implemented in the systematic experiments iDRP framework network.

Table 5: Comparison of Classification Experimentations Techniques with Standardized Datasets and Results – Preliminary Experiment

No.	Technique / Dataset	MNIST Accuracy (%)	NORB Accuracy (%)
1	SAE	98.60	86.28
2	SDA	98.72	87.62
3	DBN	98.87	88.47
4	DBM	99.05	89.65
5	MLP-BP	97.39	84.20
6	ML-ELM	99.04	88.91
7	H-ELM	99.13	91.28
8	iDRP	**100.00**	**100.00**

Key: SAE (Stacked Auto Encoders), SDA (Stacked Autoencoders), DBN (Deep Belief Networks), DBM (Deep Boltzmann Machines), MLP-BP (Multilayer Perceptron Back Propagation), ML-ELM (Machine Learning Extreme Learning Machine, H-ELM (Hierarchical Extreme Learning Machine), iDRP (Intelligent Detection Recognition Prediction)

7.4. Comparison of Classification Techniques on Standardized Datasets (BoTNetIoT) Results

The standardized IoT malware generic traffic metric and evaluation were used as parameters to compare the effectiveness of the learning and training techniques in correlations with the validation of the techniques applied to detecting benign and malicious network traffic in an IoT ecosystem. Table 6 shows the attained results.

Table 6 indicates that the proposed novel iDRP framework techniques and methods perform better with other notable techniques for IoT malware detections. Significantly, the optimal results were achieved with the synthesis of both the discriminative features and differential evolution techniques for the universal approximation of the extrapolated hidden information together with connected structures in the integrated iDRP framework.

Table 6: Comparison of Classification Techniques on Standardized IoT Malware Datasets (BoTNetIoT) Results

No.	Techniques	Accuracy	Precision	Recall	F-score
1	Resnet34 [41]	92.39%	93.57%	64.55%	76.40%
2	Resnet50 [41]	94.50%	95.78%	94.02%	94.90%
3	MobileNet [41]	91.32%	91.67%	91.03%	91.35%
4	SOINN [41]	91.75%	89.68%	95.52%	92.50%
5	iDRP	**99.98%**	**99.95%**	**99.93%**	**99.97%**

Evidently, the innovative iDRP framework network techniques performed better with even the state-of-the-art MLP NN techniques and methods etc. The flexibility, extensibility, stability, and maintainability in the architecture design of the innovative iDRP framework network that were developed have demonstrated to be optimal in helping achieve the overall outcome. Through the leveraged innovative and systematic techniques implemented in the experiments of the proposed iDRP framework network, it is safe to mention that the optimal security and protection of heavily computerized network systems and smart technologies such as an IoT ecosystem against evolving attacks and threats can be better managed with improved innovation and intelligence in the digital ecosystem.

8. Conclusion

The novelty of this work are the systematic integrations, techniques and methods of the adaptive and scalable *pre-processing* and *post-processing* components designed and developed for multi-class classifications and capturing the immutable vector state structures from categorical distributions to mapped data distributions in order to estimate and exploit hidden textural features of converted polymorphic and new IoT malware binary images in correlations with their aggregated distributions in a network. Together with, the modular *detection*, *recognition*, and *prediction* components DL approach provided new concepts through the syntheses of the three (3) key components with the *multi-class classification*, *data augmentation*, *vector scaling and retention*, *clustering*, *cross-validation*, and *resampling* techniques and methods etc., synthesized and interconnected evolutionarily by the *post-processing* techniques in the proposed novel iDRP framework, which can be multidisciplinary. In particular, considering the textural and structural changes in polymorphic and new malware variants or obfuscation, the applied techniques and methods that were designed and developed ensured the improved and accurate *detection*, *recognition*, and *prediction* of the complex code malwares that specifically changes their binary structures on execution in an IoT ecosystem. A higher level of feature extraction and classification estimation of the modular hidden layer information were achieved to intelligently and efficiently detect the present-day obfuscation of malware attacks and infections in an IoT ecosystem. Overall, the derived intelligent immutability of the stately vectorized mapped data estimation techniques and methods implemented ensured optimal results in the experiments.

We conducted and tested the proposed implemented experiments with sound scientific procedures and processes to provide the proof-of-concept by conceptualizing and applying conventional scientific processes such as conducted in the engineering, medical, animal science fields etc., to establish empirical and statistical analysis in the performed experiments i.e., structurally complex, yet similar vector images; MNIST (2-D) and NORB (3-D) datasets prior to the comparative real IoT malware (BoTNetIoT) dataset converted image files as benchmark, which is analogous to animal to human experiments in the pure science and engineering fields systematically. The premise of the research experiments is to discover conclusive scientific approaches, techniques, and methods to solving polymorphic and new malware anomalies problems in an IoT ecosystem. These were achieved by applying the aforementioned procedures and techniques to solving the IoT malware problems with new ideas in this work. Extensive analysis was conducted by synthesizing the various techniques on the big datasets utilized in the proposed experiments. The syntheses of the integrated procedures and techniques in the proposed novel iDRP framework, which consists of five (5) components; *pre-processor*, *detection*, *recognition*, *prediction*, and *post-processor* components provided the seamless evolutionary information distributions for a high-level accuracy result in the experiment. Standardized metrics such as *accuracy* rate, *ROC/AUC* rate, *precision* rate, *True Positive* (TP) rate, *False Positive* (FP) rate, *True Negative* (TN) rate, *False Negative* (FN) rate, *recall* rate, and *F1-Score* were calculated and recorded to effectively determine the performance of the finalized generative DNN models in the proposed and implemented systematic experiments of the iDRP framework research.

In the end, the resultant output of the proposed experiments was tested and evaluated to achieve the best performance and consistency, accuracy, and precision etc., of the comparative aggregated model results achieved. As a result, a 100% accuracy rate, 100% ROC/AUC rate, 99.94% precision rate, and 99.90% recall rate were successfully achieved with the preliminary experiment, while with a real IoT malware data, BoTNetIoT, a 99.98% accuracy rate, 99.99% ROC/AUC rate, 99.95% precision rate, and 99.93 recall rate were successfully attained as solution with the synthesized techniques and methods to the regression and binary classification problems for intelligent malware anomalies detection in an IoT ecosystem.

Future Work

The future work would be to evaluate the systematic '*checkpointing*' of various weightages and best performing weights measurement together with the corresponding network topography and architecture in IoT security models, which will be verified, validated and explored further for better security solution in expanding heavily computerized interconnected systems.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1]   O. Eboya, J. Juremi and M. Shahpasand, "An Intelligent Framework for Malware Detection in Internet of Things (IoT) Ecosystem," in 2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC), 1 - 6, 2020, doi: 10.1109/r10-htc49770.2020.9356961.

[2] R. AL MOGBIL, M. AL ASQAH and S. EL KHEDIRI, "IoT: Security Challenges and Issues of Smart Homes/Cities," in 2020 International Conference on Computing and Information Technology (ICCIT-1441), 1 - 6, 2020, doi: 10.1109/iccit-144147971.2020.9213827.

[3] E. Tabane and T. Zuva, "Is there a room for security and privacy in IoT?," in 2016 International Conference on Advances in Computing and Communication Engineering (ICACCE), 1-6, 2016, doi: 10.1109/icacce.2016.8073758.

[4] C. Vorakulpipat, E. Rattanalerdnusorn, P. Thaenkaew and H. Dang Hai, "Recent challenges, trends, and concerns related to IoT security: An evolutionary study," in 2018 20th International Conference on Advanced Communication Technology (ICACT), 1 - 6, 2018, doi: 10.23919/icact.2018.8323773.

[5] Z. Zheng, S. Xie, H. Dai, X. Chen and H. Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," in 2017 IEEE International Congress on Big Data (BigData Congress), 1 - 6, 2017, doi: 10.1109/bigdatacongress.2017.85.

[6] N. Neshenko, E. Bou-Harb, J. Crichigno, G. Kaddoum and N. Ghani, "Demystifying IoT Security: An Exhaustive Survey on IoT Vulnerabilities and a First Empirical Look on Internet-Scale IoT Exploitations," in 2019 IEEE Communications Surveys & Tutorials, 2702-2733, 2019, doi: 10.1109/comst.2019.2910750.

[7] Y. Seralathan et al., "IoT security vulnerability: A case study of a Web camera," in 2018 20th International Conference on Advanced Communication Technology (ICACT), 1 - 6, 2018, doi: 10.23919/icact.2018.8323685.

[8] N. Karie, N. Sahri and P. Haskell-Dowland, "IoT Threat Detection Advances, Challenges and Future Directions," in 2020 Workshop on Emerging Technologies for Security in IoT (ETSecIoT), 1 - 6, 2020, doi: 10.1109/etseciot50046.2020.00009.

[9] O. Georgiana Dorobantu and S. Halunga, "Security threats in IoT," in 2020 International Symposium on Electronics and Telecommunications (ISETC), 1 - 6, 2020, doi: 10.1109/isetc50328.2020.9301127.

[10] M. Antonakakis et al., "Understanding the mirai botnet," 1st ed. Vancouver, BC, Canada: Proceedings of the 26th USENIX Security Symposium, USENIX Association, 2017.

[11] K. Sha, R. Errabelly, W. Wei, T. Yang and Z. Wang, "EdgeSec: Design of an Edge Layer Security Service to Enhance IoT Security," in 2017 IEEE 1st International Conference on Fog and Edge Computing (ICFEC), 1 - 6, 2017, doi: 10.1109/icfec.2017.7.

[12] I. Waz, M. Sobh and A. Bahaa-Eldin, "Internet of Things (IoT) security platforms," in 2017 12th International Conference on Computer Engineering and Systems (ICCES), 1 - 5, 2017, doi: 10.1109/icces.2017.8275359.

[13] W. Iqbal, H. Abbas, M. Daneshmand, B. Rauf and Y. Bangash, "An In-Depth Analysis of IoT Security Requirements, Challenges, and Their Countermeasures via Software-Defined Security," in IEEE Internet of Things Journal, 10250-10276, 2020, doi: 10.1109/jiot.2020.2997651.

[14] E. Baccelli et al., "RIOT: An Open Source Operating System for Low-End Embedded Devices in the IoT," in IEEE Internet of Things Journal, 4428-4440, 2018, doi: 10.1109/jiot.2018.2815038.

[15] R. Johari, N. Gaurav, S. Chaudhary and A. Pramanik, "START: Smart Stick based on TLC Algorithm in IoT Network for Visually Challenged Persons," in 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 1 - 6, 2020, doi: 10.1109/i-smac49090.2020.9243517.

[16] A. Gupta and R. Johari, "IOT based Electrical Device Surveillance and Control System," in 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), 1 - 5, 2019, doi: 10.1109/iot-siu.2019.8777342.

[17] L. Tawalbeh, H. Tawalbeh, H. Song and Y. Jararweh, "Intrusion and attacks over mobile networks and cloud health systems," in 2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 1 - 6, 2017, doi: 10.1109/infcomw.2017.8116345.

[18] Y. Zheng, A. Pal, S. Abuadbba, S. Pokhrel, S. Nepal and H. Janicke, "Towards IoT Security Automation and Orchestration," in 2020 Second IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA), 1 - 6, 2020, doi: 10.1109/tps-isa50397.2020.00018.

[19] M. Al-Garadi, A. Mohamed, A. Al-Ali, X. Du, I. Ali and M. Guizani, "A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security," in IEEE Communications Surveys & Tutorials, 1646-1685, 2020, doi: 10.1109/comst.2020.2988293.

[20] F. Hussain, R. Hussain, S. Hassan and E. Hossain, "Machine Learning in IoT Security: Current Solutions and Future Challenges," in IEEE Communications Surveys & Tutorials, 1686-1721, 2020, doi: 10.1109/comst.2020.2986444.

[21] M. Bettayeb, O. Waraga, M. Talib, Q. Nasir and O. Einea, "IoT Testbed Security: Smart Socket and Smart Thermostat," in 2019 IEEE Conference on Application, Information and Network Security (AINS), 1 - 6, 2019, doi: 10.1109/ains47559.2019.8968694.

[22] J. Moos, "IoT, Malware and Security," in 2017 ITNOW, 28-29, 2017, doi: 10.1093/itnow/bwx013.

[23] A. Zahra and M. Shah, "IoT based ransomware growth rate evaluation and detection using command and control blacklisting," in 2017 23rd International Conference on Automation and Computing (ICAC), 1 - 5, 2017, doi: 10.23919/iconac.2017.8082013.

[24] E. Ronen, A. Shamir, A. Weingarten and C. O'Flynn, "IoT Goes Nuclear: Creating a Zigbee Chain Reaction," in 2018 IEEE Security & Privacy, 54-62, 2018, doi: 10.1109/msp.2018.1331033.

[25] S. Sonune, D. Kalbande, A. Yeole and S. Oak, "Issues in IoT healthcare platforms: A critical study and review," in 2017 International Conference on Intelligent Computing and Control (I2C2), 1 - 5, 2017, doi: 10.1109/i2c2.2017.8321898.

[26] M. Mekala and P. Viswanathan, "A Survey: Smart agriculture IoT with cloud computing," in 2017 International conference on Microelectronic Devices, Circuits and Systems (ICMDCS), 1 - 6, 2017, doi: 10.1109/icmdcs.2017.8211551.

[27] P. Rughoobur and L. Nagowah, "A lightweight replay attack detection framework for battery depended IoT devices designed for healthcare," in 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), 1 - 7, 2017, doi: 10.1109/ictus.2017.8286118.

[28] A. Ukil, S. Bandyoapdhyay, C. Puri and A. Pal, "IoT Healthcare Analytics: The Importance of Anomaly Detection," in 2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), 1 - 4, 2016, doi: 10.1109/aina.2016.158.

[29] S. Venkatraman and M. Alazab, "Classification of Malware Using Visualisation of Similarity Matrices," in 2017 Cybersecurity and Cyberforensics Conference (CCC), 1 - 5, 2017, doi: 10.1109/ccc.2017.11.

[30] J. Tang, C. Deng and G. Huang, "Extreme Learning Machine for Multilayer Perceptron," in 2016 IEEE Transactions on Neural Networks and Learning Systems, 809-821, 2016, doi: 10.1109/tnnls.2015.2424995.

[31] Y. Zeng, L. Qian and J. Ren, "Evolutionary Hierarchical Sparse Extreme Learning Autoencoder Network for Object Recognition," in 2018 Symmetry, 1 - 11, 2018, doi: 10.3390/sym10100474.

[32] M. Shahini, R. Farhanian and M. Ellis, "Machine Learning to Predict the Likelihood of a Personal Computer to Be Infected with Malware," in 2019 SMU Data Science Review, 1 - 24, 2019, doi: https://scholar.smu.edu/datasciencereview/vol2/iss2/9/.

[33] J. Villanueva, R. Juanatas and L. Lacatan, "Malware Predictor using Machine Learning Techniques," in 2020 Research Gate, 5665 - 5674, 2020, doi: https://www.researchgate.net/publication/339935591_Malware_Predictor_using_Machine_Learning_Techniques.

[34] A. Onal, O. Berat Sezer, M. Ozbayoglu and E. Dogdu, "Weather data analysis and sensor fault detection using an extended IoT framework with semantics, big data, and machine learning," in 2017 IEEE International Conference on Big Data (Big Data), 2037-2046, 2017, doi: 10.1109/bigdata.2017.8258150.

[35] J. Wang, C. Liu, M. Zhu, P. Guo and Y. Hu, "Sensor Data Based System-Level Anomaly Prediction for Smart Manufacturing," in 2018 IEEE International Congress on Big Data (BigData Congress), 158-165, 2018, doi: 10.1109/bigdatacongress.2018.00028.

[36] N. Zeeshan, M. Reed and Z. Siddiqui, "Three-way Security Framework for Cloud based IoT Network," in 2019 International Conference on Computing, Electronics & Communications Engineering (iCCECE), 183-186, 2019, doi: 10.1109/iccece46942.2019.8941877.

[37] B. Li, X. Ming and G. Li, "Big data analytics platform for flight safety monitoring," in 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), 1 - 3, 2017, doi: 10.1109/icbda.2017.8078837.

[38] Y. Xing, H. Shu, H. Zhao, D. Li and L. Guo, "Survey on Botnet Detection Techniques: Classification, Methods, and Evaluation," in 2021 Mathematical Problems in Engineering, 1-24, 2021, doi: 10.1155/2021/6640499.

[39] Y. Song, S. Hyun and Y. Cheong, "Analysis of Autoencoders for Network Intrusion Detection," in 2021 Sensors, 4294, 2021, doi: 10.3390/s21134294.

[40] T. Palla and S. Tayeb, "Intelligent Mirai Malware Detection for IoT Nodes," in 2021 Electronics, 1241, 2021, doi: 10.3390/electronics10111241.

[41] G. Bendiab, S. Shiaeles, A. Alruban and N. Kolokotronis, "IoT Malware Network Traffic Classification using Visual Representation and Deep Learning," in 2020 6th IEEE Conference on Network Softwarization (NetSoft), 1 - 9, 2020, doi: 10.1109/netsoft48620.2020.9165381.

[42] R. Kozik, M. Pawlicki and M. Choraś, "A new method of hybrid time window embedding with transformer-based traffic data classification in IoT-networked environment," in 2021 Springer, 1 - 9, 2021, doi: 10.1007/s10044-021-00980-2.

[43] A. Alhowaide, I. Alsmadi and J. Tang, "Towards the design of real-time autonomous IoT NIDS," in 2021 Cluster Computing, 1 - 14, 2021, doi: 10.1007/s10586-021-03231-5.

[44] K. Sudheera, D. Divakaran, R. Singh and M. Gurusamy, "ADEPT: Detection and Identification of Correlated Attack Stages in IoT Networks," in 2021 IEEE Internet of Things Journal, 6591-6607, 2021, doi: 10.1109/jiot.2021.3055937.

[45] Q. Ngo, H. Nguyen, V. Le and D. Nguyen, "A survey of IoT malware and detection methods based on static features," in 2020 ICT Express, 280-286, 2020, doi: 10.1016/j.icte.2020.04.005.

[46] P. Sudhakaran, C. Malathy, T. Vardhan and T. Sainadh, "Detection of Malware from IOT Devices Using Deep Learning Techniques," in 2021 Journal of Physics: Conference Series, 1 - 7, 2021, doi: 10.1088/1742-6596/1818/1/012219.

[47] B. Khammas, "The Performance of IoT Malware Detection Technique Using Feature Selection and Feature Reduction in Fog Layer," in 2020 IOP Conference Series: Materials Science and Engineering, 1 - 11, 2020, doi: 10.1088/1757-899x/928/2/022047.

[48] K. Han, J. Lim, B. Kang and E. Im, "Malware analysis using visualized images and entropy graphs," in 2015 International Journal of Information Security, 1-14, 2015, doi: 10.1007/s10207-014-0242-0.

Extraction of Psychological Symptoms and Instantaneous Respiratory Frequency as Indicators of Internet Addiction Using Rule-Based Machine Learning

Hung-Ming Chi[1], Liang-Yu Chen[2], Tzu-Chien Hsiao[1,3,*]

[1]*Department of Computer Science, College of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, 300, Taiwan*

[2]*Institute of Computer Science and Engineering, College of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, 300, Taiwan*

[3]*Institute of Biomedical Engineering, College of Electrical and Computer Engineering, National Yang Ming Chiao Tung University, Hsinchu, 300, Taiwan*

ARTICLE INFO	ABSTRACT
	Internet addiction (IA) has adverse effects on psychophysiological responses, interpersonal relationships, and academic and occupational performance. IA detection has received increasing attention. Although questionnaires enable long-term assessment (over 6 months) and physiological measurements to aid the short-term evaluation (over 2 min) of IA, the lack of algorithms results in an inability to detect IA in real time. A computer-aided system can address this problem. This study used the extended classifier system with continuous real-coded variables (XCSR) for rule-based machine learning to classify IA risk. Chen Internet Addiction Scale (CIAS) items were verified and instantaneous respiratory features of IA were extracted with "don't care" attribute values. The result demonstrated that the XCSR model achieved more than 95% classification accuracy. Using the "don't care" attribute values, the CIAS items were reduced from 26 to 19, and the instantaneous frequency (IF) of respiratory muscle contractions, respiratory wall movements, and body movements were extracted as IA-related features. These findings suggested that the XCSR model is a potentially useful system for detecting IA. The modified 19-item CIAS and IF of respiration can be adopted to assist in the real-time detection of IA and explore the psychophysiological developments of IA users. In future studies, more samples must be collected to validate these findings and instantaneous physiological responses investigated with different window sizes while participants with IA engage in active online gameplay.

1. Introduction

This paper constitutes a continuation of the work originally presented at the 2019 Genetic and Evolutionary Computation Conference [1]. The Internet has changed our daily lives profoundly over the last three decades. People now have easy access to the Internet by using desktops, smartphones, tablets, laptops, and other digital devices, through which they interact with friends, purchase products, play games, search for information, and send emails. Despite the convenience the Internet lends to our lives, some serious concerns remain.

Some individuals intensely indulge in and depend on the Internet, which can lead to negative ramifications, such as estrangement from family and friends, impaired mental and physical health, and low academic and work performance [2, 3]. This condition has been termed Internet addiction (IA) [4], and it has various subtypes, including problematic online game addiction [5], problematic cellphone use [6], and social network mental disorder [7]. Scholars have reported that problematic game use in particular was associated with severe clinical impairment, and Internet gaming disorder (IGD) has been preliminarily listed in the *Diagnostic and Statistical Manual of Mental Disorders (DSM), Fifth Edition* [8]. A 2019 announcement by the World Health Organization recognized gaming disorder as a mental disorder [9]. The IA prevalence rate in Asia [3], Europe [10], and the United States [11] was 5.6%, 7.97%, and 8.1%, respectively, with the risk of developing IA increasing as a result of the 2019 coronavirus

disease pandemic [12, 13]. IA is a common behavioral addiction globally and continues to receive considerable attention.

Psychological symptoms and physiological responses are key indicators for detecting addiction behavior. IA-related questionnaires are used to evaluate the symptoms of IA, but they have some limitations. First, different questionnaires must be designed that correspond to various cultures and environments worldwide. The question of how many psychological symptoms must be incorporated into an IA questionnaire remains debated. Second, questionnaires are typically adopted to assess Internet experiences over a long-term period (over 6 months) [2, 3, 5, 8, 9, 14–18] and are not ideal for examining short-term or current experiences. Third, Internet users must spend time completing these questionnaires. Fourth, subjective experiences can affect the assessment of IA. Researchers can potentially observe the physiological responses of Internet users alongside questionnaires to assist in the assessment of IA risk. In such assessments, the major physiological signals are analyzed or calculated using descriptive statistics or generic mathematics [10, 19–23], although these algorithms may be insufficient to estimate the dynamical and instantaneous nature of physiological responses. Some studies have collected user behavior or social media and online gaming records to input into machine learning systems to identify addiction behaviors [7, 24–28]. However, this type of data collection is time consuming, and the real-time detection of IA presents a considerable challenge.

The extended classifier system (XCS) method is a rule-based machine learning system that assumes an agent role to interact with the environment and produce the optimum policy [29]. The XCS method generates rules with each interaction, each in the form of an "IF condition THEN action." The conditions in the rules employ coded symbols {0, 1, #}; the symbol # involves ignoring some values and is also referred to as the "don't care" attribute. The XCS method is applied to manage binary problems, and when combined with the continuous real-coded variables (XCSR) method, it can solve real-value problems [30]. Unlike other machine learning systems, the learning process of the XCSR method can be observed directly, and the "don't care" attribute can extract the main features for classification. Therefore, this study adopted the XCSR method as a computer-aided system to detect the risk of IA using Chen Internet Addiction Scale (CIAS) items [3] as input data. The XCSR method was then applied to verify the major psychological symptoms of IA from the CIAS data. To investigate the instantaneous physiological responses of IA, the ensemble empirical mode decomposition (EEMD) method [31] was used to decompose a physiological signal into instantaneous components, and the XCSR method was employed to extract the IA-related physiological features from these components. This study has two main contributions, namely the use of a rule-based machine learning system to classify the risk of IA, and the extraction of the major symptoms and instantaneous physiological features of IA from a questionnaire and from physiological components, respectively. These developments may be used to design a new short-form IA questionnaire and instantaneous physiological model to detect IA in real time. These extracted psychological symptoms and instantaneous responses can assist in exploring the developmental process of IA behavior.

The organization of this paper is as follows: The Literature Review section outlines the psychological symptoms covered in the IA questionnaire, physiological responses of the IA group, and detection of behavioral addiction using machine learning systems. The Research Methodology section presents the overview of the XCSR and EEMD methods, and the Experiment section describes the design of experiment and data analysis. The Results section reveals the classification accuracy of the XCSR using the CIAS items and instantaneous physiological waveforms as input data, with the rate of non-"don't care" values calculated for these input data. In the Discussion section, the study findings and limitations are discussed, and the Conclusion section summarizes the study.

2. Literature Review

2.1. IA questionnaires

The design of IA questionnaires is based on substance abuse, substance dependence, and pathological gambling criteria from the *DSM, Fourth Edition*, IGD criteria from the *DSM, Fifth Edition*, and gaming disorder criteria from the *International Classification of Diseases, Eleventh Revision*. IA questionnaires cover various psychological symptoms (Table 1), including abandoning other activities (S_A), craving (S_C), compulsive symptoms (S_{CS}), deceit or cover-up (S_{DC}), health-related problems (S_H), interpersonal problems (S_I), mood modification (S_{MM}), preoccupation or salience (S_{PS}), school or occupational problems (S_{SO}), tolerance (S_T), time management problems (S_{TM}), and withdrawal symptoms (S_{WS}). Researchers have debated how many psychological symptoms can be incorporated into an IA questionnaire. Young presented an eight-item IA diagnostic questionnaire that included S_{CS}, S_{DC}, S_I/S_{SO}, S_{MM}, S_{PS}, S_{TM}, S_T, and S_{WS} [2]. Griffiths proposed that S_I/S_H, S_{MM}, S_{CS}, S_{PS}, S_T, and S_{WS} were the core components of addiction [32]. The IGD questionnaire comprises S_A, S_{CS}, S_{DC}, S_I, S_{MM}, S_{PS}, S_{SO}, S_T, and S_{WS} [14], and the CIAS consists of S_{CS}, S_I/S_H, S_T, S_{TM}, and S_{WS} [3]. Authors suggested that S_I/S_H, S_{MM}, S_{CS}, S_{SO}, S_T, S_{PS}, and S_{WS} be used to design a questionnaire [5]. Authors structured their IA questionnaire by using the psychological symptoms of S_A, S_C, S_{CS}, S_{DC}, S_H/S_{SO}, S_I, S_{MM}, S_{PS}, S_T, and S_{WS} [9].

IA questionnaires are typically translated into various languages or modified into suitable structures for application in different countries or areas. For example, Young's IA Test is a well-known questionnaire for assessing IA risk [15]. Researchers in Germany converted this 20-item test into a 12-item short-form version [16]. Researchers also modified the nine IGD criteria outlined in the *DSM, Fifth Edition*, into a new nine-item short version for UK Internet users [17]. The CIAS is a common Chinese-language tool for evaluating IA symptoms [3]. Researchers in Hong Kong reduced the 26 items of the CIAS to 19 [18]. The aforementioned studies have adopted factor analysis to extract the psychological symptoms of IA and confirm the reliability and validity of their new IA questionnaire versions [16–18]. However, the factor analysis employed for extracting these symptoms used single-dimension calculations, and thus may be unable to analyze non-Markovian problems. Non-Markovian indicates that the optimum decision cannot be determined only from the current state, and memory is needed to assist decision-making. In addition, current IA questionnaires are limited to assessing prolonged Internet use (over 6 months) after browsing Internet (also called narrative surveys). Few questionnaires

evaluate the current experiences of the Internet user during browsing Internet (also termed the experience sampling method). Therefore, the design of an IA questionnaire for the immediate detection of IA risk is a substantial challenge.

Table 1: Psychological symptoms of IA

Symptom	Explanation
S_A	stop or reduce take part in other activities due to using the Internet
S_C	desire to use the Internet
S_{CS}	try to stop or reduce Internet use repeatedly without success
S_{DC}	lie about their amount of using the Internet
S_H	produce psychosocial and physical problems due to overusing the Internet
S_I	lose relationships due to overusing the Internet
S_{MM}	use the Internet to escape from stress or relieve negative mood
S_{PS}	always think about Internet activities or expect the next Internet activities
S_{SO}	risk the educations or occupations due to overusing the Internet
S_T	need to increase time spent on the Internet to achieve the same level of satisfaction as before
S_{TM}	reduce sleep and other activities to use the Internet
S_{WS}	feel displeasure when stop or reduce using the Internet

2.2. Physiological response of IA

Individuals with IA experience dynamic emotional stimulation on the Internet and then regulate psychophysiological responses instantaneously. Cardiovascular, respiratory, body temperature, skin conductance, and other responses in IA have been observed for short periods. The aim of such monitoring is to provide additional data that can be analyzed with IA questionnaires. Relevant results indicated that the risk of IA was positively associated with respiratory responses during 6 min of Internet browsing but negatively related to peripheral temperature [19]. Compared with individuals with low-risk Internet addiction (LIA), individuals with high-risk Internet addiction (HIA) showed higher skin conductance responses after 15 min of Internet use [10]. The respiratory sinus arrhythmia responses of users with HIA while watching negative and positive videos of 2 to 3 min were lower than that of users with LIA [20]. The heart rate and systolic blood pressure of individuals with problematic Internet use both increased after 2 min of stopping Internet use [21]. When playing an online game for 5 min, the high-frequency heart rate variability and cardiopulmonary synchrony for problematic Internet users in an excessive online gaming group were lower than those of the healthy group [22]. Compared with 5 min of baseline values, the natural logarithm of high-frequency heart rate variability of gamers with IGD was lower during 5 min of online gaming with attention [23].

Among the physiological signals, respiration is the only one that reflects emotional responses and affects emotions, motivation, and other physiological signals. Patients with IA and anxiety disorder obtained alleviation for their addiction and anxiety symptoms after 10 weeks of cognitive-behavioral therapy with breathing exercises [33]. Mindful breathing has also been employed to assist gamers with IGD [34] and individuals with a gambling addiction [35] to relieve their desire to play. For physiological regulation, respiration is related to the vagus nerve, which modulates sympathetic and parasympathetic nervous system activities. Slow respiratory rates reduce sympathetic nervous activity or increase parasympathetic nervous activity. Inhalation and exhalation change the intrathoracic pressure, which in turn affects the heart rate and blood pressure. Individuals with high-risk IGD reduced their respiratory rates by using slow and deep breathing exercises while watching negative films [31]. Therefore, the respiration response is a key feature for detecting IA risk, and respiratory adjustments constitute a valuable means of affecting psychophysiological responses.

The major IA-related features are generally extracted from physiological signals using descriptive statistics and generic mathematics methods. However, these methods are limited in analyzing physiological signals with dynamic and instantaneous changes and solving non-Markovian problems. The extraction of instantaneous physiological features and investigation of the relationship between each instantaneous physiological component are crucial for determining IA.

2.3. IA detection using machine learning

Some researchers have proposed machine learning models as computer-aided systems for detecting IA and extracting IA-related psychological and physiological features. Impulsive and compulsive symptoms were adopted to construct prediction models of problematic Internet use by employing naïve Bayes, random forest, and logistic regression methods [24]. Authors used a decision tree to determine IGD and extract the features of gaming cost, gaming time on weekdays and weekends, and offline gaming social, marriage, and self-perceptions [25]. In supervised machine learning, demographic information, gamer statistics, and self-esteem scores were input to classify IGD [26]. Authors combined long short-term memory with an adversarial autoencoder to predict the gamer's IGD risk using duration, costs, feelings of depressive responses, and contextual information in games as input data [27]. Social data from Facebook and Instagram were applied to detect problematic social network use by employing a support vector machine framework [7]. The aforementioned studies were required to collect data over a long period of time. Authors proposed that 10- or 25-s biosignals, such as heart and respiratory rates, eye blinks, and saccadic movements, be input as variables into the support vector machine to determine the urge of an IGD group for gaming [28]. Although long-term and short-term information has been used to predict addiction, the lack of algorithms impedes the use of instantaneous physiological response measurements in IA detection.

3. Research Methodology

3.1. Empirical mode decomposition

Empirical mode decomposition (EMD) [36] is an adaptive filter method applied to decompose nonlinear and nonstationary signals into oscillatory components from high to low frequencies; these components are also termed intrinsic mode functions (IMFs). The decomposition process of EMD is expressed in Algorithm 1.

Algorithm 1: Decomposition process of EMD

Result:

$s_1(t) \leftarrow s(t)$

$c(t) \leftarrow s(t)$

while $c_i(t)$, $i = 1, 2, \ldots, n$ **do**

    \\ Sifting Process

    if standard_deviation $(c(t)) \geqslant 0.02$ **then**; % stop criterion

        $p_{max} \leftarrow$ aidentify_local_maxima$(c(t))$

        $p_{min} \leftarrow$ iidentify_local_minima$(c(t))$

        $e_{max}(t)$ xncompute_upper_envelop(p_{max})

        $e_{min}(t)$ npcompute_lower_envelop(p_{min})

        $m(t) \leftarrow (e_{max}(t) + e_{min}(t))/2$; % mean envelop

        $c(t) \leftarrow c(t) - m(t)$

    end

    $IMF_i(t) \leftarrow c(t)$

    $s_{i+1}(t)\ 1\ s_i(t) - IMF_i(t)$

end

EMD, however, has boundary-effect and mode-mixing problems. Boundary-effect indicates IMF endpoint distortions. Mode mixing refers to different oscillatory waveforms being present in a single IMF. These two problems lead to difficult or erroneous interpretation of the physical meaning of the IMFs but can be resolved using the EEMD method [37]. The analysis process of EEMD involves the addition of Gaussian noise to the source signal; this noise-added signal is then decomposed using the sifting process of the EMD method.

3.2. XCS

The XCS is a powerful machine learning model with high interpretability and flexibility. The interpretability of the XCS stems from the use of a set of collaborative rules known as a population set [P] to express the solution of the problem domain. The interpretation of each rule is the "IF condition THEN action," with the condition part of each rule corresponding to the detected environmental state and coded with a ternary character {0, 1, #}. The # symbol is termed "don't care," which indicates that the corresponding environmental state can be ignored regardless of whether it is 0 or 1. The action part of each rule is the decision-making that the XCS executes. If the output is a binary classification problem, the action can be encoded as a binary string {0, 1}. Furthermore, each rule contains the following three parameters to indicate the prediction and quality of each rule: 1) *prediction*, which is used to predict how much feedback can be obtained from the environment after the action is executed; 2) *prediction error*, which is used to calculate the error between the prediction and actual feedback; and 3) *fitness*, which is used to evaluate the accuracy and quality of the prediction. A modular architecture allows XCS to address many problem areas flexibly, including performance components, exploration components, and credit assignment.

The performance component is responsible for detecting the state of the environment and making decisions. First, the XCS converts the detected environmental state into a binary string. Second, the condition part of each rule in the [P] is compared with the environmental state, and the rule is moved to the matching set [M] only if the conditions of the rule match the environmental state.

If [M] is empty, the XCS initiates the *covering* mechanism to generate new rules that satisfy the environmental state. Third, the XCS must select an action to execute according to the rules in [M]. The system uses rules with the same action in [M] to calculate the expected feedback of the action and form a prediction array [PA]. The XCS adopts two mechanisms, exploration and exploitation, to select the action. In the exploration mode, the expected feedback calculated by each action in [PA] is used as the probability of random selection; in the exploitation mode, the system directly selects the action with the greatest expected feedback. Fourth, the selected action is executed through the effector, and the rules in [M] that match the action move to the action set [A] to update the parameters. The XCS repeats these four steps until convergence or the maximum number of iterations has been reached.

The exploration components are used to assist the XCS in generating new rules to search for the optimum solution in the problem domain. Evolutionary computation (EC) introduces small random factors to gradually optimize the solution and is suitable for the exploration components of the XCS. Typically, the XCS adopts the genetic algorithm (GA) in EC as its exploration components. GA solves the optimization problem by simulating natural selection, heredity, crossover, and mutation mechanisms in evolutionary biology. In the XCS, the GA randomly selects two rules from [P] for inheritance based on the fitness of each rule as the probability. Subsequently, the GA performs crossover on each part of the condition in accordance with the two rules based on the probability χ and mutates each part with probability μ. The GA generates two new rules to replace the rules with lower fitness in [P] to assist the XCS in identifying the global optimum in the problem domain.

The credit assignment assigns and updates the parameters of the rules in [A]. The XCS uses the feedback from the environment to modify the prediction of each rule in [A] as a reference parameter for the next iteration. The modified technique uses the Widrow–Hoff delta rule, as described in Eq. (1), where p is the *prediction* of the rule, P is the actual feedback obtained from the environment, t is the current iteration number, and β is the learning rate ($0 < \beta \leq 1$). The XCS can then use the absolute error to correct the *prediction error* of the rule, as detailed in Eq. (2), where ε denotes the prediction error. The last parameter *fitness* is used to evaluate the quality of the rule. The calculation of *fitness* is based on the relative accuracy of the rule in [A]. First, the XCS must calculate the accuracy k_t of each rule in [A], as expressed in Eq. (3), where α controls the attenuation of the exponential function ($0 < \alpha < 1$), ε_t is the *prediction error* of the rule at time t, and ε_0 is the tolerable error. Subsequently, the XCS calculates the relative accuracy k_t' of the rule compared with the other rules in [A], as described in Eq. (4). The system utilizes the relative accuracy k_t' as the *fitness* of the rule and employs the Widrow–Hoff delta rule to update, as detailed in Eq. (5), where F is the *fitness* of the rule.

$$p_{t+1} = p_t + \beta(P_t - p_t) \tag{1}$$

$$\varepsilon_{t+1} = \varepsilon_{t+1} + \beta(|P_t - p_t| - \varepsilon_t) \tag{2}$$

$$k_t = \begin{cases} \exp\left[\dfrac{(\ln \alpha)(\varepsilon_t - \varepsilon_0)}{\varepsilon_0}\right], & \varepsilon_t > \varepsilon_0 \\ 1, & \text{otherwise} \end{cases} \tag{3}$$

$$k_t{}' = \frac{k_t}{\sum_{k_t \in [A]} k_t} \tag{4}$$

$$F_{t+1} = F_t + \beta(k_t' - F_t) \tag{5}$$

The XCS provides the rule with a *numerosity* parameter to record how many of the same rules are in [P] and uses a rule as a representative of these rules. Therefore, the same rules must only be matched once to maximize the computation speed. The *subsumption* mechanism is also added to assist in generating the most accurate and general rules. In the *subsumption* mechanism, one rule can subsume another rule, and the rule therefore undergoes a certain amount of training that strengthens its generalizability and accuracy. If rule r_a can subsume another rule r_b, r_b is removed from [P], and the numerosity of r_a increases by one. The *subsumption* mechanism usually acts on the GA and [A].

3.3. XCSR

The encoding method used in the *condition* of the rule can only process discrete variables. Wilson proposed to modify the ternary-alphabet encoding of the *condition* into interval predicates to enable the XCS to process continuous variables [30]. The expression of interval predicates is $int_i = (c_i, s_i)$, where int_i is the rule condition that the *i*th interval predicate corresponds to the *i*th variable, c_i is the central value of the range, and s_i refers to the size of the coverage boundary. Changing the expression of the condition modifies the mechanism of components, such as *matching*, *covering*, and the GA. For *matching*, assuming the *i*th variable is x_i, if $c_i - s_i \leq x_i < c_i + s_i$, then the int_i matches x_i. For *covering*, the XCS generates a rule whose interval predicates cover all input variables. The generation method is to set c_i to x_i, and s_i is set to a random value between 0 and s_r, where s_r is defined by the user. For the GA, *crossover* exchanges the int_i with probability χ; *mutation* adds a random value from 0 to s_m to the range of int_i with probability μ, where s_m is defined by user. The remaining mechanism is the same as the XCS operation, with this modified version known as the XCSR. The modular structure allows the XCS to be flexibly modified to adapt to various problem domains.

4. Experiment

4.1. Design of experiment

(1) Objective of this experiment

The CIAS scores of participants with HIA and LIA were collected, and their respiratory signals were observed while they watched online game videos. This study was granted ethical approval by National Chiao Tung University (NCTU) Research Ethics Committee for Human Subject Protection (Approval Number: NCTU-REC-102-009-e).

(2) Method

To assess the risk of IA, we adopted the CIAS, which contains 26 items, with each item ranked from 1 (strong disagreement) to 4 (strong agreement) [3]. The total score of the CIAS ranged from 26 to 104, with a score of 26 to 63 and 64 to 104 reflecting LIA and HIA, respectively. This study also adopted the IGD questionnaire to evaluate IGD risk; this questionnaire consists of nine items requiring yes or no responses [14]. The IGD questionnaire score ranged from 0 to 9, with 0 to 4 and 5 to 9

representing non-IGD and IGD, respectively. The internal consistency of the CIAS and IGD questionnaires, as estimated using Cronbach's α, was 0.94 and 0.81, respectively. The extent of participants' experience using computers and the Internet were recorded, and the Self-Assessment Manikin (SAM) was adopted to evaluate the emotions of participants. The SAM measures emotional valence and arousal on a 9-point Likert scale. A 5-point score on valence represents neutral emotion, and anything higher indicates positive emotions. Conversely, a score lower than 5 is indicative of negative emotions. Arousal was scored from 1 (no feeling) to 9 (strong feeling). To arouse a craving for gaming, 2 min of online game videos were selected from League of Legends (LOL; Riot Games, Los Angeles, CA, USA), MapleStory (MS; Nexon, Seoul, South Korea), and Resident Evil (RE; Capcom, Osaka, Japan). Compared with the active stimuli of playing games, watching clips as passive stimuli may induce more physiological stability, potentially lowering the difference in psychological responses between participants.

(3) Indices

The 26-item scores of CIAS and instantaneous frequency (IF) of respiratory signal were used to detect the risk of IA.

(4) Participant

Participants aged 20–40 years and with no symptoms of depression and anxiety, and cardiovascular diseases were included in the study. Each participant provided written informed consent. A total of 50 participants (36 men and 14 women) were recruited from NCTU.

(5) Procedure

Each participant stared at a gray picture to familiarize themselves with the environment and relax physically at the beginning of the experiment. To induce a feeling of calm, the participant looked at a gray picture for 2 min before each video was played. For emotional stimuli, participants watched three videos (LOL, MS, and RE) in a random order. After each video, participants completed the SAM questionnaire, after which time the gray picture was displayed for another 2 min to allow the participant to recover from emotional arousal. During the experimental procedure, respiratory inductance plethysmography (RIPmate Inductance Belt, Ambu, Ballerup, Denmark) and electrocardiography (Ag/AgCl spot electrodes attached to the left and right subclavian and left lumber regions) were used to measure abdominal wall movement and electrocardiography signals, respectively. These signals were acquired through a DAQCard (USB 6218, NI, Austin, TX, USA) with a 1000 Hz sampling rate. To investigate instantaneous respiratory responses, this study only analyzed abdominal wall movement signals.

4.2. Data analysis

(1) Method

The XCSR model was used to detect the risk of IA by using the CIAS data and the IF of respiration as the input features (Figure 1). The abdominal movement signal was analyzed as the IF according to the signal processing step described in a previous study [36], including integral calculation, signal baseline correction, downsampling, signal decomposition, and transformation.
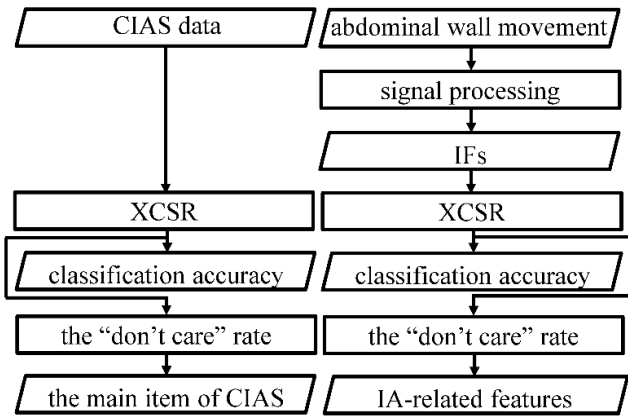
Figure 1: Data analysis procedure.

For statistical calculation, the Mann–Whitney U test was performed to compare differences in the questionnaire data and average IF between the HIA with IGD and LIA with non-IGD groups. The statistical significance was set as $p < 0.05$. To assess the performance of the average IF of respiration in discriminating between HIA with IGD and LIA with non-IGD, the receiver operating characteristic curve (ROC) was calculated, and the area under the curve (AUC) of the ROC indicated the predictive accuracy. The AUC ranged from zero (reflecting incorrect prediction) to one (representing correct prediction); the AUC was 0.7, indicating acceptable predictive power.

(2) Procedure

The integral calculation converted the abdominal movement signal into the volume change of the abdomen by using Simpson's rule. The volume changes of the abdomen were then used to perform baseline correction using the linear fit. To investigate the frequency band of respiratory muscle contractions, the corrected signal was downsampled from a rate of 1000 to 50 Hz. Subsequently, the EEMD method was used to decompose the downsampled signal into eight IMFs. With reference to the frequency bandwidth [38], IMF_1–IMF_2, IMF_3–IMF_4, IMF_5–IMF_6, and IMF_7–IMF_8 represented the diaphragm, respiratory muscle cluster on the abdomen, respiratory wall movement on the abdomen, and body movement components, respectively. In the transformation step, each IMF was transformed into IF by using the normalized direct quadrature [39]. The average value of IF during the viewing of the LOL, MS, and RE videos was treated as the probability of crossover operation (χ) = 0.8, the probability of the instantaneous psychological features that were input into the XCSR model.

Parameters in the XCSR were set as follows [29, 30, 40]: the population size (N) constituted 350 classifiers, probability of # (P#) = 0.33, initial rule $\varepsilon = 0$, initial rule $p = 10$, initial rule $F = 10$, learning rate (β) = 0.2, accuracy function (ε_0) = 10, $\nu = 5$, $\varepsilon_0 = 10$, $\alpha = 0.1$, $\phi = 0.1$, mutation operation (μ) = 0.04, threshold parameter of the GA = 50, deletion thresholds of the GA = 20, and subsumption thresholds of the GA = 20; moreover, the reward setup for accurate prediction of HIA with IGD and LIA with non-IGD was set as 1000, or otherwise 0. The iteration number of the XCSR method was 1950 for both the 26-item score in the CIAS and 24 IF values of respiration as input features. The moving average per 50 exploitations was used to calculate the

classification accuracy of the XCSR. To verify the items in the CIAS and extract the crucial instantaneous respiratory components for detecting the IA risk, rules with a reward of 1000 in [P] were selected, and then the rate of non-"don't care" values ("non-#") in these rules was calculated. To effectively evaluate the training quality of the XCSR and avoid bias, 10-fold cross-validation was used. To eliminate the unevenness of the IF at different frequency bandwidths, the IF underwent min–max normalization. The XCSR does not update any parameters nor generate new rules for the testing data and only uses the exploitation mode when selecting and performing actions on this data. All data management, respiratory signal processing, and XCSR execution operation were conducted using LabVIEW (v.2021, NI, Austin, TX, USA).

5. Results

Of all participants, 34% were from the College of Electrical and Computer Engineering, 28% from the College of Computer Science, 14% from the College of Management, 8% from the College of Humanities and Social Sciences, 6% from the College of Hakka Studies, 6% from the College of Engineering, and 4% were from other departments. Participants were stratified into four groups according to their CIAS and IGD questionnaire scores, namely the LIA with non-IGD, LIA with IGD, HIA with non-IGD, and HIA with IGD groups. This study focused on investigating the psychological symptoms and physiological responses of the LIA with non-IGD and HIA with IGD groups only. Table 2 lists the demographic information and questionnaire responses of the two groups. No significant differences in age, computer and Internet use experience, or emotional valence and arousal were observed between the two groups. The total scores of the CIAS and IGD questionnaires of the HIA with IGD group were significantly higher than those of the LIA with non-IGD group. The emotional valence results indicated that the LOL and MS videos evoked positive emotions, whereas the RE video prompted negative emotions.

Table 2: Demographic information and questionnaire responses

		LIA with non-IGD	HIGD with IGD
Gender (men, women)		12, 9	14, 4
Age (years old)		22.62 ± 1.79	23.67 ± 5.27
Computer experience (year)		13.19 ± 2.71	12.44 ± 4.50
Internet experience (year)		11.71 ± 2.63	10.89 ± 4.19
CIAS		52.10 ± 8.03	79.33 ± 8.88**
IGD questionnaire		1.48 ± 1.25	6.58 ± 1.54**
SAM	valence_LOL	5.95 ± 1.32	6.61 ± 1.33
	arousal_LOL	4.33 ± 2.01	5.28 ± 1.53
	valence_MS	5.76 ± 1.79	5.56 ± 1.46
	arousal_MS	4.33 ± 1.88	4.22 ± 1.73
	valence_RE	3.90 ± 1.87	4.39 ± 1.61
	arousal_RE	6.14 ± 1.80	5.78 ± 1.63

** $p < 0.01$ in the Mann–Whitney U test

The results of classification accuracy of the XCSR with an average of 30 repetitions using 26-item CIAS scores and 24 IFs of respiration in response to online video stimuli are illustrated in Figure 2 and Figure 3, respectively. The classification accuracy of the XCSR model was over 95%. Table 3 presents the mean ± standard deviation scores of each item in the CIAS for the two

groups and the "non-#" rate of each item. The score of each item for the HIA with IGD group was significantly higher than that of the LIA with non-IGD group. The "non-#" rate of the 11th, 14th, 19th, and 22nd items in S_{CS}; 12th, 13th, 15th, 17th, 18th, and 21st items in S_I/S_H; 6th item in S_T; 1st, 8th, 23rd, and 25th items in S_{TM}; and 2nd, 4th, 10th, and 16th items in S_W were higher than the 0.50 value for the CIAS. The 10-fold cross-validation of the XCSR using the 26-item CIAS score and 24 IFs of respiration in response to online video stimuli is described in Table 4 and Table 5, respectively; the average accuracy of 10-fold cross-validation for the XCSR was 87% and 78%, respectively. Table 6 presents the mean ± standard deviation of the IFs of respiration of the two groups while watching the LOL, MS, and RE videos as well as the "non-#" rate of each IF of respiration. The IF_2 of the HIA with IGD group while watching the RE video stimuli was significantly higher than that of the LIA with non-IGD group. Except for IF_5 during the MS video and IF_5 and IF_7 during the RE video, the "non-#" rate of all IFs was higher than the 0.50 value. The ability to predict the risk of IA using IF values, as calculated with the AUC of the ROC, is detailed in Table 7.

6. Discussion

In this study, the XCSR method as a rule-based machine learning system was used to aid in detecting the risk of IA. The "don't care" attribute of the XCSR was employed to verify the items of the CIAS and extract the instantaneous respiratory features for the classification of IA. The results demonstrated that the XCSR model could identify HIA with IGD group or LIA with non-IGD group with a classification accuracy of over 95%. The average accuracy of the 10-fold cross-validation was also over 78%. Researchers have proposed that self-esteem scores [26],

impulsive and compulsive symptoms [24], and physiological signals [28] can serve as the input variables of machine learning systems to predict behavioral addictions. Our findings suggested that the XCSR model is an appropriate computer-aided tool for detecting IA. The psychological symptoms in the CIAS and instantaneous respiratory responses are vital indexes for assessing risky Internet use.
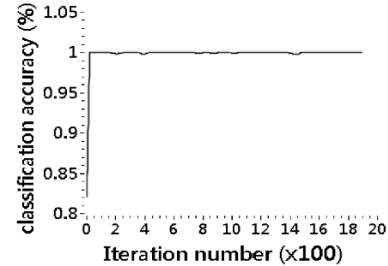
Figure 2: Classification accuracy of the XCSR with the CIAS data as the input feature.
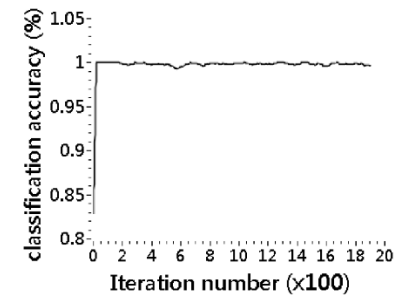
Figure 3: Classification accuracy of the XCSR with the IF of respiration data as the input feature.

Table 3: Psychological symptoms in the CIAS and their corresponding items (modified from [1])

Symptom	item	LIA with non-IGD	HIA with IGD	not # rate
S_{CS}	11. I unable control my impulse of Internet use	1.81 ± 0.68	2.89 ± 0.74**	0.58
	14. When waking up every morning, my first thought is to use the Internet	2.00 ± 0.89	3.05 ± 0.78**	0.75
	19. After stopping Internet use, I crave it again	2.33 ± 0.73	3.32 ± 0.58**	0.67
	20. Without the Internet, my life would be joyless	1.76 ± 0.77	2.79 ± 0.85**	0.42
	22. I try to spend less time on the Internet, but failure	2.14 ± 0.57	3.05 ± 0.71**	0.67
S_I/S_H	7. Even if the Internet negatively affects my interpersonal relationships, my use remains unreduced	1.43 ± 0.60	3.21 ± 0.71**	0.42
	12. I find that I reduce the time spent with friends due to the Internet	1.81 ± 0.51	2.95 ± 0.78**	0.67
	13. I feel aches and soreness in the back or other discomfort due to Internet use	2.00 ± 0.84	3.11 ± 0.94**	0.92
	15. Using Internet has negative effects on my education or work	1.86 ± 0.73	2.84 ± 0.60**	0.75
	17. I reduced interaction with my family due to Internet use	1.76 ± 0.77	3.05 ± 0.85**	0.58
	18. I reduce my recreational activities to use the Internet	2.00 ± 0.63	3.16 ± 0.83**	0.67
	21. Using the Internet has negative effects on my health	2.10 ± 0.70	2.95 ± 0.78**	0.83
S_T	3. I perceive my Internet use as getting longer and longer	2.43 ± 0.51	3.05 ± 0.71**	0.50
	6. I spend more time on the Internet than I originally intended	2.71 ± 0.64	3.53 ± 0.51**	0.58
	9. I have spent more time on the Internet since last semester	2.00 ± 0.63	2.95 ± 0.71**	0.33
	24. I need to increase the amount of time I use the Internet to achieve the same satisfaction as before	1.67 ± 0.48	2.63 ± 0.83**	0.42
S_{TM}	1. I have been told more than once that I spend too much time on the Internet	2.48 ± 0.75	3.21 ± 0.79**	0.58
	8. My sleep time is less than 4 hours from using the Internet	1.76 ± 0.89	3.16 ± 0.96**	0.67
	23. I am used to reducing my sleep time to use the Internet	1.52 ± 0.75	2.79 ± 0.79**	0.67
	25. I do not eat on time due to using the Internet	2.05 ± 0.80	2.84 ± 0.90**	0.58
	26. I use the Internet all night, which causes daytime tiredness	2.33 ± 1.02	3.00 ± 0.58**	0.50
S_{WS}	2. I feel displeasure when I stop using the Internet for a period of time	2.05 ± 0.67	3.26 ± 0.73**	0.58
	4. I feel restless and irritable when the Internet is unavailable	2.52 ± 0.87	3.37 ± 0.68**	0.75
	5. I feel energetic upon using the Internet regardless of a fatiguing experience	1.62 ± 0.59	2.84 ± 0.76**	0.50
	10. I feel distressed when stopping the Internet use for a period of time	1.57 ± 0.51	2.89 ± 0.74**	0.83
	16. I feel like I am missing something when stopping the Internet use for a period of time	2.38 ± 0.97	3.37 ± 0.60*	0.75

* p < 0.05 and ** p < 0.01 in the Mann–Whitney U test

The results indicated that the 11th, 14th, 19th, and 22nd items in S_{CS}; 12th, 13th, 15th, 17th, 18th, and 21st items in S_I/S_H; 6th item in S_T; 1st, 8th, 23rd, and 25th items in S_{TM}; and 2nd, 4th, 10th, and 16th items in S_W were maintained in the CIAS item verification. This result, however, is inconsistent with that of other works. Authors modified the items in the CIAS for grade 7-to-13 Hong Kong students, retaining the 14th and 20th items in S_{CS}; 12th, 13th, 15th, 17th, 18th, and 21st items in S_I/S_H; 3rd, 6th, 9th, and 24th items in S_T; 1st, 8th, 23rd, and 26th items in S_{TM}; and 2nd, 5th, and 16th items in S_W [18]. Authors changed the 26-item CIAS into a 10-item smartphone addiction inventory, retaining the 22nd item in S_{CS}; 7th, 13th, and 15th items in S_I/S_H; 3rd, 6th, and 9th items in S_T; 23rd item in S_{TM}; and 2nd and 4th items in S_W for Taiwan students aged 18 to 31 years [41]. These inconsistent results are likely attributable to differences in age, country, analysis method, or assessment of psychological symptoms between studies. Our findings must be validated with more data.

In this study, the "don't care" rate of IF_1–IF_8 of respiration during LOL video stimuli; IF_1–IF_4 and IF_6–IF_8 of respiration during MS video stimuli; and IF_1–IF_4, IF_6, and IF_8 of respiration during RE video stimuli were lower than 0.50. IMF_1–IMF_2, IMF_3–IMF_4, IMF_5–IMF_6, and IMF_7–IMF_8 represented the diaphragm, respiratory muscle cluster on the abdomen, respiratory wall movement on the abdomen, and body movement components, respectively [38]. The LOL and MS videos aroused positive emotions in the participants, whereas the RE video evoked negative emotions.

Table 4: The 10-fold cross-validation of the XCSR with the CIAS data as the input feature

Testing	1	2	3	4	5
Accuracy	1.00	1.00	1.00	0.50	1.00
Testing	6	7	8	9	10
Accuracy	1.00	0.75	0.75	1.00	0.67
Average accuracy			0.87		

Table 5: The 10-fold cross-validation of the XCSR with the IF of respiration data as the input feature

Testing	1	2	3	4	5
Accuracy	1.00	0.25	0.5	0.75	0.75
Testing	6	7	8	9	10
Accuracy	1.00	0.75	1.00	0.75	1.00
Average accuracy			0.78		

Emotions affect respiratory responses, and respiratory muscular contraction and relaxation drive the thoracic and abdominal wall movements. Compared with the users with LIA, those with HIA exhibited a higher abdominal muscle contraction amplitude and lower abdominal wall movements whilestaring at positive or negative pictures of 3- to 12-s [42]. The respiratory responses positively related to the scores of the CIAS [19], and the respiratory rate was used to predict the craving of individuals with IGD to play online games [28]. Those with HIA are on the Internet excessively, which can cause aches and soreness in the back [2, 3] or hands; they may change their posture to relieve this pain. Therefore, we determined that the instantaneous responses

of respiratory muscle clusters, respiratory wall movements, and body movements constitute potential biomedical indexes through which to observe the psychological reaction of the IA group, with the various emotional stimuli eliciting different respiratory responses. These responses can assist in detecting IA symptoms. We also demonstrated that IF_2 during the RE video stimuli was an acceptable predictor of IA risk, which was calculated using the ROC. However, this finding was inconsistent with the feature extraction of the XCSR model and can be attributed to several possible explanations. The ROC employs conventional mathematical calculation, but the XCSR method observes the relationship between features using multiple dimension attributes. In addition, the sampling size and parameter setup of the XCSR may influence feature extraction. Therefore, future works should employ more data samples and test different parameters of the XCSR method.

Table 6: Mean ± standard deviation of IF of respiration

Feature		LIA with non-IGD	HIGD with IGD	not # rate
LOL	IF_1	12.21 ± 0.15	12.19 ± 0.19	0.61
	IF_2	7.42 ± 0.27	7.32 ± 0.38	0.67
	IF_3	3.64 ± 0.21	3.53 ± 0.33	0.64
	IF_4	1.18 ± 0.20	1.16 ± 0.20	0.59
	IF_5	0.35 ± 0.08	0.34 ± 0.07	0.64
	IF_6	0.18 ± 0.03	0.18 ± 0.04	0.57
	IF_7	0.08 ± 0.02	0.07 ± 0.02	0.59
	IF_8	0.04 ± 0.01	0.04 ± 0.01	0.58
MS	IF_1	12.21 ± 0.20	12.24 ± 0.15	0.72
	IF_2	7.33 ± 0.29	7.33 ± 0.40	0.56
	IF_3	3.55 ± 0.27	3.57 ± 0.38	0.58
	IF_4	1.22 ± 0.28	1.31 ± 0.30	0.58
	IF_5	0.35 ± 0.11	0.36 ± 0.13	0.49
	IF_6	0.18 ± 0.04	0.18 ± 0.05	0.56
	IF_7	0.08 ± 0.02	0.08 ± 0.03	0.54
	IF_8	0.04 ± 0.01	0.04 ± 0.01	0.55
RE	IF_1	12.31 ± 0.17	12.29 ± 0.14	0.63
	IF_2	7.74 ± 0.16	7.61 ± 0.28*	0.53
	IF_3	3.93 ± 0.16	3.81 ± 0.32	0.56
	IF_4	1.54 ± 0.28	1.45 ± 0.35	0.55
	IF_5	0.38 ± 0.09	0.41 ± 0.17	0.37
	IF_6	0.25 ± 0.06	0.21 ± 0.04	0.56
	IF_7	0.10 ± 0.02	0.10 ± 0.03	0.44
	IF_8	0.05 ± 0.01	0.05 ± 0.02	0.66

* $p < 0.05$ in the Mann–Whitney U test

This study has some limitations. First, the small sample size may lead to overfitting and affect the verification of items in a questionnaire or the feature extraction of physiological responses. Second, the parameter setup of the XCSR was based on empirical studies and must be tested with different values. Third, the physiological responses of participants while watching online game videos may differ from those actively playing online games. Fourth, we observed respiratory responses during 2 min of emotional stimuli. Future studies can investigate second-by-second psychological responses under varying window sizes. Fifth, the unequal sample of men and women may affect the analysis of the XCSR. Notwithstanding these limitations, the XCSR model provides researchers with a computer-aided detection method for IA. The verification of items in a questionnaire can assist in the

design of a questionnaire for experience sampling method and the real-time detection of IA. The extraction of instantaneous respiratory features can provide psychological insight into the development of IA behavior.

Table 7: Predictive accuracy for the IF value in individuals watching online game videos, calculated with the area under the curve (AUC) of the ROC

Feature		AUC	p-value	95% confidence interval	
				lower bound	upper bound
LOL	IF_1	0.52	0.82	0.33	0.71
	IF_2	0.52	0.80	0.33	0.71
	IF_3	0.58	0.41	0.39	0.76
	IF_4	0.53	0.74	0.35	0.72
	IF_5	0.56	0.55	0.37	0.74
	IF_6	0.49	0.96	0.31	0.68
	IF_7	0.63	0.17	0.45	0.81
	IF_8	0.51	0.96	0.32	0.69
MS	IF_1	0.56	0.52	0.33	0.71
	IF_2	0.51	0.91	0.38	0.74
	IF_3	0.53	0.78	0.32	0.70
	IF_4	0.59	0.35	0.33	0.72
	IF_5	0.52	0.84	0.41	0.77
	IF_6	0.53	0.74	0.33	0.70
	IF_7	0.55	0.61	0.34	0.72
	IF_8	0.56	0.55	0.36	0.74
RE	IF_1	0.54	0.63	0.36	0.73
	IF_2	0.70	0.03	0.51	0.89
	IF_3	0.67	0.07	0.48	0.86
	IF_4	0.64	0.14	0.45	0.82
	IF_5	0.52	0.87	0.32	0.71
	IF_6	0.67	0.07	0.50	0.84
	IF_7	0.58	0.41	0.39	0.76
	IF_8	0.52	0.82	0.33	0.71

7. Conclusion

This study aimed to detect IA using the XCSR model. This model was also applied to verify the items of the CIAS and extract instantaneous respiratory features. The XCSR achieved over 95% classification accuracy for both the CIAS data and IF of respiration during online game video stimuli. The average accuracy of the 10-fold cross-validation was also higher than 75% for both input data. Furthermore, using the XCSR model, the CIAS was modified from a 26-item to 19-item questionnaire and the IFs of respiratory muscle contractions on the abdomen, respiratory wall movements on the abdomen, and body movements were extracted as the IA-related features. The current findings indicated the suitability of the XCSR method as a computer-aided system for real-time IA detection. The extracted psychological symptoms and instantaneous respiratory responses are key features in determining the development of IA. Further studies should collect more samples to validate these findings; design experiments incorporating active gameplay stimuli; and measure the instantaneous psychophysiological responses per second under varying window sizes.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] H.-M. Ji, L.-Y. Chen, T.-C. Hsiao, "Real-Time detection of Internet addiction using reinforcement learning system," in 2019 Genetic and Evolutionary Computation Conference (GECCO), 1–9, 2019, doi:10.1145/3319619.3326882.

[2] K.S. Young, "Internet addiction: The emergence of a new clinical disorder," CyberPsychology and Behavior, **1**(3), 237–244, 1998, doi: 10.1089/cpb.1998.1.237.

[3] S.-H. Chen, L.-J. Weng, Y.-J. Su, H.-M. Wu, P.-F. Yang, "Development of Chinese Internet addiction scale and its psychometric study," Chinese Journal of Psychology, **45**(3), 279–294, 2003, doi: 10.1037/t44491-000.

[4] C. Chou, L. Condron, J.C. Belland, "A review of the research on Internet addiction," Educational Psychology Review, **17**(4), 363–388, 2005, doi: 10.1007/s10648-005-8138-1.

[5] J.S. Lemmens, P.M. Valkenburg, J. Peter, "Development and validation of a game addiction scale for adolescents," Media Psychology, **12**, 77–95, 2009, doi: 10.1080/15213260802669458.

[6] M. Beranuy, U. Oberst, X. Carbonell, A. Chamarro, "Problematic internet and mobile phone use and clinical symptoms in college students: The role of emotional intelligence," Computers in Human Behavior, **25**, 1182–1187, 2009, doi:10.1016/j.chb.2009.03.001.

[7] H.-H. Shuai, C.-Y. Shen, D.-N. Yang, Y.-F. C. Lan, W.-C. Lee, P. S. Yu, M.-S. Chen, " A comprehensive study on social network mental disorders detection via online social media mining," IEEE Transactions on Knowledge and Data Engineering, **30**(7), 1212–1225, 2018, doi: 10.1109/TKDE.2017.2786695.

[8] American Psychiatric Association, Diagnostic and statistical manual of mental disorders, fifth edition (DSM-5), American Psychiatric Association Press Inc., Arlington, VA, 2013.

[9] Y.-S. Jo, S.-Y. Bhang, J.-S. Choi, H.-K. Lee, S.-Y. Lee, Y.-S. Kweon, "Clinical characteristics of diagnosis for Internet gaming disorder: comparison of DSM-5 IGD and ICD-11 GD diagnosis," Journal of Clinical Medicine, **8**, 945, 2019, doi:10.3390/jcm8070945.

[10] M. Romano, A. Roaro, F. Re, L.A. Osborne, R. Truzoli, P. Reed, "Problematic internet users' skin conductance and anxiety increase after exposure to the internet," Addictive Behaviors. **75**, 70–74, 2017, doi: 10.1016/j.addbeh.2017.07.003.

[11] J. Morahan-Martin, P. Schumacher, "Incidence and correlates of pathological Internet use among college students," Computers in Human Behavior, **16**(1), 13–29, 2000, doi: 10.1016/S0747-5632(99)00049-7.

[12] Y. Sun, Y.-Y. Li, Y.-P. Bao, S.-Q. Meng, Y.-K. Sun, G. Schumann, T. Kosten, J. Strang, L. Lu, J. Shi, "Brief report: Increased addictive Internet and substance use behavior during the COVID-19 pandemic in China," American Journal on Addictions, **29**, 268–270, 2020, doi: 10.1111/ajad.13066.

[13] M.C. Zara, L.H. A. Monteiro, "The negative impact of technological advancements on mental health: An epidemiological approach," Applied Mathematics and Computation, **396**, 125905, 2021, doi: 10.1016/j.amc.2020.125905.

[14] N.M. Petry, F. Rehbein, D.A. Gentile, J.S. Lemmens, Hans-Jürgen Rumpf, T. Mößle, G. Bischof, R. Tao, D.S.S. Fung, G. Borges, et al., "An international consensus for assessing Internet gaming disorder using the new DSM-5 approach," Addiction, **109**, 1399–1406, 2014, doi: 10.1111/add.12457.

[15] K.S. Young, Caught in the net: How to recognize the signs of internet addiction--and a winning strategy for recovery: John Wiley & Sons, 1998.

[16] M. Pawlikowski, C. Altstötter-Gleich, M. Brand, "Validation and psychometric properties of a short version of Young's Internet addiction test," Computers in Human Behavior, **29**(3), 1212–1223, 2013, doi: 10.1016/j.chb.2012.10.014.

[17] H. M. Pontes, M.D. Griffiths, "Measuring DSM-5 internet gaming disorder: Development and validation of a short psychometric scale," Computers in Human Behavior, **45**, 137–143, 2015, doi: 10.1016/j.chb.2014.12.006.

[18] K.-K. Mak, C.-M. Lai, C.-H. Ko, C. Chou, D.-I. Kim, H. Watanabe, and

R.C.M. Ho, "Psychometric properties of the revised Chen Internet addiction scale (CIAS-R) in Chinese adolescents," Journal of Abnormal Child Psychology, **42**, 1237–1245, 2014, doi: 10.1007/s10802-014-9851-3.

[19] D.-W. Lu, J.-W. Wang, A.C.W Huang, "Differentiation of Internet addiction risk level based on autonomic nervous responses: The Internet-addiction hypothesis of autonomic activity," Cyberpsychology, Behavior, and Social Networking, **13**, 371–378, 2010, doi: 10.1089/cyber.2009.0254.

[20] D.-L. Hsieh, T.-C. Hsiao, "Respiratory sinus arrhythmia reactivity of internet addiction addicts in negative and positive emotional states using film clips stimulation," BioMedical Engineering OnLine, **15**, 69, 2016, doi: 1186/s12938-016-0201-2.

[21] P. Reed, M. Romano, F. Re, A. Roaro, L.A. Osborne, C. Viganò, R. Truzoli, "Differential physiological changes following internet exposure in higher and lower problematic internet users," PLoS ONE, **12**(5), e e0178480, 2017, doi: 10.1371/journal.pone.0178480.

[22] J.-S. Chang, E.-Y. Kim, D. Jung, S.-H. Jeong, Y. Kim, M.-S. Roh, Y.-M. Ahn, B.-J. Hahm, "Altered cardiorespiratory coupling in young male adults with excessive online gaming," Biological Psychology, **110**, 159–166, 2015, doi: 10.1016/j.biopsycho.2015.07.016.

[23] S.-J. Hong, D. Lee, J. Park, K. Namkoong, J. Lee, D.P. Jang, J.E. Lee, Y.-C. Jung, and I.Y. Kim, "Altered heart rate variability during gameplay in Internet gaming disorder: The impact of situations during the game," Frontiers in Psychiatry, **9**, 429, 2018, doi: 10.3389/fpsyt.2018.00429.

[24] K. Ioannidis, S.R. Chamberlain, M.S. Treder, F. Kiraly, E.W. Leppink, S.A. Redden, D.J. Stein, C. Lochner, J.E. Grant, "Problematic internet use (PIU): Associations with the impulsive-compulsive spectrum. An application of machine learning in psychiatry," Journal of Psychiatric Research, **83**, 94–102, 2016, doi: 10.1016/j.jpsychires.2016.08.010.

[25] M.J. Rho, J.-E. Jeong, J.-W. Chun, H. Cho, D.J. Jung, I.Y. Choi, D-J. Kim, " Predictors and patterns of problematic Internet game use using a decision tree model," Journal of Behavioral Addictions, **5**(3), 500–509, 2016, doi: 10.1556/2006.5.2016.051.

[26] S. Aggarwal, S. Saluja, V. Gambhir, S. Gupta, S.P.S. Satia, "Predicting likelihood of psychological disorders in PlayerUnknown's Battlegrounds (PUBG) players from Asian countries using supervised machine learning," Addictive Behaviors, **101**, 106132, 2020, doi: 10.1016/j.addbeh.2019.106132.

[27] D. Seth, S. Eswaran, T. Mukherjee, M. Sachdeva, "A deep learning framework for ensuring responsible play in skill-based cash gaming," in 2020 IEEE International Conference on Machine Learning and Applications (ICMLA), 454–459, 2020, doi: 10.1109/ICMLA51294.2020.00077.

[28] H. Kim, J. Ha, W.-D. Chang, W. Park, L. Kim, C.-H. Im, "Detection of craving for gaming in adolescents with Internet gaming disorder using multimodal biosignals," Sensors, **18**, 102, 2018, doi:10.3390/s18010102.

[29] S.W. Wilson, "Classifier fitness based on accuracy," Evolutionary Computation, **3**, 149–175, 1995, doi: 10.1162/evco.1995.3.2.149.

[30] S.W. Wilson, Get real! XCS with continuous-valued inputs. Lanzi et al. (eds) Learning Classifier Systems: From Foundations to Applications. Springer-Verlag, Berlin, 1813 of Lecture Notes in Artificial Intelligence, 209–219,

2000.

[31] H.-M. Ji, T.-C. Hsiao, "A novel cue-induced abdominal reaction analysis for Internet gaming disorder," Journal of Medical Systems, **43**, 94, 2019, doi: 10.1007/s10916-019-1221-9.

[32] M. Griffiths, "Nicotine, tobacco and addiction," Nature, **384**, 18, 1996, doi: 10.1038/384018a0.

[33] V.A. Santos, R. Freire, M. Zugliani, P. Cirillo, H.H. Santos, A.E. Nardi, A.L. King, "Treatment of Internet addiction with anxiety Disorders: Treatment protocol and preliminary before-after results involving pharmacotherapy and modified cognitive behavioral therapy," JMIR Research Protocols, **5**, e46, 2016, doi: 10.2196/resprot.5278.

[34] W. Li, E.L. Garland, J.E. OʹBrien, C. Tronnier, P. McGovern, B. Anthony, M.O. Howard, "Mindfulness-oriented recovery enhancement for video game addiction in emerging adults: preliminary findings from case reports," International Journal of Mental Health Addiciton, **16**, 928–945, 2018, doi: 10.1007/s11469-017-9765-8.

[35] A.R. Melero Ventola, J.R. Yela, A. Crego, M. Cortés-Rodríguez, "Effectiveness of a mindfulness-based cognitive therapy group intervention in reducing gambling-related craving," **20**(1), 107–134, 2020, doi: 10.24193/jebp.2020.1.7.

[36] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proceedings of the Royal Society A, **454**(1971), 903–995, 1998, doi: 10.1098/rspa.1998.0193.

[37] Z. Wu, N.E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," Advances in Adaptive Data Analysis, **1**, 1–41, 2009, doi: 10.1142/S1793536909000047.

[38] Y.-C. Chen, T.-C. Hsiao, "Towards estimation of respiratory muscle effort with respiratory inductance plethysmography signals and complementary ensemble empirical mode decomposition," Medical & Biological Engineering & Computing, **56**(7), 1293–1303, 2018. doi: 10.1007/s11517-017-1766-z.

[39] N.E. Huang, Z. Wu, S.-R. Long, K.C. Arnold, X. Chen, K. Blank, "On instantaneous frequency," Advances in Adaptive Data Analysis, **1**, 177–229, 2009, doi: 10.1142/S1793536909000096.

[40] M.V. Butz, S.W. Wilson, An algorithmic description of XCS. In Lanzi, P., Stolzmann, W., and Wilson, S., Ed., Advances in Learning Classifier Systems. IWLCS 2000. Lecture Notes in Computer Science, vol 1996. Springer, Berlin, Heidelberg, 253–272, 2001, doi: 10.1007/3-540-44640-0_15.

[41] Y.-H. Lin, Y.-C. Pan, S.-H. Lin, S.-H. Chen, "Development of short-form and screening cutoff point of the smartphone addiction inventory (SPAI-SF)," International Journal of Methods in Psychiatric Research, **26**(2), e1525, 2017, doi: 10.1002/mpr.1525.

[42] D.-L. Hsieh, H.-M. Ji, T.-C. Hsiao, B.-S. Yip, "Respiratory feature extraction in emotion of Internet addiction abusers using complementary ensemble empirical mode decomposition," Journal of Medical Imaging and Health Informatics, **5**, 1–9, 2015, doi: 10.1166/jmihi.2015.1405.

Electrification of a Bus Line in Savona Considering Depot and Opportunity Charging

Michela Longo[*,1], Carola Leone[1], Luise Lorenz[1], Andrea Strada[1], Wahiba Yaici[2]

[1]*Politecnico di Milano, Energy Department, Milan, 20056, Italy*

[2]*CanmetENERGY Research Centre - Natural Resources Canada, Ottawa, Ontario, Canada*

A B S T R A C T

A transition towards electrification of the public transport sector is ongoing in many cities around the world, as a response to global warming and pollution. However, the question is whether the current state of technology is already sufficient to replace the conventional buses with electric ones and if the existing charging facilities are appropriate to ensure the smooth operation of the buses. Therefore, this work aims to verify the technical feasibility of the electrification of an existing urban line. The purpose is achieved by evaluating a case study on a public transport bus line in the city of Savona, Italy. The average energy consumption of an electric bus operating in the considered line path is estimated in order to investigate the possible locations and sizes of the charging systems to install. The results show that the correct service operation of the electric buses can be achieved by installing one opportunity charger of at least 300 kW in one of the terminals or by installing three 43 kW charging ports in the depot.

1. Introduction

With the Paris Agreement in 2015, the members of the UNFCCC (United Nations Framework Convention on Climate Change) agreed to undertake ambitious efforts to keep the global average temperature rise well below 2°C above pre-industrial levels and to strive for an increase of less than 1.5°C within this century [1]. To reach this ambitious long-term goal, the nations must take action and reduce their overall carbon dioxide (CO_2) emissions drastically. Regarding the European Union (EU) total emissions, the transportation sector makes a share of 21%, whereas road buses in combination with other heavy-duty vehicles contribute to 6% of the total emitted CO_2 [2]. Following the electrification of the public transport sector could have a significant role in the transition towards more sustainable mobility. Evaluating the trend in the last years is quite clear how electric buses continue to witness a dynamic development with more than 460 000 vehicles on the world's road in 2018, almost 100 000 more than in 2017 [3]. The electrification of public transport is not just an important step towards a free-CO_2 transport sector, but it also entails positive effects on the living conditions of the inhabitants of the city. Indeed, it enhances air quality as well as noise pollution. [4]. Therefore, the transformation of the public transport sector is not just an important step towards sustainability but also an important sign to the future generation as a livable city of the future.

This electric transition however is bringing, opportunities as well as challenges. As a matter of fact, this technology, in addition to having very high capital costs, also required the installation of a proper design charging infrastructure [5]. In this global context, this paper proposes the electrification of a currently operated bus line in Savona, a small city in the northern part of Italy. More in detail, the different possible charging infrastructures that can be used to correctly electrified the line are evaluated. To achieve this scope, first of all, the energy consumed by an electric bus running on the considered line must be carried out. In the literature, the estimation of an electric bus energy consumption along a route is typically obtained considering standard driving cycles, for instance in [6] the energy consumption of city transit electric buses is computed using four international driving cycles. However, these general driving cycles do not take into account all the statistical uncertainties of the studied bus route. In other works, driving cycles are developed from real-world data, as reported in [7] presents a freeway driving cycle developed based on the traffic information in California. In the present work, instead, the energy consumption is found by simulating the driving cycle of an electric bus on two different sections of the considered line. The test sections are chosen to represent the entire characteristics of the road structure in the overall line. Therefore, this method allows us to develop a very precise driving cycle for the studied route, which considers also factors such as road slope and signs.

This paper is organized as follows. In section 2 the methodology used to estimate the consumption of electric buses along the considered lines is proposed. The methodology is applied

*Corresponding Author: Michela Longo, via La Masa 34, 20156, Milan (Italy), michela.longo@polimi.it

to the case study in section 3 and the energy consumption of the line is estimated based on the previous considerations. In section 4 the possible charging systems solutions are discussed. Finally, in section 5 the conclusions are presented.

2. Methodology to assess the energy consumption

The overall energy consumption is determined by different factors. Not only the technical characteristics of the bus itself define the consumption of the bus, but also external factors and the characteristic of the route have a huge influence [8]. The following section describes the technical properties of the chosen bus, as well as the assumptions and equations used to calculate the overall energy consumption.

2.1. Test conditions

To get a proper assessment of the e-bus consumption, it is necessary to define the operating conditions assume for this study, such as vehicle occupancy level, service constraints, and climatic conditions.

1. Occupancy rate: the e-bus is assumed to be ballasted at 2/3 of its payload which means it is transporting about 60 passengers (n_{pass}). Therefore, the overall mass of the loaded vehicle (m) is obtained in (1), by adding to the mass of the vehicle (m_v), the mass due to the presence of passengers, assuming an average weight of 70 kg for each passenger (m_{pass}).

$$m = m_v + n_{pass} \cdot m_{pass} \qquad (1)$$

2. The external temperature is assumed to be equal to 5°C (average minimum temperature in Savona). This choice has been made to consider the worst-case scenario in terms of auxiliary power consumption, as a matter of fact, it has been widely proved in the literature that the highest secondary is achieved by the heating system.

3. The actual existing service provided by TPL (Trasporto Pubblico Locale) for Line 6 cannot be degraded by the electrification of the line.

2.2. Energy consumption computation model

In the following the model and the equations necessary to evaluate the speed, power, and energy consumption profiles of the e-bus over the chosen route are described. The theoretic model used is based on the motion equation reported in (2) [9].

$$F - R = m_e \cdot a \qquad (2)$$

where F is the resultant of active forces, we will refer to this quantity as tractive effort, R is the resultant of passive forces also called resistances, m_e is the equivalent mass and a is the vehicle acceleration.

Through (3) the equivalent mass of the e-bus is calculated. Therefore, the equivalent mass is the gross mass of the vehicle m augmented by a factor β which must weigh up the inertial momentum of all the rotating masses inside the vehicle such as wheels, shafts, and axles [10].

$$m_e = m \cdot (1 + \beta) \qquad (3)$$

As regards the total resistance to the bus motion, in the present analysis three terms have been taken into account: the rolling resistance, the aerodynamic resistance, and the grade resistance shown respectively in (4), (5), and (6).

The rolling resistance is due to the continuous deformation of the tire/wheel during its rotation and is computed through (4).

$$R_1 = \left[0.005 + \left(\frac{1}{p}\right)\left(0.01 + 0.0095\left(\frac{v_1}{100}\right)^2\right)\right] mg \qquad (4)$$

where p is the pressure of the wheels in bars, v_1 is the speed in m/s, m is the vehicle mass in kg and, g is the acceleration of gravity in m/s^2.

The aerodynamic resistance represented in (5), also called wind or air resistance, depends on the square of the vehicle speed (v), the density of the medium in which the vehicle is moving which in this case is air (ρ_{air}), on the frontal area of the vehicle (A) and its shape, considered through a drag coefficient (C_x). In the calculations, for reasons of simplicity, the air density is assumed constant and the wind speed equal to zero.

$$R_2 = \frac{1}{2} \rho_{air} A C_x v_2^2 \qquad (5)$$

Finally, the grade resistance appears when the vehicle is traveling over a surface with a positive or negative slope α and it is computed by (6).

$$R_i = mg \sin \alpha \qquad (6)$$

Regarding the analysis of the power consumption of the bus, the formula varies depending on the sign of the tractive effort. If a positive tractive effort is applied by the motors to the wheels, it means that the vehicle is accelerating, on the contrary, a negative tractive effort corresponds to a braking condition; finally, a nil tractive effort implies that the motor is turned off (or decoupled from the wheels) and the vehicle naturally decelerated because of the external resistances. Equations (7) and (8) describe the interrelations that apply in each case in order to compute the electric power P_e provided by the motor.

$$P_e = F_T \frac{v}{\eta_T} + P_{aux} \qquad for \ F_T > 0 \qquad (7)$$

$$P_e = F_T v \eta_B \eta_{RBS} + P_{aux} \qquad for \ F_T < 0 \qquad (8)$$

Where F_T is the value of the tractive effort, v is the speed, η_T represents the efficiency in traction, P_{aux} is the value of the power absorbed by the auxiliaries, η_B is the efficiency in braking and η_{RBS} represents the amount of recoverable energy. Since during braking the value of the tractive effort is negative, consequently the value of the power is also negative and this means that there is a recovery of energy in that instant of time.

The value of the current provided by the supply system (in this case the battery) is calculated in (3), where V is the supply battery voltage.

$$I = \frac{P_e}{V} \qquad (9)$$

Lastly, the cumulative energy required by the bus is obtained by taking the integral of the electric power over time. However, since in this study, the procedure is based on discrete-time values with a time interval Δt of 1 s, for the energy computation a backward integration is used and reported in (10).

$$E_n = P_{e-n} \frac{\Delta t}{3600} + E_{n-1} \qquad (10)$$

where n is the discretization index and P_{e-n} indicates the electric power exchanged at the n time instant.

2.3. Characteristics of electric bus

The chosen electric bus model, for the analysis, is the eCitaro developed by Mercedes-Benz. This model is a 12 m long bus with a mass of 15800 kg. It is completely electric and equipped with an onboard lithium-ion battery composed of 15 modules, each one made of 12 battery cells. The nominal maximum power this bus model can provide is 250 kW, while the supply battery voltage is 400 V. The traction battery must provide the energy also to supply the consumption of the auxiliary service, i.e., rear and front lights, climatization, and radio [11]. The maximum amount of recoverable energy through the regenerative braking system is set equal to 65 %, while the efficiency in traction and braking are considered both equal to 80% in order not to have unbalanced distributions of power between the acceleration mode and the deceleration one. The mass of the vehicle, with the assumed number of passengers declared previously, is 20 t, thus leading to a value for the equivalent mass equal to 23.2 t. For safety and comfort reasons, the maximum acceleration of the bus is limited to 0.8 m/s^2 while for braking deceleration the constraint is less restrictive, and its value can reach values up to 1.5 m/s^2. Given the values of the equivalent mass and maximum acceleration, it is possible to compute the maximum tractive effort at the starting phase by (11), which is equal to 18.56 kN.

$$F_T^{max} = m_e \cdot a_{max} \qquad (11)$$

Lastly, the obtained maximum tractive effort must be compared with the maximum adherence force, which is a limit of the tractive effort which must not be overcome otherwise slippage occurs. However, since the speed of the bus is rather small because it is limited to 50 km/h by the Italian laws about speed limits in the urban environment, and because the value of adhesion coefficient is high (equal to 0.6) due to the contact between the tires and the street, the adhesion limit value is never reached and so the bus will never work in slip conditions.
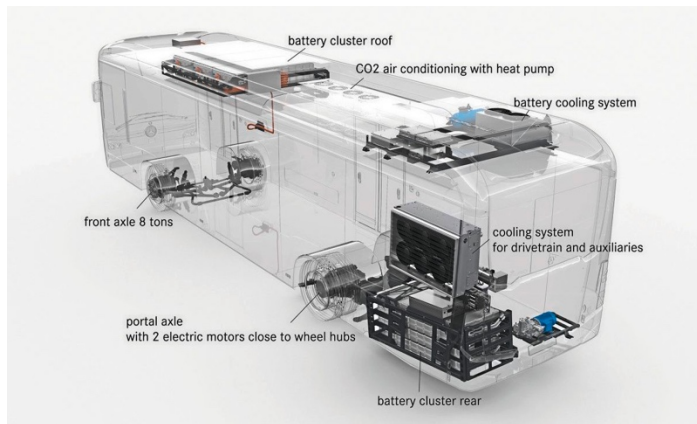
Figure 1: e-Citaro scheme [11].

The eCitaro manufactured by Mercedes-Benz has been considered in this study, since during the last two years, this model has taken over the market of electric buses and it is common in different European countries like Germany, Luxembourg, Sweden,

and Norway [12]. Moreover, it has the most similar characteristics to the conventional bus currently employed by TPL in the selected line. Indeed, the buses currently working in Savona for the public transport services are the Citaro C2 models, produced by Mercedes-Benz too. They are gasoline buses having an internal combustion engine able to provide a nominal power of 220 kW.

The main advantages of choosing an electric bus fleet instead of a gasoline one are the reductions in pollution and fuel costs. Furthermore, the electric motors are less loud than the internal combustion engine and if well-displayed, the recharging process for an electric bus is more efficient and safer than the gasoline one in terms of reliability. The feature of regenerative braking must not be forgotten as it represents a great advantage compared to the conventional buses which are not able to recover energy braking. All these advantages combined confirm that if a modernization of the bus lines is needed, one better option will be to replace the gasoline buses with electric ones.

All the parameters value necessary for the simulation are reported in Table I. Using an Excel Spreadsheet, the overall electric energy consumed by the vehicle along the chosen sections was estimated.

Table 1: Simulation values

l	Bus length	12 m
m_v	Mass vehicle	15 800 kg
n_{pass}^{max}	Maximum number of passengers	85
m_{pass}	Average mass of passengers	70 kg
a_{br}^{max}	Maximum comfortable deceleration	-1.5 m/s^2
a_{max}	Maximum comfortable acceleration	0.8 m/s^2
P_e^{max}	Maximum motor power	230 kW
V	Battery voltage	400 V
η_{RBS}	Percentage of recoverable energy	65%
β	Equivalent mass factor	0.16
η_B	Braking efficiency	80%
η_T	Traction efficiency	80%
F_T^{max}	Maximum tractive effort	18.56 kN
P_{aux}	Power absorbed by auxiliaries	22 kW
C_x	Shape coefficient	0.6
m	Overall mass of the vehicle (considering 60 passengers)	20 000 kg
A	Frontal area	8.7 m^2
p	Tyre pressure	9 bar

3. Energy consumption estimation

The considered Line 6, depicted in Figure 2, connects the bus stations in Porto Vado to "Via Alessandria" in Savona for an overall trip 7.9 km long. While the first part of the line is passing through the city center, the other and longer part of the line heads to the west following the seafront of Savona. For the energy consumption estimation, two representative and very different sections in terms of ground characteristics have been analyzed. We will refer to the first one (between the yellow pointers) as Section A and to the second one (between the red pointers) as Section B.
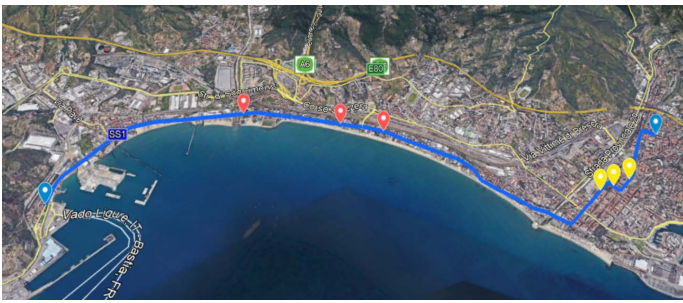
Figure 2: Bus Line 6: From Porto Vado to Via Alessandria in Savona (Italy).

3.1. Analysis of Section A

The section analyzed in the following paragraph is an extract from the downhill part of Line 6 TPL, starting from Via Paolo Boselli and reaching its destination of "Via XX Settembre", with an intermediate stop in "Piazza del Popolo" after 300 m. In the 700 m long section, some crossings and turnings are present. Therefore, to make the simulation more realistic and consequently much more consistent and reliable, some braking sections are considered in correspondence of pedestrians crossing and traffic lights. For instance, as can be seen from the speed profile reported in Figure 4, 3 deceleration and 2 complete stops are simulated along the section. The first deceleration is performed after 140 m from the beginning, and it is due to the fact that the bus must decrease its speed to perform a right turn. Then it accelerates until it reaches a speed of about 40 km/h but at that moment another deceleration phase starts after 270 m because the bus must stop at the intermediate station of "Piazza del Popolo". The stopping time at this station has been set equal to 20 sec. After few seconds from the restart, the bus must stop again at a red traffic light. This second stopping time is set equal to 5 seconds. Finally, in the last part of Section A, the driver can reach the maximum speed of 50 km/h and keep it for a few seconds before arriving at the final stop of this section.
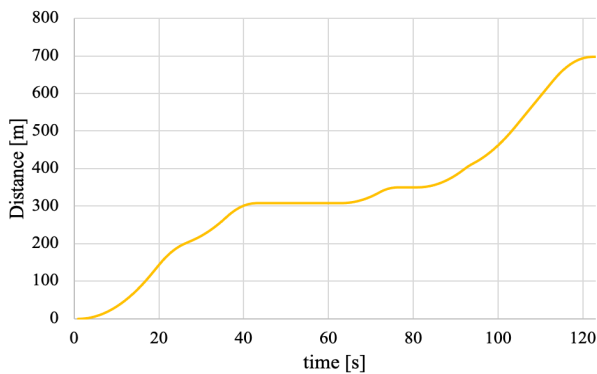
Figure 3: Space Profile of Section A

In Figure 3 the space profile of the section is shown, and we can see that the total time spent to complete the path is equal to 123 seconds.

In Figure 5 the evolution of the electric power as a function of time, from which the electric energy consumption is derived. As expected, in the first part electric power increases up to its maximum value since at this moment the motor must provide the highest tractive effort to let the vehicle reach the speed limit of 50 km/h. Instead, negative values of electric power indicate that at that moment the regenerative braking phase is ongoing.

(a)

(b)
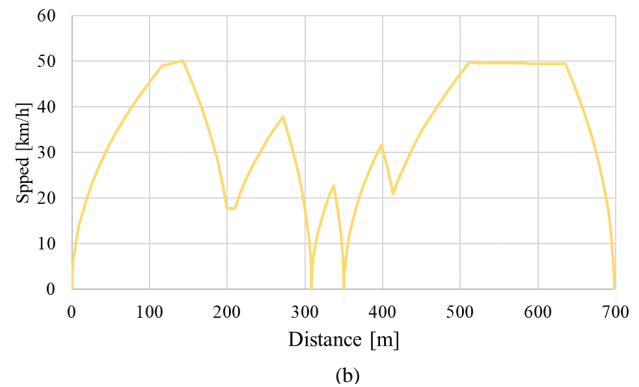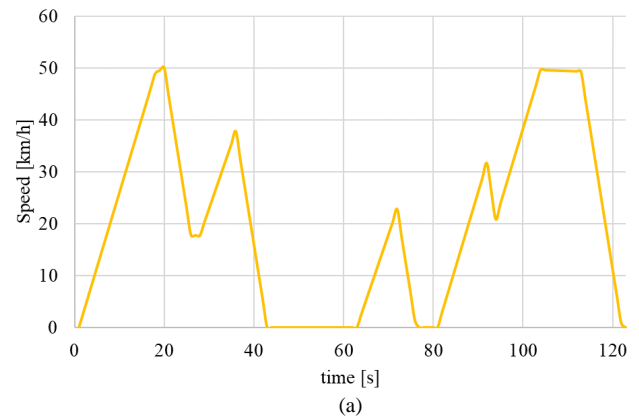
Figure 4: Speed Profile of Section A a) speed vs time and b) speed vs distance.
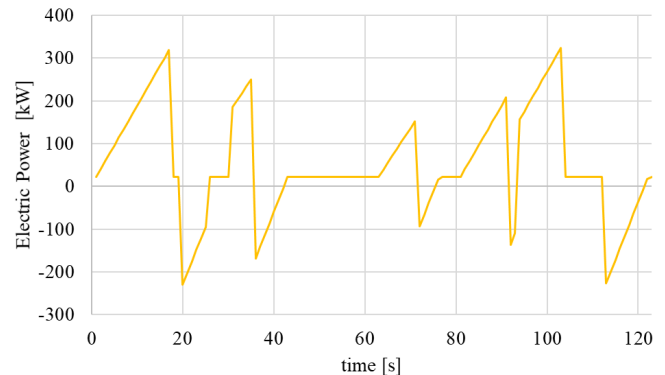
Figure 5: Electric Power profile of Section A
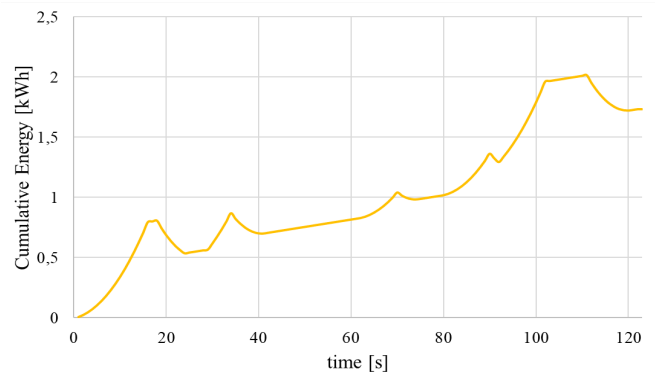
Figure 6: Cumulative energy consumption in Section A.

Finally, the total energy required by the bus is reported in Figure 6 and it can be concluded that to cover the 700 m related to

Section A, 1.73 kWh are necessary, which correspond to consumption of about 2.47 kWh per km, in line with that find in [13].

3.2. Analysis of Section B

The second considered section of Line 6 TPL goes from the bus stop Villa Zanelli to Via Nizza (Camping) until the final destination Via Nizza 126. This path differs from section A primarily because of its location along the seaside, on Strada Statale 1 Via Aurelia (SS 1), instead of the city center. Consequently, Section B is flatter and characterized by a straight path on which, however, the speed limit is 50 km/h, since it is always an urban road [14]. The whole Section B has a total length of 1,450 m and the middle bus station Via Nizza (Camping) is located after 270 m starting from the first bus station Villa Zanelli.

The whole Section B lays on a priority road with no traffic lights intersections but with two rotaries and several crosswalks. Within the model, a deceleration down to 20 km/h is assumed before entering the rotaries that are located after 200 m and 968 m starting from the first bus stop 'Villa Zanelli'. In the whole path of Section B, there are ten crosswalks on which the bus has to decelerate in case of a pedestrian wants to cross the street. However, since the path does not have any curves, the driver of the vehicle has a wide forward view on the road, therefore, in the computation, it is not assumed to stop on every crosswalk but only to slow down to 10 km/h or less. Finally, before the final bus stop, the bus must pass a fork in the road. However, as the bus has the right of way the intersection is not taken further into account.

a)

b)

Figure 7: Speed profile in Section B a) speed vs time and b) speed vs distance.

An approximated gradient profile is included in the calculation, even though the slope is very low. The complete profile is generated by Google Earth and then is approximated with a

positive value of 0.23% for the first 1300 m and a negative value (which represent a downhill) of 0.67% for the resting 300 m. Finally, the bus is assumed to stop at 'Via Nizza (Camping)' for 20 seconds.

From the results reported from Figure 7 to Figure 9, it is possible to conclude that the bus needs 3 min and 34 seconds to cover this section. In particular, the bus reaches Via Nizza (Camping) after 1 min and 11, and after a 20 second stop, it takes another 2 minutes to run the last part. The speed profile in Figure 7 is characterized by many peaks, which means many changes in velocity caused by the rotaries and crossings the bus is passing by. The maximum speed of 50 km/h is reached three times.

Overall it is visible, that due to the number of crossings and rotaries, the bus does not have the driving distance in between the traffic interactions to accelerate much further than 50 km/h. Regarding the fact that the road is located close to the beach and right next to camping opportunities, it can be assumed that it is a rather touristic area that many walkers pass. Additionally, the fact that there is a cycle path right next to the road clarifies that driving faster than 50 km/h would bring dangers with it.

As Figure 8 shows, the electric power taken by the supply system (the battery) is strongly varying over the path. The fluctuation can be traced back to the situation, whether the bus is in traction (tractive effort > 0) or braking (tractive effort < 0) – as it can be observed in Figure 9. If the bus is in traction mode, the supply system is providing not only the needed power for the wheels but also the power absorbed by the auxiliaries. The bus is always consuming the power needed for its climate system and other extra facilities, even though the tractive effort is zero or the bus is not moving (speed = 0 km/h). The electric power assumes negative values when the bus is braking because in those moments a percentage of energy is sent back from the wheels to the battery thanks to the regenerative braking technology. However, due to technical limitations, there is no regenerative braking when electric vehicles are running at low speed; in this study, this limit is set at 10 km/h.
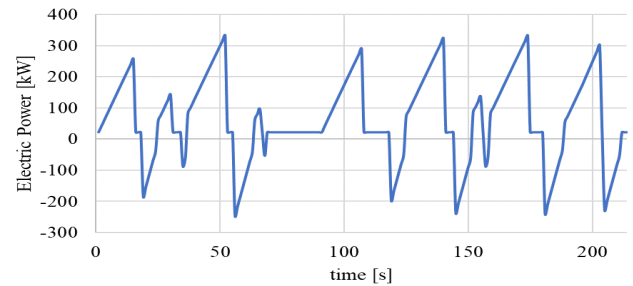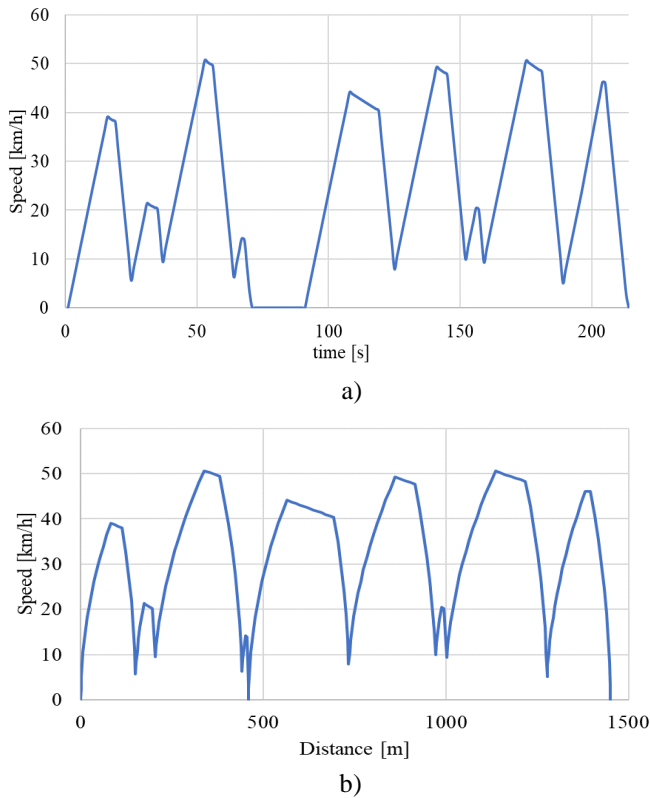
Figure 8 Electric power along Section B.
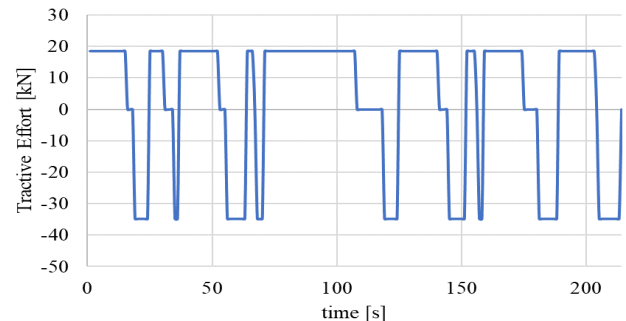
Figure 9: Tractive effort along Section B.

As can be seen from Figure 10, the overall energy consumption along Section B is 3.8 kWh which corresponds to 2.62 kWh per km. Compared to the average energy consumption of 2 kWh/km found in [15], it is a quite higher result which can be justified by the presence of many influencing factors. First of all, as already mentioned, the bus is consuming power, even when it is stopped at the bus station because it has to supply the auxiliaries (22 kW). For instance, the 20 seconds stopping time at the station Via Nizza (Camping) results in power consumption of 0.12 kWh. Additionally, the high number of crossings and rotaries present along the path forces the bus to continuously decelerate and accelerate. The high power required to accelerate leads to high energy consumption. A smoother driving style that reduces the deceleration and acceleration rates could lead to energy savings.

In order to extrapolate the energy consumption of the entire line, a weighted average of the two energy consumptions found for Section A and Section B is computed. Since the line is prevalently located along the seaside rather than in the center of the city, for the computation of the overall consumption Section B weights more (75%) than section A (25%). Therefore, the total energy consumed along the 7.9 km trip between Porto Vado and Via Alessandria is estimated to be about 20.3 kWh. The results coming from this section are summarized in Table.
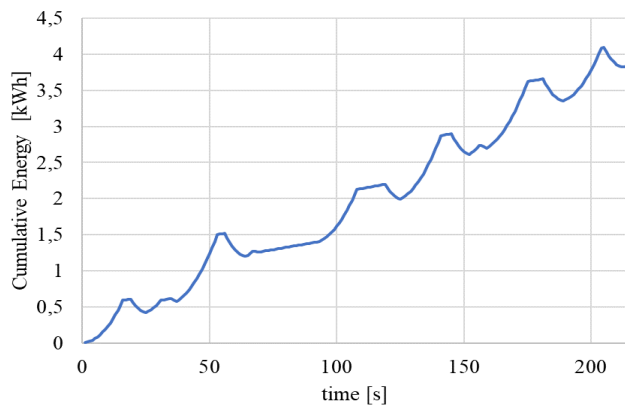
Figure 10: Cumulative energy consumption at section B.

Table 2: Average energy consumption

Section	Length	Request [kWh/km]	% overall path	Weighted Average [kWh/km]
A	700	2.47	25%	2.57
B	1 450	2.62	75%	

4. Charging Systems Sizing

An electric bus can be charged with two main charging technologies: depot and opportunity charging. Depot charging technology fills the fleet of electric buses at the depot through plug-in connectors [16]. The charging power for each connector ranges from 40 up to 150 kW. The rated power of this charging system mainly depends on whether AC or DC charging system is installed. If an AC technology is chosen, the onboard charger of the vehicle must be used to operate the AC-DC conversion, and the off-board structure will only include power and communication cables, metering, and protection devices. AC charging systems can provide a maximum power of 43 kW (86 kW if two Type2 plugs are inserted in the vehicle). If higher charging rates are necessary or if the electric buses which compose the fleet are not equipped

with the onboard charger, a DC charging system (mode 4) must be used. In this case, the conversion stage is performed offboard the vehicle, inside the charging system. On one side this will result in higher achievable power rates on the other hand this will lead to higher investment costs. In DC charging systems, the power limit is mainly imposed by the cable and the connector; as a matter of fact, the Combo 2 (CCS2) connector can provide a maximum of 200 A without the need for special and expensive liquid cooling systems. To not have a too high impact on the public utility grid, the output current is limited to 150 A, which corresponds to a nominal power of about 100 kW considering the average battery voltage of the actual electric buses on the market. It can occur that to complete and provide a proper service electric buses require additional occasional recharges that take advantage of the halt times at the terminus and/or at the stops. The charging system which provides this type of facility takes the name of opportunity charging. The opportunity charging uses overhead pantographs which can support charging powers up to $600 \div 750$ kW [17].

In this paragraph, the two charging technologies for the chosen electric buses of Line 6 TPL will be sized. The most important data necessary to display the analysis are the total energy consumption on the entire line, its timetable, and the battery and charging characteristics of the vehicle.

Following the information found on the timetable of Line 6 TPL, it can be understood that to ensure the service 3 buses must be used at the same time. In the following, we will refer to these three buses as Bus 1, Bus 2, and Bus 3. All the buses start and end in the two stations of Via Alessandria and Porto Vado, however, some of them perform small deviations along the day. The total number of one-way trips between Porto Vado and Via Alessandria for these three buses are respectively equal to 32, 30, and 25.

The eCitaro buses are provided with lithium-ion batteries having different levels of battery capacity: 150, 200, and 250 kWh, in fact, they are made of a variable number of modules ranging from 6 up to 10. The standard charging system for eCitaro buses is the Combo2 plug-in system, but there is also the possibility to adopt the pantograph system which is sold as special equipment.

4.1. Depot charging

In this section, the charging system that must be installed in the depot in order to assures a correct service of Line 6 is analyzed. An example of the layout that the depot charging infrastructure can assume is shown in Figure 11 b). In Figure 11 a), instead, the energy distribution scheme is sketched.

TPL already rents a depot in a public area installed in Via Valletta San Cristoforo, Savona, approximately located in the midpoint of the trip. Therefore, this depot will be the location of the charging system. Since the depot is located in a public area, some technical issues are to be discussed properly [13]. First, it must be considered that other large consumers may also take energy from the same grid thus limiting the amount of energy that can be used to recharge the fleet. Then, the European Union stated restrictions on the maximum power to be installed in a charging depot and this threshold cannot be overcome for legal reasons. Consequently, in the depot, it will be installed an AC charging system with rated power for each port equal to 43 kW.

To not oversize the number of required charging ports, and hence to not waste money, the schedule of all the electric buses recharging times becomes a crucial feature of this study. In order to size the depot charging, the following assumptions are made:

1. All the electric buses of the fleet made have the same battery capacity which in the first case is set at 250 kWh, then to 200 kWh, and finally to 150 kWh.

2. A safety SoC threshold has been set at 20%, and this amount of energy is necessary for example to return to the depot in case of emergency conditions.

3. All buses leave the depot fully charged (SoC 100%).

4. The buses rely only on the depot charging for the operation.

(a)
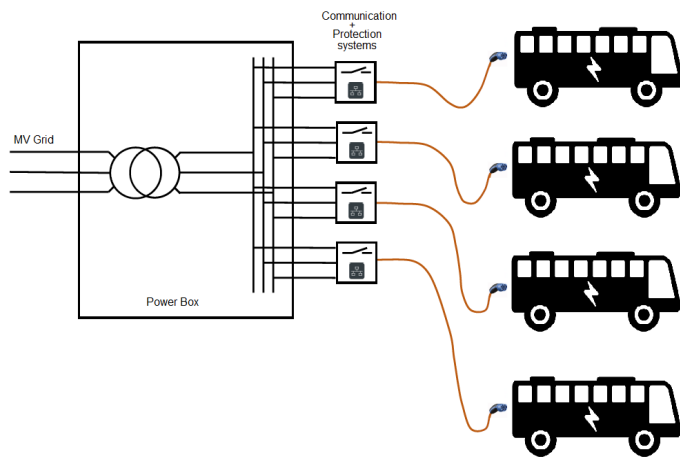
(b)

Figure 11: Example of depot charging infrastructure a) layout and b) electric scheme.

CASE 1: The buses are equipped with a 250 kWh battery. Under the before-mentioned assumptions, a single bus is able to do nine one-way trips before reaching a battery percentage level below the threshold of 20%. As a bus can approximately drive four hours before having to drive back to the depot, another bus can use this time for charging its battery. Within that time, the battery level of the bus can reach values up to 80%, considering that the bus arrives at the depot having 20% of battery and the depot has a recharging system of 43 kW with a charging efficiency equal to 88%. So, after the second bus has covered its part of the runs, the bus that was waiting at the depot becomes the one in service while the other heads to the depot to recharge its batteries. By keeping this alternate on during the whole day, two electric buses for each currently employed diesel bus are needed, so the fleet will be

composed of six electric buses in total. Three charging ports must be installed in the depot, one for every two electric buses.

CASE 2: The battery capacity is decreased at 200 kWh. In this case, a bus can cover only seven trips without reaching the level of 20% of battery and according to the timetable, this means a service time equal to three hours. By repeating the same calculations as in the previous case, the fleet will be composed of thirteen buses. If instead the daytime charging at the depot is allowed, the number of buses needed will decrease as described. In three hours, the battery of 200 kWh can recharge to a level of 80% by making the same assumption made before for a battery of 250 kWh. So, after the second block of three hours, the two buses overlap and this continuous cycle is repeated throughout all the daytime, leading to the conclusion that again two electric buses for each currently employed diesel bus are needed. Therefore, even using electric buses equipped with batteries having a capacity of 200 kWh, the electric fleet must be composed of six vehicles.

CASE 3: finally, the case of batteries with a capacity of 150 kWh is brought into the analysis. In this case, a single bus can cover five trips with a corresponding amount of time of two hours. Therefore, a bus can recharge up to 70% of its battery at the 43kW charging pole at the depot with an efficiency of 88%. Since the battery capacity is very low, the overlapping of two buses is not fitting for the case, as the iterative decrease of the battery percentage during the service for the two buses leads to an incompletion of the daytime service (two buses are able to cover only 65% of the daytime service). For this reason, three buses are needed, with two buses recharging simultaneously while the other one is providing the service. In this case, the buses can always run into service having a 100% battery capacity, thanks to the fact that having two buses waiting in the depot, the overlapping of one of the two with the exhausted one enables the second bus to further recharge its batteries for supplementary two hours. By extending this observation to the three currently employed buses, the result is an electric fleet composed of nine buses, which means 3 electric buses for each diesel one.

Table 2: Results depot charging

#Case	Battery Capacity	# necessary e-buses	# installed 43 kW charging ports
Case 1	250 kWh	6	3
Case 2	200 kWh	6	3
Case 3	150 kWh	9	3

In conclusion, the battery capacity has a significant impact on the number of buses needed to substitute a conventional bus if the electrification relies only on the depot charging technology.

4.2. Opportunity charging

As an alternative to the traditional depot charging system, opportunity charging is gaining in popularity. The difference between the two charging technologies can be simplified in a way that opportunity charging is taking place during the service at on-street charging stations [18]. Given the higher power level, this charging system can provide, it follows that the charging duration gets significantly shortened [17]. The main components of an opportunity charging system are sketched in Figure 12.

Regarding the timetable of the 3 routes of Line 6 TPL, the vehicles are having regular breaks between five and twenty minutes at the bus station Porto Vado. The parking area in that bus

station is spacious, and hence the buses can take their break without disturbing the traffic. Therefore, the installation of a pantograph for opportunity charging at this place is technically possible. To electrify the routes, without passengers noticing a change in the schedule of Line 6 TPL, the break at Porto Vado is consequently chosen for opportunity charging.
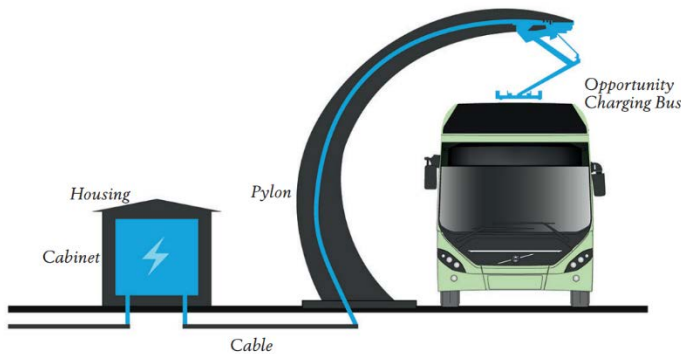
Figure 12: Example of opportunity charging infrastructure.

To determine the necessary battery capacity of the bus as well as the charging power of the opportunity charging station, the timetables of the three routes are evaluated. In particular, the overall timetable is characterized by regular stops at the station Porto Vado for 10 minutes. Between two consecutive stopping times in Porto Vado, therefore the vehicle must perform a round trip for an overall length of 15.8 km (2 times 7.9 km). Based on the energy consumption coming from the previous section, to cover this path the bus consumes 40.6 kWh (2 times 20.3 kWh).

The charging time is determined by the breaking times scheduled through the timetable of Line 6 TPL. This time is assumed to be 2 min shorter than the scheduled one in order to take into account the connection and the disconnection maneuvers to connect to the pantograph. Therefore, the bus charges every time it reaches Porto Vado for an average time of 8 minutes (10 min minus 2 min as a buffer). Therefore, in this time interval, the charging system must provide the bus enough energy to manage the next trip without exceeding the minimum battery SoC limit of 20%. The opportunity charging system can usually have different power rated: 150 kW, 300 kW, 450 kW up to 600 kW [17]. Considering the available charging time in Porto Vado t_{ch}, and the required energy to cover a round trip E_{cons}, in (12) the necessary charging power is computed. The result shows that a pantograph of at least 300kW is needed.

$$P_{ch} = \frac{E_{cons}}{t_{ch}} = \frac{40\ kWh}{8\ min} \approx 300\ kW \qquad (12)$$

Looking at the overall path of the buses operating on line 6 reported in [19], as before mentioned, the three conventional buses must make some small deviations from the usual path. However, all the three buses during their operation stop at Porto Vado station for some minutes, and this interval of time is used to charge the battery through the installed opportunity charger. Figure 13 illustrates the SoC of each of these three buses as a function of the covered distances. Each vertical upwards course represents a charging operation at Porto Vado. As the halt time is most of the time 10 minutes long, the trend of the battery SoC is characterized by a regular pattern due to the same amount of energy supplied in each charging event. While the charging-discharging cycle of the Bus 3 is repetitive, Bus 2 is showing more irregularities. Those can be traced back to the higher number of deviations that this bus has

to perform, which will cause more irregularities in the stopping times at Porto Vado. For instance, Bus 2 is facing the absence of stops in Porto Vado in the evening runs, so the longer stop it has after 250 km is used to counterbalance the lack of charging events from 213 km up to 244 km. Bus 1, instead, is facing shorter breaks in the morning hours. To ensure sufficient SoC, Bus 1 must completely charge its battery at Porto Vado before starting its last trip in the evening. Finally, Bus 3, which is starting its journey from Porto Vado in the morning has to arrive at the bus station around 15 minutes earlier in order to completely fill its battery. The extra time needed for charging at the Pantograph at Porto Vado as well as driving there off-schedule causes extra costs for personnel that needs to be considered. For the calculations, it is, in addition, assumed that all the buses during night drive to their depot at Via Valletta San Cristoforo, so the extra energy necessary to cover this path to go to the depot has been considered.
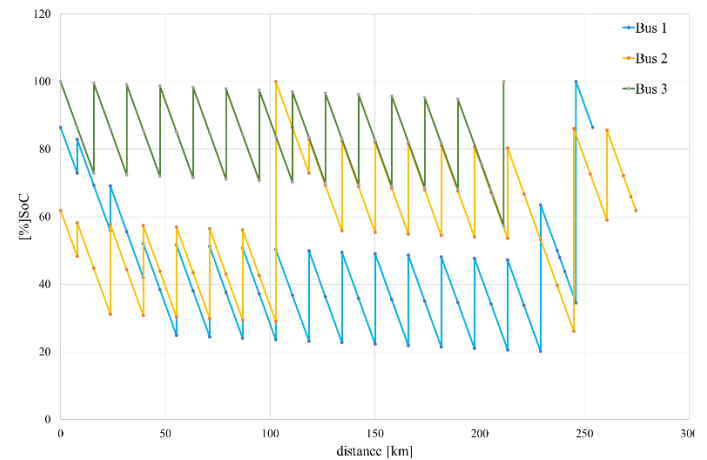
Figure 13: Battery Charge Level of the buses on the investigated routes.

The eCitaro is available in different versions and with different battery capacities. As Figure 13 shown, the basic model (150kWh) of the lithium-ion battery is already sufficient to meet the requirements of Line 6 TPL. As well it can be observed that according to the results the battery SoC is never lower than 20 %. So, the opportunity charging system fulfills the set safety requirements.

Concluding, it is technically achievable to fulfill the timetable using only one opportunity charging station with a capacity of 300kW at Porto Vado in combination with the eCitaro equipped with a standard battery of 150 kWh.

5. Conclusion

This work aims to analyze the electrification of a conventionally operated bus line. Precisely, the focus is on the bus Line 6 operating between the stations Via Alessandria and Porto Vado in Savona, Italy. Therefore, the overall energy consumption on this specific bus route is first computed and then used to evaluate possible depot and opportunity charging infrastructure.

The model results show that it is possible to substitute the conventional buses of Line 6 TPL connecting Porto Vado and Via Alessandria with electric buses and corresponding charging infrastructure. Based on the energy consumption of two characteristically different sections on the assigned Line 6 TPL path, the overall energy consumption of an eCitaro shuttling between Porto Vado and Via Alessandria is upscaled. The

calculations lead to the energy consumption of 20.3 kWh for the 7.9 km long one-way trip between Porto Vado and Via Alessandria. Corresponding the electric bus has an average energy consumption of 2.57 kWh/km.

In order to stick to the present timetable of the chosen routes, there are three possible solutions for electrification. According to the first approach, the conventional buses could be replaced by e-buses using depot charging technology. The depot charging system is composed of three charging ports of 43 kW each. From the results, it can be seen that this scenario would imply that each conventional bus needs to be replaced by at least two electric ones with a battery capacity of 250 kWh. As an alternative to reduce the number of needed vehicles, opportunity charging is introduced as the second possible electrification approach. Based on the halt times available at the terminals in the timetable of Line 6 and the energy consumption along the route, the results prove that it is possible to substitute each conventional bus with one e-bus exclusively relying on opportunity charging systems installed in Porto Vado terminal with a power rate of at least 300 kW. Merging both approaches leads to a third possible solution which could ensure sufficiently charged batteries, increase the safety against failures as well as the efficiency in operation. The analysis of this mixed scenario as well as the cost analysis of the different solutions will be part of further research.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Adoption of the Paris Agreement, Paris, 2015.
[2] Regulation (EU) 2019/631 of the European Parliament and of the Council, 2019.
[3] IEA, Global Ev Outlook 2019 – Analysis - IEA, Paris, 2019.
[4] The Impact of Electric Buses On Urban Life, 2019.
[5] S. Majumder, K. De, P. Kumar, B. Sengupta, P.K. Biswas, "Techno-commercial analysis of sustainable E-bus-based public transit systems: An Indian case study," Renewable and Sustainable Energy Reviews, 144, 111033, 2021, doi:10.1016/J.RSER.2021.111033.
[6] X. Zhao, Y. Ye, J. Ma, P. Shi, H. Chen, "Construction of electric vehicle driving cycle for studying electric vehicle energy consumption and equivalent emissions," Environmental Science and Pollution Research 2020 27:30, 27(30), 37395–37409, 2020, doi:10.1007/S11356-020-09094-4.
[7] H. He, J. Guo, N. Zhou, C. Sun, J. Peng, "Freeway Driving Cycle Construction Based on Real-Time Traffic Information and Global Optimal Energy Management for Plug-In Hybrid Electric Vehicles," Energies 2017, Vol. 10, Page 1796, 10(11), 1796, 2017, doi:10.3390/EN10111796.
[8] X. Ma, R. Miao, X. Wu, X. Liu, "Examining influential factors on the energy consumption of electric and diesel buses: A data-driven analysis of large-scale public transit network in Beijing," Energy, 216, 2021, doi:10.1016/J.ENERGY.2020.119196.
[9] M. Brenna, F. Foiadelli, D. Zaninelli, "Electrical railway transportation systems."
[10] P. Fajri, R. Ahmadi, M. Ferdowsi, "Equivalent vehicle rotational inertia used for electric vehicle test bench dynamic studies," IECON Proceedings (Industrial Electronics Conference), 4115–4120, 2012, doi:10.1109/IECON.2012.6389231.
[11] eCitaro – Mercedes-Benz Buses, Sep. 2021.
[12] Export success for the Mercedes Benz eCitaro with fully electric drive: orders from Luxembourg, Norway and Sweden - Daimler Global Media Site, Sep. 2021.
[13] C. Leone, M. Longo, F. Foiadelli, S. Bracco, G. Piazza, F. Delfino, "Opportunity fast-charging of e-buses: A preliminary study for the city of Savona," 2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive, AEIT AUTOMOTIVE 2020, 2020, doi:10.23919/AEITAUTOMOTIVE50086.2020.9307434.
[14] Going abroad - European Commission, Sep. 2021.
[15] S. Ceccon, M. Longo, R. Mazzoncini, A. Panarese, "Analysis of the Implementation of Full Electric and Hydrogen Hybrid Buses in Two Lines of the City of Milan," in 2020 15th International Conference on Ecological Vehicles and Renewable Energies, EVER 2020, Institute of Electrical and Electronics Engineers Inc., 2020, doi:10.1109/EVER48776.2020.9243074.
[16] S.M. Arif, T.T. Lie, B.C. Seet, S.M. Ahsan, H.A. Khan, "Plug-In Electric Bus Depot Charging with PV and ESS and Their Impact on LV Feeder," Energies 13(9), 2139, 2020, doi:10.3390/EN13092139.
[17] M.M. Hasan, M. Ranta, M. El Baghdadi, O. Hegazy, "Charging management strategy using ECO-charging for electric bus fleets in cities," 2020 IEEE Vehicle Power and Propulsion Conference, VPPC 2020 - Proceedings, 2020, doi:10.1109/VPPC49601.2020.9330970.
[18] O. Barraza, M. Estrada, "Battery Electric Bus Network: Efficient Design and Cost Comparison of Different Powertrains," Sustainability 13(9), 4745, 2021, doi:10.3390/SU13094745.
[19] LINEA 6 - LINEA 6/ Timetable, Sep. 2021.

Design Optimization and Life Cycle Cost Assessment of GRC Shading Screens for Office Buildings in Cairo

Ghada Shedid*, Osama Tolba, Sherif Ezzeldin

Department of Architectural Engineering and Environmental Design, Arab Academy for Science, Technology and Maritime Transport, El Moshir Ahmed Ismail street, Cairo 2033, Elhorria, Egypt

A R T I C L E   I N F O

A B S T R A C T

Office buildings commonly use fully glazed façades to reflect a luxurious appearance and to maximize natural light of high solar exposure and high-energy consumption due to cooling and heating. There is a great abundance in constructing shading screens as they are part of the modern movement in field of Energy Conservation, renewable energies, and architectural design. This paper studies the impact of various perforated Glass Reinforced Concrete (GRC) shading screens for different orientation and the Life Cycle Cost Assessment (LCCA) in a prototypical office space in Cairo. We have simulated a wide range of perforated shading screens using Design Builder to identify optimal shading screens with the highest energy savings for different façade orientation. In this paper, we suggest a methodology to achieve better energy saving in office buildings, knowing the façade's orientation and perforation percentage of shading screen. Simulation results show that shading screens with 10% perforation percentage commonly achieve the highest energy savings for all façade orientations and reaches up to 53% energy savings for the southwest façade. The LCCA of this shading screen in that southwest façade saves 52% in LCCA compared to the base case.

1. Introduction

Contemporary buildings in the world require massive amounts of energy for cooling and heating, while the cost of electrical energy is ceaselessly increasing. The amount of energy needed for providing comfortable living conditions within a specific region depends on climate conditions [1]. Based on Köppen's climate classification system, Egypt's climate is classified as hot desert arid climate, characterized by high direct solar radiation and clear skies [2]. These climate characteristics demand special façade treatments to minimize heat gain, solar exposure and energy consumption.

Shading screens, as a type of shading devices, help in reducing glare and excessive solar heat gain through fenestration. When properly designed, shading screens will improve not only indoor visual and thermal comfort, but also will greatly cut back electrical lighting, cooling, and heating energy [3].

Commercial buildings consume energy around 5,000 GWh, which represents nearly 4% of all electricity consumed in Egypt.

Around 60% of energy consumption in commercial building is due to HVAC and lighting. Façade configurations account for up to 45% of the cooling loads of the building [4]. Heat gain is transmitted through building façades either by radiation through

transparent surfaces or by conduction through the whole building envelope.

The Life Cycle Cost Assessment (LCCA) primary goal is to provide a tool with the ability for appropriate financial appraisal of buildings and to substitute existing approaches focused on the actual costs of the construction project [5].

2. Research Objective

The objective of this research is to understand the value of integrating perforated GRC shading screens for different orientations of office building façades in Cairo using energy simulation and LCCA. Previous related studies focused on either predicting energy savings as a result of using shading screens for different facades or LCCA of different facade components other than shading screens. Therefore, the paper considers both energy savings and LCCA for the assessment of shading screens. In addition, we chose GRC as the material of the shading screen in order to minimize the initial cost.

3. Literature review

3.1. Types of shading screen

In Egypt, shading screens have been widely used for centuries. The *Mashrabiya*, an example of shading screen is used for controlling the passage of light, regulating air flow, reducing the

*Corresponding Author: Ghada Shedid , Architectural department , AAST, Cairo, Egypt E-mails: eng.ghada.ashraf@gmail.com

temperature of the air current, increasing air humidity and guaranteeing a great amount of privacy [6]. Nowadays, designers allowed addressing the design exploration at both the macro and micro levels of resolution by the application of parametric modeling to architectural design. Figure 1 shows a type of shading screen that allows engineers or fabricators to immediately have input into the modelling process, to ensure structural integrity, while analyzing the resulting impact on the component form [7].
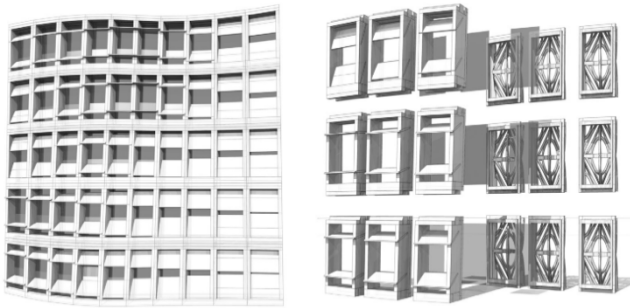
Figure 1: Exploded view of individual elements [7]

In [8], The authors evaluated various types and models of shading treatments based on used evidence review of systematic method to identify patterns and trends through classification and comparison. Three main categories of shading treatments were identified based on the energy involvement and the design approach: (i) passive systems with zero energy use, (ii) active systems that use mechanical devices and (iii) hybrid systems integrated with a biomimetic approach. Passive systems are fixed shading devices, which are cost-effective and easy to implement. Their performance presented limitations to address changeable weather conditions throughout the year.

The most common materials for shading screens are wood, aluminum and GRC [9]. The GRC is considered a durable, flexible and a cost effective material for shading screens.

3.2. Shading screens varieties and orientations

In hot climates, building energy use for lighting, heating and cooling, as well as the occupants' visual and thermal comfort are affected by external perforated shading screens [10]. Studies have explored screen geometries, perforation percentages, and orientations in hot climate zone. Examples of these perforated screens are Mashrabeya, Jali, and Guss-Walls. The implementation of these screen solutions was to mitigate excessive solar radiation on window openings that result in overheating and therefore reducing cooling loads. [11].

In [12], The authors assessed the effect of shading screens with different rotation angles, aspect to depth ratio and perforation percentages on the annual cooling, heating and lighting energy loads using Energy Plus (Figure 2). The study proved that external fixed deep perforated solar screens could effectively achieve energy savings up to 30% of the total energy consumption in the west and south orientations. The highest and most significant savings were achieved in screens having a depth ratio of 1:1depth/opening width ratio.

This study shows that the highest energy consumption is on the west and south in which energy saving could reach the highest

level by using a shading screen and the shape of the shading screen use is the orthogonal grid screens with square fenestrations with 1:1 depth ratio.
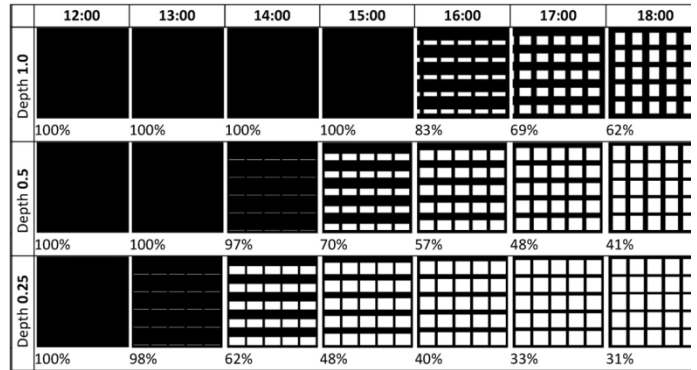
Figure 2: Shadow patterns and resultant shading percentage of shading screens with different depth ratio [12].

3.3. Shading screens and energy consumption in office buildings

In [13], The authors examined a wide range of perforation percentages of shading screen in Kharga Oasis for different orientations using Energy Plus simulation software to recommend the range of shading screens perforation percentages that reduces monthly and annual cooling loads. Simulation results showed that shading screen with perforation rate of 90% reduces cooling energy by 25% in average, while using more dense screens with perforation decreasing from 90% to 10% saves about 16% more, giving a total saving of 41% in average of the west, south and east façade.

This study shows that as we decrease the opening perforation we get high energy saving, and insure that the west and south facades have the highest energy consumption.

3.4. LCCA estimation

LCCA is determined by initial manufacturing, repair and replacement costs of the residual value of the commodity at the end of the study period, the sum of all these costs except the residual value determining LCC (1) defined by [14]

$$LCC= \text{INTIAL COST} + \text{OPERATING COST} + \text{MAINTAINANCE COST} - \text{RESIDUAL VALUE} \quad (1)$$

3.5. Previous studies of LCCA of building façade

In [15], The authors studied the value of kinetic aluminum louvers installed for office building south façades in a hot-dry climate, such as that of Egypt. The study compared kinetic aluminum louvers, fixed aluminum louvers to the non-shaded cases in terms of energy consumption and building Life-Cycle Cost. LCCA, using equation 1, finds out that fixed louvers are the most cost effective than kinetic louvers (Figure 3).The study shows the variation in both initial and operation cost but the difference in the initial cost is larger than the difference in the operation cost. The main challenge facing the application of kinetic louvers in office buildings in Egypt is the high initial cost due to importing the aluminum louvers and the additional automation components. This makes the payback period of kinetic

louvers in Egypt longer than 20 years. Alternative scenarios of using local manufacturer of fixed louvers achieve net savings and cost payback just before the end of the study period. This study shows that the fixed shading screen are more affordable than the kinetic louvers.
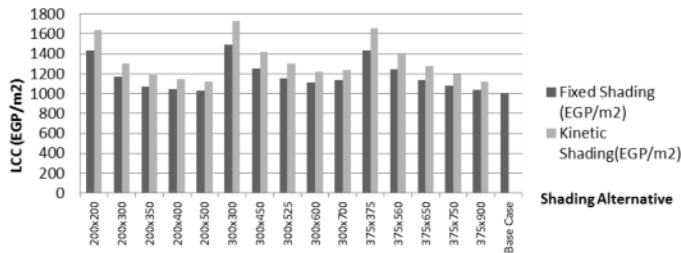
Figure 3: LCCA of kinetic and fixed louvers [15]

From the literature review, preferences using fixed shading screen with orthogonal grid screens with square fenestrations having high-density perforation in the west and south are the optimum for simulation.

4. Methodology

Simulation was performed for using Design Builder version 4.5 as a graphical user interface for energy plus. The base case model, the shading screen modeling and parameters are discussed below.

4.1. Base Case Model

The base case is a typical cellular office space in the hot arid desert environment of Cairo, International airport, Egypt, a closed office having a floor area of 4 m by 4 m, with a height of 3 m (Figure 4). This room was isolated from the external environmental thermal changes by assuming that all surfaces, other than the external tested wall, were adiabatic. Thus, three walls, floor and roof were assumed adiabatic. The fourth wall was defined as a fully glazed curtain wall and defined as double clear 6mm glass and 13mm air gap.
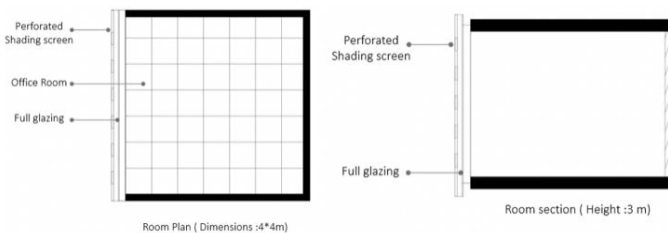
Figure 4: Office room plan and section show the relation between curtain wall and shading screen

Typical occupancy schedule was set to start at 8 am and end by 5 pm with a two-day weekend break. Mechanical cooling was simulated using an A/C chilled water system. Cooling and heating set points were 26°C and 20°C respectively, which achieve the required thermal comfort, set by ASHRAE. Artificial lighting was set to use LED bulbs (10 W/m²), which is considered as optimum lighting efficiency for office spaces. The annual simulation runs were conducted for west, southwest, south, southeast and east facade orientations and testing the perforation of shading screens from 5% to 90%.

4.2. Shading screen modeling approach

In order to find the screen configuration with highest energy saving potential, a range of shading screen perforation was simulated. The main focus of the simulation was to evaluate the effect of alternative configurations on energy consumption due to cooling, heating and lighting. The shading screen was externally mounted at a distance of 20 cm from the curtain wall. The screen dimensions cover the whole room's façade with extended 1min each side.

Alternatives of different perforation percentages were studied for Curtain walls facing west, southwest, south, southeast and east orientations. We performed the study on screens with perforation percentages ranging from 5% to 90%, and depth of shading screen 5 cm. The number of simulation runs reached 55 runs.

4.3. Shading screen parameters

- Material: GRC

- Dimensions: Width 4 m and height 3 m and extended to cover the office façade.

- Position: At a distance 20 cm from the curtain wall.

- Shape: Orthogonal grid 90 degree of rectangular section bars.

- Perforation Ratio: Ranging from 5% to 90% (Figure 5).

- Material Depth: 5 cm

- Fixed number of opening: 5 rows and 6 columns

- Orientation: west, southwest, south, southeast and east façades.
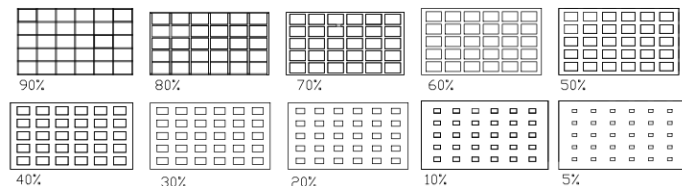
Figure 5: Shading screen used in simulation from 90% to 5% perforation

5. Annual energy consumption of alternative screen configurations

The annual energy consumption of the base cases was calculated first. It was a closed office having all the above defined parameters, with daylighting controls and without any shading screens. Simulation results of the base case revealed that the highest energy consumption was in the southwest (255 kWh/m²), west (231 kWh/m²), south (215 kWh/m²), southeast (214 kWh/m²) and east (183 kWh/m²), as shown in (Figure 6).
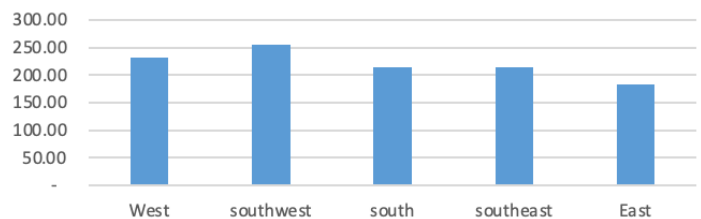
Figure 6: Total energy consumption of the base case in all façade orientations

As expected for hot environments, cooling represented the highest component from 185 kWh/m^2 to 136 kWh/m^2, which was almost 80% of the energy consumption. As shown in (Figure 7) energy consumption due to lighting is kept to the minimum since daylighting controls minimized the operation of artificial lighting.

Annual energy consumption has been presented in kWh/m^2 for alternative shading screen perforation percentages and orientations.
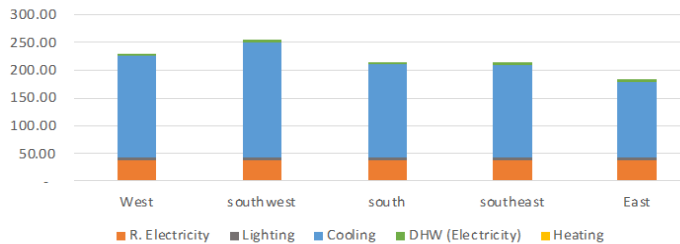
Figure 7: Breakdown of energy consumption in the base case

5.1. West orientation:

West orientation in the base case considered as 10% less energy consumption compared to southwest. Therefore, as shown in (Figure 8) the energy saving reached 48% with 10% perforated shading screen, which saved up to 111 kWh/m^2

	Base case	90%	80%	70%	60%	50%	40%	30%	20%	10%	5%
West	230.95	199.22	191.94	176.11	168.96	154.33	153.74	140.24	127.07	119.72	127.48
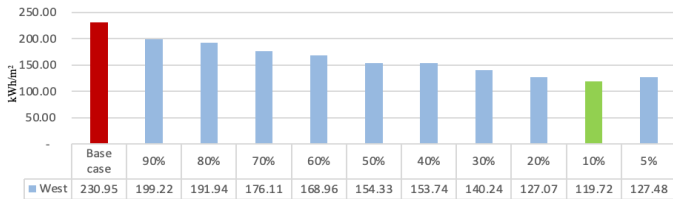
Figure 8: Total Annual Energy Consumption of the case study without and with shading screens (90% to 5% Perforations) for the west facade

The graph below in (Figure 9) shows the relation between the perforation percentage and the energy consumption in the southwest façade. The cooling is the main factor of saving energy in which it decrease around 45% and saving 86 kWh/m^2 from 90% perforated shading screen to 5% perforation. The inverse of this relation shown in lighting but with less impact on energy saving in which 90% perforation is 5.6 kWh/m^2 and the 5 % perforation is 19.8 kWh/m^2. The difference is around 14 kWh/m^2 whichis considered a little impact on energy saving compared to the cooling saving.
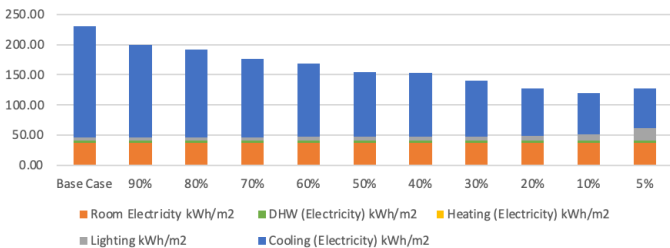
Figure 9: Energy consumption breakdown for the 5% to 90% perforated shading screens compared to the base case for the west facade

5.2. Southwest orientation:

As the southwest orientation was the highest energy consumption in the base case, in order to save as much energy as

possible, we used perforation percentage from 90% - 5%. As shown in (Figure 10) that the energy saving reach 136 kWh/m^2 saving in the 10% perforation shading screen, which is almost 53% saving of energy consumption.

	Base case	90%	80%	70%	60%	50%	40%	30%	20%	10%	5%
southwest	254.64	215.01	205.96	187.36	178.34	162.93	162.39	146.38	128.62	118.58	125.74
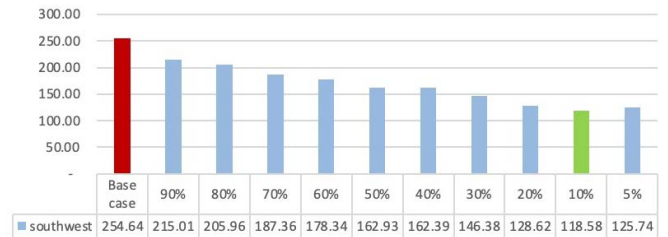
Figure 10: Total Annual Energy Consumption of the case study without and with shading screens (90% to 5% Perforations) for the Southwest façade

The graph below in (Figure 11) shows the relation between the perforation percentage and the energy consumption break down in the southwest façade. Heating and room electricity does not change, or almost fixed. The cooling is the main factor of saving energy in which it decrease around 40% of total cooling energy which save 100 kWh/m^2 from 90% perforated shading screen to 5% perforation. The inverse of this relation shown in lighting but with less impact on energy saving in which 90% perforation is 5.5 kWh/m^2 and the 5% perforation is 18 kWh/m^2. The difference is around 13 kWh/m^2 that considered as little impact on energy saving in compared to the cooling saving. The most important factor in the analysis is the cooling energy consumption that has a great impact on the saving energy.
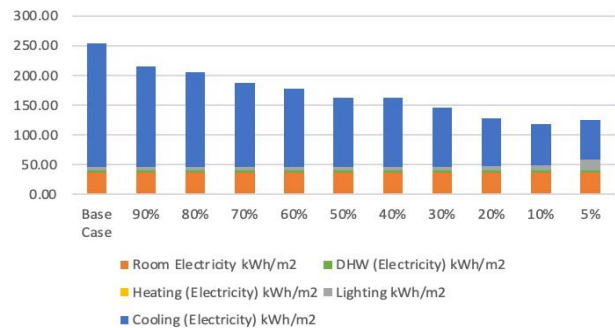
Figure 11: Energy consumption breakdown for the 5% to 90% perforated shading screens compared to the base case for the Southwest facade.

5.3. South orientation

South orientation in the base case considered as 19% less energy consumption compared to southwest. As shown in (Figure 12) the energy saving reached 47% with 10% perforated shading screen, which saved up to 100 kWh/m^2

	Base case	90%	80%	70%	60%	50%	40%	30%	20%	10%	5%
south	214.64	185.61	179.98	167.87	160.69	151.41	151.17	138.67	123.41	114.52	125.49
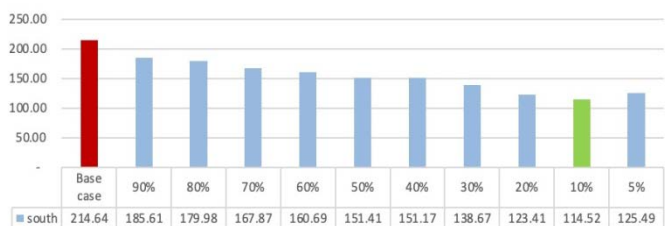
Figure 12: Total Annual Energy Consumption of the case study without and with shading screens (90% to 5% Perforations) for the South façade

The graph below in (Figure 13) shows the relation between the perforation percentage and the energy consumption in the south façade. The cooling is the main factor of saving energy in which it decrease around 47% saving 73 kWh/m² from 90% perforated shading screen to 5% perforation. The inverse of this relation shown in lighting but with less impact on energy saving in which 90% perforation is 5.5 kWh/m² and the 5% perforation is 18.8 kWh/m². The difference is around 13 kWh/m² that considered as little impact on energy saving in compared to the cooling saving.
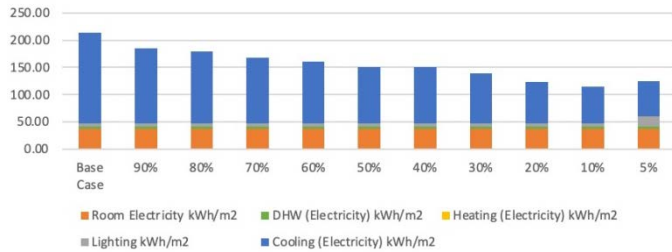
Figure 13: Energy consumption breakdown for the 5% to 90% perforated shading screens compared to the base case for the south facade

5.4. Southeast orientation

South East orientation in the base case considered as 19% less energy consumption compared to south West. Therefore, as shown in (Figure 14) the energy saving reached 47% with 10% perforated shading screen, which saved up to 101 kWh/ m².
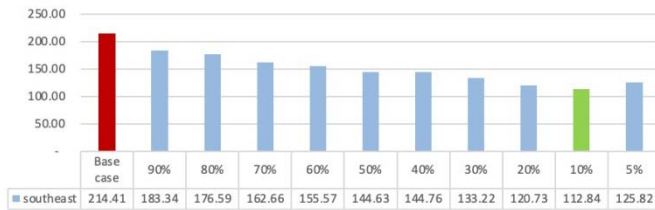
Figure 14: Total Annual Energy Consumption of the case study without and with shading screens (90% to 5% Perforations) for the South façade

The graph below in (Figure 15) shows the relation between perforation percentage and the energy consumption in the southeast façade. The cooling is the main factor of saving energy in which it decrease around 48% saving 72 kWh/m² from 90% perforated shading screen to 5% perforation. The inverse of this relation shown in lighting but with less impact on energy saving in which 90% perforation is 5.6 kWh/ m² and the 5% perforation is 20 kWh/m². The difference is around 15 kWh/m² that considered as little impact on energy saving in compared to the cooling saving.
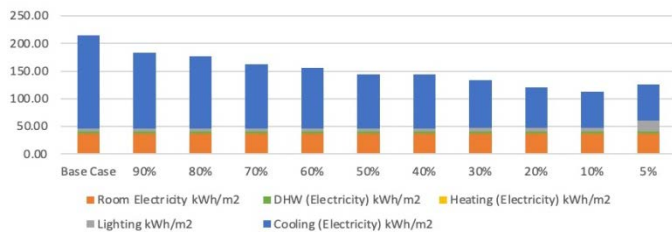
Figure 15: Energy consumption breakdown for the 5% to 90% perforated shading screens compared to the base case for the Southeast facade

5.5. East orientation

East orientation in the base case considered as 39% less energy consumption compared to southwest. Therefore, as shown in

(Figure 16) the energy saving reached 39% with 10% perforated shading screen, which saved up to 72 kWh/ m².
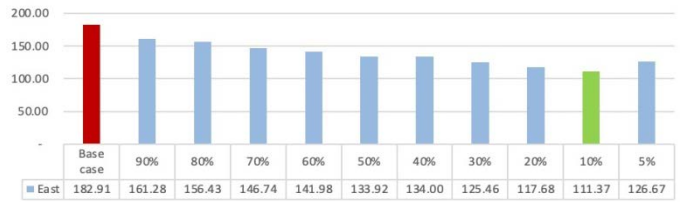
Figure 16: Total Annual Energy Consumption of the case study without and with shading screens (90% to 5% Perforations) for the East façade

The graph below in (Figure 17) shows the relation between the perforation percentage and the energy consumption in the east façade. The cooling is the main factor of saving energy in which it decreases around 56% saving 50 kWh/m2 from 90% perforated shading screen to 5% perforation.
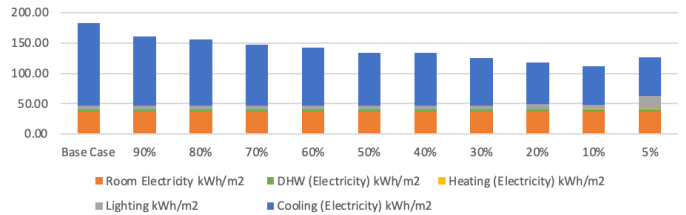
Figure 17: Energy consumption breakdown for the 5% to 90% perforated shading screens compared to the base case for the east facade

5.6. Overall energy savings

The total annual energy savings resulting from using a shading screen with 10% perforations compared to the base case. Shown in (Figure 18) the orientations with highest and least energy savings (53% and 39%) are the southwest and east respectively.
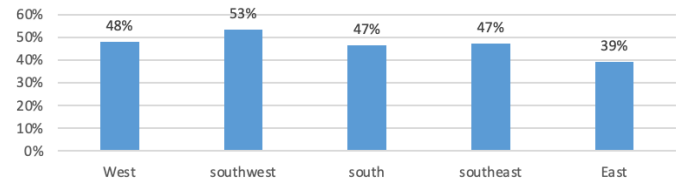
Figure 18: Percentage of saving energy for 10% perforation shading screen compared to base case

6. Life Cycle Cost of GRC Shading Screen with 10% perforations in southwest Façade.

6.1. Initial Cost

The Initial cost of GRC shading screen was based on quotations obtained from a local supplier submitted on July 2021. The initial cost of GRC is 450 EGP / m² and the total area need to cover the curtain wall of the office room is 12m2, so the total price for the room is 5,400 EGP (Egyptian Pounds), which is 338 EGP/ m².

6.2. Operation Cost

As discussed earlier, operation costs are mainly the electrical energy costs. Operation cost includes all energy costs required for cooling, lighting, heating and office equipment for the test cases, previously simulated. The cost of energy consumption was calculated according to the Egyptian electrical cost (tariff) by the

operation year for commercial buildings 2021which equals 1.6 EGP/kWh [16].

$$\text{Annual Operation Cost } (A_0) = \text{Annual Electric Consumption /m}^2 \text{ (E) x Tariff (T)} \quad (2)$$

In order to apply a present cost of 20 years of operation in today's value, the present value equation in Equation (3) need to be considered where the present value is defined as: "the time equivalent value of past, present or future cash flows as of the beginning of the base year" [14]. The main inputs of the present values are annual costs and the annual real discount rate (d) which defines annual escalation in energy prices, which is set to be 4.3% [17] resulting that present value of the GRC shading screen is 2,520 EGP/m². GRC material does not have a maintenance cost over the 20 years operation.

$$PV = A_0 \ x \ \frac{(1+d)^t - 1}{d \ x \ (1+d)^t} \quad (3)$$

Where:
PV = present value
A_0 = annual recurring costs (operation costs)
d = real discount/interest Rate
t = time (number of years)

6.3. Residual Value

Residual value is defined as the net cost of a building or building component after at the end of the LCC study period [14]. The residual value will be equivalent to the resale price of the metal studs as scrap. The current local price of metal scrap is equal to 5 EGP/kg based on local market survey in 2021. The Weight of the resale value of metal studs used in GRC fixation is 35 Kg/m² that represent 176 EGP/ m² (Table 1).

6.4. LCCA Results

LCC has been calculated by applying Equation (1), adding the initial cost per unit area, which is 338 EGP/m² and operation costs per unit area, which is 2,519 EGP/m², and then subtracting the residual value 175 EGP/m2 (Table 2). Since, the base case has no initial or residual value. The LCC of the base case is based on operation costs only, which is 5,411 EGP/m². The LCC of 10% perforation shading screen is 2,682 EGP/ m².

Table 1: Residual Value Breakdown for the GRC Shading Screen

volume of metal frame	cross sectional area	studs L*w	m³
	0.0015	16	0.024
Weight of single metal	volume of metal	metal density	kg
	0.024	2600	62.4
weight per room surface	Weight of single metal frame	No. of Metals	kg
	62.4	9	561.6
weight/ unit room area	weight of metal in Room	Room Area	kg/m²
	561.6	16	35.1
Price of kg	5		
Residual value	175		

Table 2: LCC of the Base case and 10% perforation of Southwest façade

	Base Case	10% perforation in southwest
Initial Cost of GRC shading screen	0	338
Operation Cost	5,411	2,519
Maintenance cost for 20yrs	0	0
Residual value	0	175
LCC	5,411	2,682

The main objective of calculating LCC was to calculate the savings and the cost payback of shading screen with reference to the base case. The results shows that the difference between LCC in the base case and the 10% perforation shading screen in the south west façade save 2,623 EGP/m2 during the study period, which actually save up to 52%.

6.5. Cost Payback period

The payback period is the time you need to recover the cost of your investment. As the total initial cost of GRC Shading Screen is 338 EGP/ m2 and the LCC is 2,682 EGP/ m2. The payback period will be in the second year of operation (Table 3)

Table 3: LCCA difference between the Base case and 10% perforation of Southwest façade

	1st Year	2nd Year
LCCA-Base Case	391	766
LCCA-10% Southwest	182	357
Savings	209	409

7. Conclusion

In order to understand the value of integrating perforated GRC shading screens for different orientations of office building façades in Cairo, A range of GRC shading screen perforations was assessed for office building façades with different orientations using Design Builder. The shading screen perforations follow an orthogonal grid as advised by [12].

The utilization of perforated shading screens for protection of curtain walls in the severe hot desert environments proved to provide considerable energy savings. The simulation of a typical office space in Cairo, Egypt illustrated this. The energy savings resulting from the use of these screens reached 53% for the southwestern orientations respectively in comparison a non-shaded curtain wall, which match the findings of [13].

The LCCA compared the southwest oriented 10% perforation GRC shading screen to the base case. LCC savings reached 52% and the return of investment of the shading screen is on the second year of operation. The research used the GRC as a local affordable material for the static shading screen that provides minimum initial cost and fast return of investment as advised by [15].

7.1. Research Recommendation

As the research shows, the most effective allocation of the shading screen is in the southwest as it is the location with the

highest energy consumption. Thus, by adding the shading screen it will have the highest energy saving. As for the perforation percentage, the narrow opening is more recommended for highest energy and cost saving percentages especially for south-west and west façade orientation.

7.2. Further Research

This paper reported on orthogonal grid screens with square fenestrations. More shading screen shapes with different dimensions could be examined with the effect of the lighting quality.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] H. Eskandari, M. Saedvandi, M. Mahdavinejad, "The Impact of Iwan as a Traditional Shading Device on the Building Energy Consumption," Buildings, **8**(1), 1-28, 2018, doi:10.3390/buildings8010003.

[2] B.L. Finlayson, T.A. McMahon, M.C. Peel, "Updated world map of the Köppen-Geiger climate classification," Hydrology and Earth System Sciences, **11**, 1633–1644, 2007, doi:10.5194/hessd-4-439-2007 .

[3] E.S. Lee, D.L. DiBartolomeo, S.E. Selkowitz, "Thermal and daylighting performance of an automated venetian blind and lighting system in a full-scale private office," Energy and Buildings, **29**(1), 47–63, 1998, doi:10.1016/S0378-7788(98)00035-8.

[4] N. Hamza, A. Gomaa, C. Underwood, "Daylighting and thermal analysis of an obstructed double skin façade in hot arid areas", Proceedings of Clima 2007 WellBeing Indoors, 7-10,2007.

[5] S. Fuller, "Life-Cycle Cost Analysis (LCCA)| WBDG—Whole Building Design Guide", National Institute of Standards and Technology (NIST), http://www.wbdg.org.html, (accessed June 2021), Washington, 1-13, 2016.

[6] Y.M. El Semary, H. Attalla, I. Gawad, "Modern Mashrabiyas with High-tech Daylight Responsive Systems," in 2017 CITAA Cities' Identity Through Architecture and Arts, 1–11, 2017, doi:10.21625/archive.v1i1.113.

[7] G. Pitts, S. Datta, "Parametric modelling of architectural surfaces," in 2009 CAADRIA: the 14th International Conference on Computer Aided Architectural Design Research in Asia, 635-644, 2009.

[8] S.M. Al-Masrani, K.M. Al-Obaidi, N.A. Zalin, M.I. Aida Isma, "Design optimisation of solar shading systems for tropical office buildings: Challenges and future trends," Solar Energy, **170**, 849–872, 2018, doi:10.1016/j.solener.2018.04.047.

[9] A.Watts, A Modern construction envelopes, Ambra Verlag, 2014. doi:10.1515/9783990436042.

[10] A. Batool, I.M.K. Elzeyadi, "From romance to performance: assessing the impacts of jali screens on energy savings and daylighting quality of office buildings in lahore, Pakistan," in 2014 PLEA Proceedings of the 30th International PLEA Conference in India, 1-10, 2014.

[11] I. Elzeyadi, A. Batool, "Learning from the Vernacular : The Impacts of Massive Perforated Screen Shades on Building Energy Savings and Thermal Comfort in Two Different Hot Climate Zones," in 2018 ASHRAE Winter Conference, 1-8, 2018.

[12] A. Sherif, A. El-Zafarany, R. Arafa, "External Perforated Window Solar Screens: The Effect of Screen Depth and Perforation Ratio on Energy Performance in Extreme Desert Environments," Energy and Buildings, **52**, 1–10, 2012, doi:10.1016/j.enbuild.2012.05.025.s.

[13] A. Sherif, A. Faggal, R. Arafa, "External Perforated Solar Screens for Thermal Control in Desert Environments : The Effect of Perforation Percentage on energy loads", in proceedings of Renewable energy, 1-5, 2010.

[14] S.K. Fuller, S.R. Petersen, LIFE-CYCLE COSTING MANUAL for the Federal Energy Management Program, NIST publication; U.S. department of commerce technology Administration National Institute of Standards and Technology, 1996.

[15] A. Essam Eldeen, O. Tolba and S. Ezz Eldin." Life-cycle cost assessment of kinetic aluminum louvers on office building façades in Egypt." Cairo University's Journal of Engineering and Applied Sciences, **62**(1), 1–22, 2015.

[16] EGYPTERA, "http://egyptera.org/ar/t3reefa.aspx", (accessed June 2021), Ministry of Electricity, Cairo, 2021.

[17] Central Bank of Egypt, "http://www.cbe.org.eg/English/Statistics/" (accessed June 2021), Cairo, 2021.

Cyberbullying Detection by Including Emotion Model using Stacking Ensemble Method

Natasia[*], Sani Muhamad Isa

Computer Science Department, BINUS Graduate Program – Master of Computer Science Bina Nusantara University, Jakarta, Indonesia 11480

A B S T R A C T

Cyberbullying is a serious problem and caused an immense impact to the victim. To prevent the cyberbullying, the solution is to develop an automatic detection system. In this research, we propose a combined model for cyberbullying detection and emotion detection by using stacking method. The experiment is to create a better model for cyberbullying detection using SVM (support-vector machine), KNN (K-Nearest Neighbors), and Naive Bayes method, then combine the best model with the emotion model. The result conducted that using SVM classifier give the best accuracy for both emotion and cyberbullying detection. Emotion detection yields an accuracy of 96.7%. Cyberbullying detection using SVM classifier yields an accuracy of 72.73%. Then, both model are combined using the stacking ensemble method and yield an average accuracy of 77.8%. It concluded that including the emotion model would improve the accuracy of detection.

1 Introduction

Technologies have grown rapidly and give a big impact on our social life. It makes communication easier and become a big role in life, especially in communication and relationship among people [1]. However, the more people use these platforms, the more crime and offense occurs. Bullying in social media become common and afford a great impact to many people, especially a teenager. In 2016, showed that 41% - 50% of the teenager had experienced cyberbullying [2].

Cyberbullying could make a great impact on victims, especially young adults. Research shows that the ability to regulate emotional responses are influenced by age [3]. However, the use of the internet for social networks are around 41.7% - 46.7% for young adult [4]. This would create a long-lasting impact on mental health. Some who experienced cyberbullying shows emotional distress and a likelihood of acting out. Some also experiencing mental health problems, such as drugs, abuse, and suicide [5].

Cyberbullying is a serious problem, which could harm the mental of social media users. The increase of social media users and how social media is easy to use made it more difficult to validate the posts and comments. Moreover, many users left to be unknown are showing a bad habit in commenting. To reduce cyberbullying and the bad manner in comments, the solution is to automatically detect and validate the content.

To prevent cyberbullying, the solution is to develop an automatic detection system to categorize which content is a bully content and reporting if the system detects the kind of bullying [6]. By using natural language processing and text mining, the cyberbullying feature are extracted and classified [1], [6].

The previous study has conducted the cyberbullying detection with a psychological feature by using big five personality traits and dark triad features. The used features are number of followers, following, popularity, user favorite count, status count, extraversion, agreeableness, neuroticsm, psychopathy, and sentiment. The experiment resulting 90.1 – 91.7% accuracy, where the baseline + psychopathy have the highest accuracy [7].

In this paper, we approach to determine how user got bullied or not by including emotion features. The emotion features are extracted into 4 classes and resulting a model of cyberbullying detection. Then, the emotion model and cyberbullying model will be used as a feature to create a new model for better detection.

The study is organized as follows. In section 2, we introduce some of related works. Then, we propose our methodology for the emotion detection, cyberbullying detection, and the ensemble method. In section 4, the experiment and evaluation method are detailed. In section 5, experimental results are detailed and analyzed. Finally, concluding study results are provided in section 6.

[*]Corresponding Author: Natasia, Bina Nusantara University, Jakarta, Indonesia 11480, & natasia001@binus.ac.id

2   Background and Related Work

Considering the influence of cyberbullying on victim's mental health, several studies have been conducted. With the improvement of technology, the automatically cyberbullying detection could be applied. The main goal of researching this is to maintain the utilization of social media and reduce bullying activity in social media.

Some previous works use a heterogeneous technique to perform the detection using text mining principal. Using text mining and natural language processing concept, the step to do are preprocessing, feature generation (Bag of Words), feature selection, classification, and analyzing results [1], [6].

There also a detection by using Big Five and Dark Triad features [7], [8]. The effectiveness of cyberbullying detection is compared to the baseline model among the Big Five and Triads models. In the experiment, higher accuracy is produced for three personalities in Big Five and one personality in Dark Triad, which are extraversion, agreeableness, neuroticism, and psychopathy [7].

2.1   Text Processing

Text processing or text mining is used to extract knowledge, information, or pattern and convert the unstructured information into structured data to solve the problem [9], [10]. Preprocessing in text mining could be done in several ways. The most commonly used for preprocessing are tokenization, stemming and lemmatization, stop word and n-gram [10], [11].

Tokenization is used to split or segmented text into tokens, such as words and punctuation by separating each text from whitespace and punctuation. Stemming and lemmatization are used to derivationally the word into the base form. The stemming works by cutting off the end of the word into a list of common prefixes or suffixes. Lemmatization work by analyzing the word and link the text with its basic form.

Stop word is used to filter out the word that does not use for further process. For high-frequency words, such as "the", "I", "we", etc, the word is filtered to improve the accuracy in data. N-gram is used to determine the connection with each word with a contiguous sequence of n-words. This is used to calculating the probabilities and improving the predictions of a word that is connected.

After the pre-processing, the next step is information extraction or feature extraction. Feature extraction is a process for raw data to be converted into a manageable resource. The feature extraction contains clustering, where the extracted data will be clustered based on their feature [9].

This process will convert the text into base-features and effectively reducing the amount of data to be processed while still accurately describing the original data. Feature extraction could be using several methods. In the previous work, it used POS-Tag, BoW, EboW [1], Word2Vec [11], or TF-IDF [12], [11].

TF-IDF (term frequency-inverse document frequency) is a measurement that determined the usefulness of word collection in a document [13]. With given D as a document, w as word, f as frequency, the importance of a word is defined in equation 1 [14].

$$wd = f(w, d) * log(|D|/f(w, d)) \qquad (1)$$

The next step is classification. Classification is a process to categorize the set of data into classes. In text mining, there are many algorithms to classified data, for example, using SVM, KNN, Decision Tree, Naïve Bayes, and many else [15], [16], and [11].

SVM uses kernel trick to map the data into a higher-dimensional space. The simplest SVM formula is linear SVM, where the hyperplane are placed on the space of the input data. The objective of using SVM algorithm is to find a boundary / hyperplane that classifies the data features (Figure 1) [17], [18].
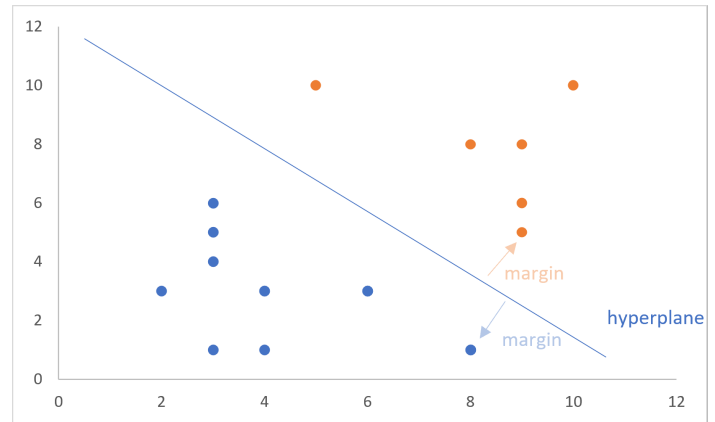
Figure 1: SVM Example

When SVM uses the hyperplane, KNN relies on the labeled data and builds a learning method to obtain the prediction of given unlabeled data. Before inputting the data into KNN classification, the data must be extracted into weight feature, by using TF-IDF, frequency, or a score. KNN will create a mapping for each class, with a border value to the other class.

2.2   Emotion Detection

Social media are frequently used to shared emotional feelings and opinions in form of a text message. Through the text message, the user show their emotion [19]. To detect the emotion automatically, the text message must be processed with several methods.

In [19] research, emotions are divided into four major classes, which are Happy-Active, Happy-Inactive, Unhappy-Active, Unhappy-Inactive [20]. The detection is developed using a supervised machine learning approach and including two tasks, an offline training task and an online classification task. The first task is called Emotex and the second task is called EmotexStream.

The dataset is obtained from Twitter tweet data. The collected data are selected when the comment would indicates an emotion. The preprocessing is to mitigate the common text which not indicate an emotion feature.

To convert the feature, the researcher use unigram, emoticon, and negation features. Unigram features are used to capture emotion in text. Emoticon features are used to portrait emotion by using an emoticon. For example, using ":)" to illustrated "happy" emotion.

Negation features are converted by using a single-phase by selecting the list of phases in LIWC dictionary. For example, "not happy" as the negation feature will be converted into "sad". Lastly,

by using Naïve Bayes as a probabilistic classifier, SVM as a decision boundary, and KNN classifier, the emotions are classified.

2.3 Cyberbullying Detection

Many researchers have researched on cyberbullying detection. For instance, using bullying features [1], create a distinction between cyberbullying and cyber aggression [21], investigate cyberbullying behavior [21], using a socio-linguistic model [22], using preprocessing of natural language processing [23] and even using pronunciation based CNN (Convolutional Neural Network) [24]. Summarizing the previous work, we attain that the best solution is using SVM as a classifier for the learning method [1] and [21].

A various dataset is taken from social media comment and post, such as Twitter and Instagram [1], [23], [24]. The TF-IDF is applied for feature extraction and used as input to classifiers. Bullying features will be pre-defined based on linguistic resources, which are named as insulting seeds [1].

After that, feed the learned feature into a linear SVM pattern classifier. The features are classified into bullying features and non-bullying features. Bullying features are separated into a threat, insult, defense, sexual talk, curse / exclusion, defamation, and encouragement of harasser [23].

For testing, the feature extraction is using text mining preprocessing and TF-IDF, which contains unigram to label every text which contains the bullying features. Then, we use SVM classifier to classify and label the data into non-bullying trace and bullying trace [1].

2.4 Ensemble Method

Several researchers have proven that the ensemble method could improve the performance in classification and regression problems [25], [26], [27], and [28]. This method is introduced by combining different learning models and create a new prediction. The most used ensemble methods are bagging, boosting, and stacking.

In the stacking method, the idea is to learn from several weak learners and combine them by training a meta-model. The output from base models is used as input for the meta-model. Unlike bagging, the models in the stacking method are typically different and fit on the same dataset.

3 Methodology

The proposed method in this research are defined in Figure 2. The stages are creating emotion detection model, cyberbullying detection model and ensemble the model.

The data are collected from Formspring and Twitter posts, which have been labeled as bully content or not. Afterward, implementation of the model will be done in 2 steps: emotion detection and bullying detection. The emotion model is divided into 4 classes: Happy-Active, Happy-Inactive, Unhappy-Active, and Unhappy-Inactive. When the victims show a sign of unhappiness, the probability of certain content is highly bully content.

In this research, we proposed the cyberbullying detection by using emotion detection through text. This method is performed by combining the model of cyberbullying detection and emotion detection. By ensemble the result of emotion detection and cyberbullying detection, we assume to have a better result. The fusing method is using stacking ensemble method.
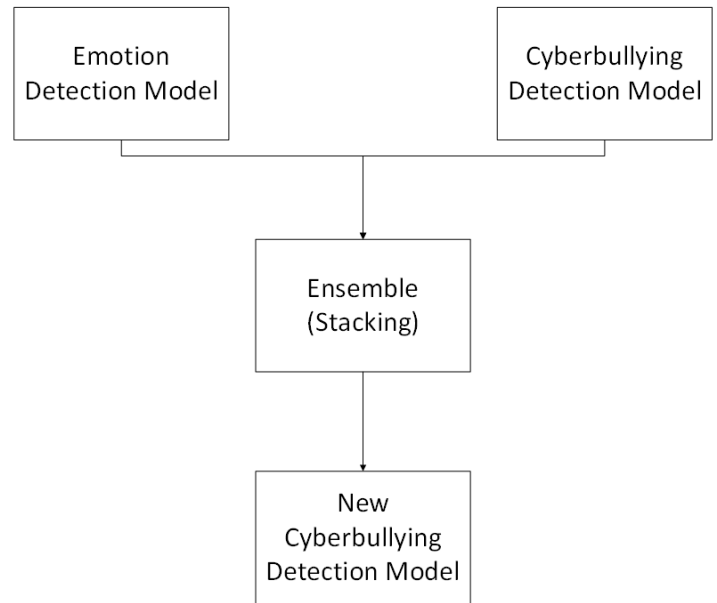
Figure 2: Proposed Model

3.1 Emotion Detection

We perform three steps, which are preprocessing, feature selection, and classifier [29]. By using WordNet synsets, the emotions are expanded. Then from Twitter's tweet, the data are collected with the following criteria:

1. Contain one or more hashtags that defined emotion hashtag.

2. Ignore the retweet (which begins with "RT" keyword).

After collecting the data, the labeled tweets are processed by using the following rules:

1. User ID and URL are separated.

2. Text normalization.

3. Remove conflicting hashtag. For tweet which contains different emotion class will be removed.

4. The hashtag at the end of tweets is considered an emotion label.

5. Change negation word into the antonym word.

For capturing emotion, we performed three parts to convert the features, which are unigram, emoticon, and negation feature. In unigram feature selection, each word is classified into emotion categories in LIWC (Linguistic Inquiry and Word Count) lexical, which contains a dictionary of thousand words with specified emotion indicative.

The emoticon features are used to express the emotion, such as using :) as happy, :(as sad, :/ as angry, :0 as afraid/surprised ,

etc., obtained from [29] and combined with nltk sentiment documentation. The negation list are obtained from the nltk sentiment documentation.

Table 1: Emoticon List

Emotion	Emoticon
Happy	:) ;) =) :] :P :-P ;P :D ;D :> :3 :-) ;-) :ˆ) :o) :˜) :ˆ) ;o) :') :-D :-> :-) :o) :] :3 :c) :> =] 8) :} :ˆ) :-D 8-D 8D x-D xD X-D XD =-D =D =-3 =3 :-)) :'-) :* :ˆ* >:P :-P :P X-P x-p xp XP :-p :p =p :-b :b >:) >;) >:-) <3
Sad	:(=(:-(:ˆ(:o(:-(:'(:-< :L :-/ >:/ :S >:[:@ :-(:[=L :< :-[:-< =\\ =/ >:(:'-(:'(:\\ :-c :c :{ >:\\ ;(: : −‖
Angry	>:S >:{ >: x-@ :@ :-@ :-/ :-\ :/
Afraid	:-o :-O o_O O_o :$
Sleepy	-_- ˜_˜

Table 2: Negation List

Negation List	never, no, nothing, nowhere, noone, none, not, havent, hasnt, hadnt, cant, couldnt, shouldnt, wont, wouldnt, dont, doesnt, didnt, isnt, arent, aint

There are possibilities that tweets have a word that conflicting with others' word. One word could be classified as happy emotion, while others are sad. For example, "Finally the last exam! Must study hard to get a better score #depressed #excited". This example shows that the post has exited feeling of upcoming exam and at the same time are depressed. To reduce these ambiguities, the negation features are separated and converted into the antonym word.

We use TF-IDF for the feature extraction [14]. Then, SVM and KNN Classifier are chosen as its level of accuracy are highest compared to Naïve Bayes and Decision Tree [19].

Table 3: Preprocessing for Emotion Detection

Feature	Process
'@' feature	A symbol that represented USERID
url links	Replaced into URL
Tweets contain more than one '#' and represented as two different classes	Remove the content to reduce ambiguities
Tweets contain more than one emoticon and represented as two different classes	Remove the content to reduce ambiguities
Tweets contain conflict between emoticon and '#'	Remove the content to reduce ambiguities
'#' in the end of content	Stripped the tags

Figure 3: Emotion Detection

3.2 Cyberbullying Detection

The data for cyberbullying are obtained from Kaggle.com, which provide Formspring data for cyberbullying detection. The data contains a question-and-answer field. Every content is reviewed by 3 annotators, with "yes" if the content is detected as a bully and "no" if not. Also, the annotator indicated the content severity with a 0-10 level.

The preprocessing is defined in 7 steps, which are data cleansing and data balancing, tokenization, transform case, stop word removal,

filter token, stemming and generate n-gram [12].

We use 2 and 4 classes for data cleansing and data balancing, which shown better accuracy [12]. The next step is tokenization, where the words are sliced into separated for each word (separated by a special character). Next, transform case will transform all the words into lower case.

Stop word removal will delete the unnecessary words based on nltk library. Filter token is used to remove the word that contains below 3 or more than 25 characters. Using Porter Stemmer, the words are converted into basic words. Lastly, we use n-gram to form a set of the word.

To produce an input that can be used as an input classifier, feature extraction is required. The preprocessing output is transformed into a vector space model and represented with a vector, with the calculation of weight. We use TD-IDF to extract the feature. The classifier of cyberbullying is using SVM method with the kernel poly. The classification is defined in 2 classes: non-bully and bully and 4 classes: class no, class low, class middle, and class high.

3.3 Stacking Ensemble Method

The idea of stacking is to learn several weak learners and combines them by training a meta-model to create a new prediction based on the new model (the combined model). The stacking architecture involves two or more base-models (level-1) and meta-model. Meta-models are trained using the predictions made by the base model (the output of the basic model is used as input in the model meta). The training dataset are prepared using k-fold cross-validation of the base-models.

The stacking method has three steps. The first step is creating the base model classifiers. Then, construct a new dataset for the meta-model. Lastly, construct the meta-model classifier using the output from the base model prediction.

In this research, we use the best model from cyberbullying and emotion detection. After both models are constructed, the models are combined using the stacking ensemble method.

This stacking method is done in three steps: learning basic classifiers, loading a new dataset and creating a meta-model. The first step is to create a base classifier, including emotion detection using SVM and cyberbullying prediction using SVM. In the emotion detection, the unhappy-active will be labeled as cyberbullying and the others will be labeled as non-bullies. Unhappy-active emotions are classified as bully content since these emotions are constituted by displeasure and aggressive form.

The next step is to load a new dataset for the meta-model. The data will be pre-processed like in cyberbullying detection. Lastly, the meta-model is constructed from the base classifier and resulting in a new prediction.

Figure 4: Cyberbullying Detection
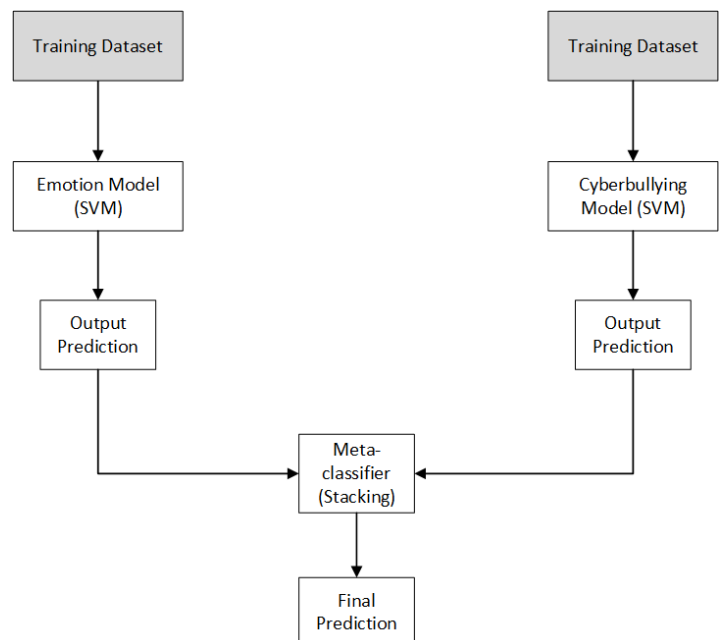
Figure 5: Stacking Ensemble Method for Cyberbullying Detection

---

Algorithm 1: Stacking Ensemble Method

Result: Meta-model Prediction
Step 1: Learn base classifier ;
Initialize models ;
Build Emotion Classifier (SVM) ;
Build Cyberbullying Classifier (SVM) ;
Step 2: Load dataset ;
Load and pre-processing dataset ;
Step 3: Build meta-model ;
Add all classifier as input for stacking classifier ;
Return new prediction ;

---

4   Experiment and Evaluation Method

The experiment for cyberbullying and emotion detection is using SVM, KNN, and Naïve Bayes classifier. The dataset will be divided into 3 parts, training data, validation data, and testing data. 60% will be used for training data, 20% for validation, and 20% for testing. The validation is using K-Fold Cross-validation. This validation is used to create a less biased model. The K-Fold Cross Validation will randomly divide a dataset into k disjoint folds, then each fold will be used to test the model [30].

The dataset for emotion is obtained from Twitter data and the cyberbullying dataset is obtained from Formspring data. The Formspring data are labeled with severity level for each comment. The amount of data after cleansing are 10,890, with 10,504 non-bullying data, 146 labeled as low, 151 labeled as middle, and 89 labeled as high severity.

Figure 6: Total Data for Each Severity Level

To create an unbiased model, the processed data are separated with the following details:

1. 2 Class

   - Class No: 330 data with severity 0
   - Class Yes: 330 data with severity 1-10

2. 4 Class

   - Class No: 80 data with severity 0

- Class Low: 80 data with severity 1-3
- Class Middle: 80 data with severity 4-7
- Class High: 80 data with severity 8-10

For the emotion detection, the data is obtained from Twitter tweets with 316,341 data after cleansing. The data are separated into 4 labels: happy-active 131,646 data, happy-inactive 34,036 data, unhappy-active 124,278, and unhappy-inactive 26,381 data. For creating a balanced model, we decide to use only 25,000 data for each class.

5   Result and Discussion

The result of the classification for cyberbullying is detailed in Table 4 and Table 5. The classification is using SVM linear, KNN with 3 neighbors, KNN with 5 neighbors, and Naïve Bayes. For the cyberbullying with four classes, the classification is using SVM linear. In the result, it could be seen that using two classes and SVM linear classification have the highest accuracy (72.73%) with an F1 score is about 72.7% and a recall score is 73.7%.

Table 4: Result for 2 Class Cyberbullying Detection

Classifier	Accuracy
SVM linear	**72.73%**
KNN 3 neigh	54.96%
KNN 5 neigh	50.94%
Naïve Bayes	66.66%

Table 5: Result for 4 Class Cyberbullying Detection

Classifier	Accuracy
SVM linear	52.5%

We use a confusion matrix to evaluate the accuracy of the model. Accuracy rates are calculated by using equation 2, where true positive (tp) and true negative (tn) is counted as true prediction, false positive (fp) and false negative (fn) is counted as a false prediction.

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{2}$$

For the emotion detection, the result is detailed in Table 6. The experiment is using two methods of classification, which are SVM linear and KNN (with 3 and 5 neighbors). The result shown that using SVM linear have the most accuracy (96.6 %) with an F1 score is about 94.81% and a recall score is 94.89%.

Table 6: Result for 4 Class Emotion Detection

Classifier	Accuracy
SVM linear	**96.6%**
KNN 3 neigh	62.38%
KNN 5 neigh	71.46%

Based on the results obtained, it appears that the best classification for cyberbullying and emotion detection is using SVM

classification. From several studies, it is also shown that SVM is better when used for text classification [1] and [12] . Besides, there are not many training data available, so when using the KNN, the results obtained are not optimal.

After constructing both models from emotion and cyberbullying, we create an ensemble model by using the stacking method. The result and comparison for the ensemble model are detailed in Table 7.

Table 7: Result and Comparison for Ensemble Model

Model	Accuracy
Emotion	96.6%
Cyberbullying	72.73%
Ensemble (Emotion + Cyberbullying)	**77.8%**

The results showed that by using the ensemble method for cyberbully and emotion models, the accuracy of cyberbullying detection could increase by about 5%. Detection on cyberbullying increased when combined with emotion detection using a method of stacking. This method is done by making each base model and making predictions, then the prediction results will be used as a feature for the meta-model. From this meta-model, new predictions will be generated.

By using this ensemble method, prediction errors from cyberbullying can be covered by predictions from emotion detection, thereby increasing the accuracy of the predictions. Accuracy can be increased due to the features of emotion. For words that have unhappy-active elements, such as angry, annoyed, frustrated will be the feature that determines whether the tweet is a bully content. This is because in most cases the bully's words constituted by displeasure and shown aggressive.

Also, the result shows that emotion detection has the highest accuracy when compared to cyberbullying detection model and an ensemble model. However, cyberbullying detection may not simply use the emotion model, as there are tweets that have no emotion feature but could be labeled as a bully tweet. Hence, the best solution is to use the stacking ensemble model from emotion and cyberbullying model to predict the content.

6 Conclusion

By including emotion feature in cyberbullying detection and using ensemble method, we ought to have a better accuracy. For cyberbullying, the most optimal classification is by using 2 class (bully and non-bully) and use SVM classifier with linear kernel, with an average accuracy is 72,73%. For emotion detection, the most optimal classification is by using 4 class (happy-active, happy-inactive, unhappy-active, unhappy-inactive) and use SVM classifier with linear kernel, with an average accuracy is 96,6%. Then, by ensemble both model using stacking method, the accuracy for cyberbullying improve to 77,8%.

This research has proven that including emotion in cyberbullying detection could improve the accuracy of the detection. This is because emotions that are indicated as unhappy-active generally show words that are less wearing and emit emotions such as anger, displeasure, distress, and other negative emotions. By including

these emotional features, accuracy, and precision of detection of cyberbullying increases.

For future works, the challenge is to try to make a model of cyberbullying detection with more data or in a different language. It is also challenging to consider sarcasm in tweet content for the possibility of bullying content.

References

[1] R. Zhao, A. Zhou, K. Mao, "Automatic detection of cyberbullying on social networks based on bullying features," ACM International Conference Proceeding Series, **04-07-Jan**(January), 2016, doi:10.1145/2833312.2849567.

[2] A. Sinring, F. Aryani, N. H. AR, "Development of Guidance and Counseling Videos to Reduce Cyber Bullying among Vocational Students," **227**(Icamr 2018), 324–328, 2019, doi:10.2991/icamr-18.2019.80.

[3] V. Orgeta, "Specificity of age differences in emotion regulation," Aging and Mental Health, **13**(6), 818–826, 2009, doi:10.1080/13607860902989661.

[4] A. Weinstein, D. Dorani, R. Elhadif, Y. Bukovza, A. Yarmulnik, P. Dannon, "Internet addiction is associated with social anxiety in young adults," Annals of Clinical Psychiatry, **27**(1), 4–9, 2015.

[5] T. Safaria, "Prevalence and impact of cyberbullying in a sample of indonesian junior high school students," Turkish Online Journal of Educational Technology, **15**(1), 82–91, 2016.

[6] H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, S. Mishra, "Analyzing labeled cyberbullying incidents on the instagram social network," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **9471**, 49–66, 2015, doi: 10.1007/978-3-319-27433-1_4.

[7] V. Balakrishnan, S. Khan, T. Fernandez, H. R. Arabnia, "Cyberbullying detection on twitter using Big Five and Dark Triad features," Personality and Individual Differences, **141**(September 2018), 252–257, 2019, doi: 10.1016/j.paid.2019.01.024.

[8] V. Balakrishnan, S. Kan, H. R. Arabnia, "Improving Cyberbullying Detection using Twitter Users' Psychological Features and Machine Learning," 2019.

[9] S. A. Salloum, M. Al-emran, A. A. Monem, "Using Text Mining Techniques for Extracting Information from Research Using Text Mining Techniques for Extracting Information from Research Articles," (November), 2018, doi: 10.1007/978-3-319-67056-0.

[10] S. Vijayarani, R. Janani, "TEXT MINING: OPEN SOURCE TOKENIZATION TOOLS – AN ANALYSIS," Advanced Computational Intelligence: An International Journal (ACII), **3**(1), 37–47, 2016, doi:10.5121/acii.2016.3104.

[11] A. Muneer, S. M. Fati, "A comparative analysis of machine learning techniques for cyberbullying detection on twitter," Future Internet, **12**(11), 1–21, 2020, doi:10.3390/fi12110187.

[12] Noviantho, S. M. Isa, L. Ashianti, "Cyberbullying classification using text mining," Proceedings - 2017 1st International Conference on Informatics and Computational Sciences, ICICoS 2017, **2018-January**, 241–245, 2018, doi: 10.1109/ICICOS.2017.8276369.

[13] A. Rajaraman, J. D. Ullman, "Data Mining," Mining of Massive Datasets, 1–17, 2011, doi:doi:10.1017/CBO9781139058452.002.

[14] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," Proceedings of the first instructional conference on machine learning, **242**, 133–142, 2003.

[15] M. Ikonomakis, S. Kotsiantis, V. Tampakas, "Text classification using machine learning techniques," WSEAS Transactions on Computers, **4**(8), 966–974, 2005, doi:10.11499/sicejl1962.38.456.

[16] H. Rosa, N. Pereira, R. Ribeiro, P. C. Ferreira, J. P. Carvalho, S. Oliveira, L. Coheur, P. Paulino, A. M. Veiga Simão, I. Trancoso, "Automatic cyberbullying detection: A systematic review," Computers in Human Behavior, **93**(October 2018), 333–345, 2019, doi:10.1016/j.chb.2018.12.021.

[17] T. Evgeniou, M. Pontil, "Workshop on Support Vector Machines : Theory and Applications," Machine Learning and Its Applications: Advanced Lectures, (January 2001), 249–257, 2001, doi:10.1007/3-540-44673-7.

[18] M. Awad, R. Khanna, "Efficient learning machines: Theories, concepts, and applications for engineers and system designers," Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, (July 2018), 1–248, 2015, doi:10.1007/978-1-4302-5990-9.

[19] M. Hasan, E. Rundensteiner, E. Agu, "Automatic emotion detection in text streams by analyzing Twitter data," International Journal of Data Science and Analytics, 2018, doi:10.1007/s41060-018-0096-z.

[20] J. A. Russell, "A Circumplex Model of Affect," Journal of Personality and Social Psychology, **39**(6), 1161–1178, 1980, doi:http://dx.doi.org/10.1037/h0077714.

[21] H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, S. Mishra, "Detection of Cyberbullying Incidents on the Instagram Social Network," 2014.

[22] S. Tomkins, L. Getoor, Y. Chen, Y. Zhang, "A socio-linguistic model for cyberbullying detection," Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018, 53–60, 2018, doi:10.1109/ASONAM.2018.8508294.

[23] C. Van Hee, G. Jacobs, C. Emmery, B. DeSmet, E. Lefever, B. Verhoeven, G. De Pauw, W. Daelemans, V. Hoste, "Automatic detection of cyberbullying in social media text," PLoS ONE, **13**(10), 1–22, 2018, doi:10.1371/journal.pone.0203794.

[24] X. Zhang, J. Tong, N. Vishwamitra, E. Whittaker, J. P. Mazer, R. Kowalski, H. Hu, F. Luo, J. Macbeth, E. Dillon, "Cyberbullying Detection with a Pronunciation Based Convolutional Neural Network," 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), 740–745, 2017, doi:10.1109/icmla.2016.0132.

[25] T. G. Dietterich, "Ensemble Method in Machine Learning," Springer, 1–15, 2000.

[26] A. Onan, S. Korukoğlu, H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification," Expert Systems with Applications, **57**, 232–247, 2016, doi:10.1016/j.eswa.2016.03.045.

[27] I. Perikos, I. Hatzilygeroudis, "Recognizing emotions in text using ensemble of classifiers," Engineering Applications of Artificial Intelligence, **51**, 191–201, 2016, doi:10.1016/j.engappai.2016.01.012.

[28] Yan-Shi Dong, Ke-Song Han, "A comparison of several ensemble methods for text categorization," 419–422, 2004, doi:10.1109/scc.2004.1358033.

[29] M. Hasan, E. Rundensteiner, E. Agu, "EMOTEX: Detecting Emotions in Twitter Messages," SocialCom Conference, 27–31, 2014.

[30] T. T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," Pattern Recognition, **48**(9), 2839–2846, 2015, doi:10.1016/j.patcog.2015.03.009.

ASTES

Leveraging Energy Efficiency Investments: An Innovative Web-based Benchmarking Tool

Filippos Dimitrios Mexis*, Aikaterini Papapostolou, Charikleia Karakosta, Elissaios Sarmas, Diamantis Koutsandreas, Haris Doukas

School of Electrical and Computer Engineering, Decision Support Systems Laboratory, Energy Policy Unit National Technical University of Athens, Athens, 104 31, Greece

A R T I C L E   I N F O

A B S T R A C T

Energy Efficiency (EE) plays a key role in decreasing energy consumption at a European level, while it is considered as one of the most cost-efficient means to achieve carbon reduction and reinforce energy sufficiency and security. EE financing is imperative to implement measures that will lead to achieving the desired carbon neutrality and, thus, avert climate change. The majority of EE investments ideas are abandoned during the first stages of investment generation as there is not enough interest by the involved actors to support the maturing of the idea. The present paper aims to boost EE investments by developing a web-based Tool that evaluates project ideas, connecting them with real financing proposals. All the above are being realised through standardised procedures, establishing a concrete typology of five (5) EE sectors, a well-structured risk assessment methodology of five (5) risk categories and (9) risk factors, and a benchmarking procedure that takes into account four (4) broadly used economic criteria and eleven (11) verified sustainability indicators. All the parameters are calculated using the candidate project data and EU official statistics, formulated into four (4) main criteria that are fed into a MultiCriteria Decision Analysis that performs the project's benchmarking. The presented methodology is being practically tested through the development of three (3) innovative Tools (Assess, Agree, Assign) and a stakeholder consultation process with around 200 participants. The Tools filter and benchmark candidate project ideas, based on the standardised benchmarking and the EU Taxonomy sustainability principles, while connecting the most promising project ideas with state-of-the-art financing methods, such as the Green Loans, the Green Bonds and the Energy Efficiency Auctions. By this token, the developed Tools provenly provide added value to the respective stakeholders, offering standardisation in EE project benchmarking and financing, building trust between investors and projects developers.

1. Introduction

This paper is an extension of work originally presented in 2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA 2020) [1]. The conference was realised virtually in 15-17 July 2021, originally organized in Piraeus, by the University of Piraeus and the University of Thessaly. The conference aims to create a forum for researchers and professionals in all areas of Information, Intelligence, Systems and Applications, while its proceeding are published by the Institute of Electrical and Electronics Engineers (IEEE) and uploaded on the its online platform, IEEEXPLORE.

The attainment of the energy and environmental targets set by the European Union (EU) in view of moving to a carbon-free economy by 2050 requires that more intense policies and measures be put into effect, ensuring that the target trajectory is on track [2]. Out of the available policy measures, the improvement of energy efficiency (EE) can play a key role in this transition, leading to the decrease of energy demand and emissions, as well as bringing about economic and social benefits [3] [4]. For this reason, EE has been recognised as the "first fuel" [5], indicating its essential role and the large-scale effects that it could result in.

Mainstreaming EE is the most feasible solution to maintain increased levels of economic and social activities while being a more sustainable and less energy demanding society [6]. The increase of EE implementation goes together with the upscale of EE investments since a prerequisite for realising EE projects is that the capital needed is secured at their development phase. However,

according to the current state of play, investments in EE are well below the required level to meet the established long-term targets [7].

Interestingly enough, despite the existence of many EE investments with great prospects at the development phase, only a few of them achieve to get financed; something that has been called as "efficiency paradox" or "energy efficiency gap" [8]. This phenomenon can be ascribed to the transparency gap that EE investments involve [9]. First, there is a shortage of evidence about successfully implemented EE projects, such as about their financial performance and the stakeholders involved within their context. The main factors that halt the financing of EE are the lack of evidence on the performance of such projects, standardised procedures and standards for energy efficiency investments, and the underestimation of its multiple benefits [10]. Moreover, postulating the financial performance of EE investments is a stochastic task at its nature since it is related to several uncertainties (e.g., energy prices), while their evaluation presents high complexity [11].

To the best of our knowledge, a standardised procedure for this scope does not exist up to date [12]. As a result, project developers do not have the expertise or resources to make a convincing financing case for investors, and, on the other hand, investors face difficulties in identifying the EE projects that merit attention [12]. Likewise, financial institutions, given the high transaction costs for evaluating EE investments with respect to their entailed technical aspects, focus only on the creditworthiness of the borrower for deciding whether they provide a loan or not [13]. Therefore, they may exclude from their portfolios many EE projects of great potential. In order to bridge the transparency gap that exists at the initial phases of EE investments, it is of paramount importance that innovative decision support schemes and standardisation Tools will become available towards providing the required information to the involved actors and key players in EE financing [14], also enabling their connection with each other.

The energy efficiency gap is a great field of interest for energy policy scientists, policymakers and research initiatives. Several efforts have been made to mitigate the energy efficiency gap and stimulate the energy efficiency project implementation in Europe by making use of a variety of different financing and funding methods. In [15], the author have developed a benchmarking framework to encourage EE Investments in South Europe under the financing method of Energy Performance Contracting, while the Energy Efficiency Financing Institutions Group (EEFIG) has evolved the EEFIG Underwriting Toolkit to assist financial institutions in scaling up the deployment of capital into energy efficiency [11]. A different approach is followed in [16] by introducing public-private partnerships to finance EE projects.

In addition, the European Union has funded projects and initiatives so as to research and develop benchmarking and standardisation methods for EE. Some examples are the LAUNCH project [17], which aims to accelerate deal closure and pipeline growth for Sustainable Energy Assets through standardised material and the RenonBill project [18] that provides tools to address the residential sector's energy renovation financing demand and to assess bundle investments. Also, the EEnvest project [19] aims to secure investors´ trust in EE actions for

existing buildings by developing a combined technical-financial risk evaluation framework focused on the renovation of commercial buildings and the E2DRIVER [20] project is developing a training platform that will boost the collective intelligence of the automotive industry on EE. In the scope of certification, X-tendo [21] and its toolbox introduce features of energy performance certificates to provide public authorities with improved compliance, reliability, usability and convergence of next-generation energy performance assessment and certification.

In this context, this paper proposes a methodological framework that supports, on the one hand, capital providers, i.e., financiers and investors, in finding an attractive project idea that merits attention, giving them access to a pool of project ideas whilst providing a detailed assessment of them. On the other hand, it provides a better understanding of the EE projects' framework to those looking for capital (e.g., project developers), enabling them to acquire the evidence towards performing complete proposals to investors [22].

The presented approach is implemented across three steps, each of which is materialised via a respective standardised Tool. At the first step ("*Assess Tool*"), the initial filtering and identification of attractive project ideas take place, while also the entailed uncertainty is postulated through the calculation of the total risk of failure of the EE investments under examination. Following, the benchmarking of the EE projects that qualified from the first step is realised (*"Agree Tool"*), taking into consideration multiple criteria of diverge nature (e.g., economic, energy), thus assessing them from a holistic point of view. In this effort, a multi-criteria approach is employed, resulting in the categorisation of investments. Finally, at the last step (*"Assign Tool"*), investors are connected to EE project ideas with a detailed description of their characteristics, while the most suitable financing instruments are proposed (e.g., Green Bonds), according to their needs. The pool of available investments is composed of the ones that have passed the first two steps successfully.

Apart from this introductory section, the rest of the paper is organised into five sections. Section 2 analyses the methodological framework of the proposed methodology and Tools' development, while Section 3 describes the materialisation of this methodology along with the methods and data that have been used. Section 4 presents the standardised Tools developed, and, finally, Section 5 concludes the paper and suggests the key points for future research.

2. Methodological Framework

The methodology has been structured into five steps, as depicted in the figure below:

Step 1: Identification of the eligible projects.

The proposed methodology aims to benchmark project ideas, which can be considered potential investments at an early elaboration stage. The project ideas should consist of EE measures and interventions that will provenly lead to a decrease in the energy demand of the site implemented. The eligible sectors of activity and type of EE measures should be adequately defined.

Step 2: Review/ Selection of Problem Parameters:

When assessing an EE project, the parameters that should be taken into account are diverse and multivariant, similar to the

nature of EE measures. The indicators and parameters had to be carefully reviewed and selected so as to include and quantify all aspects of EE measures properly, but also reflect in a standardised way the expected financial performance of these potential investments at the early stage of the project's conceptualisation. State-of-the-art typologies methods have been used in order to produce an integrated and standardised benchmarking.
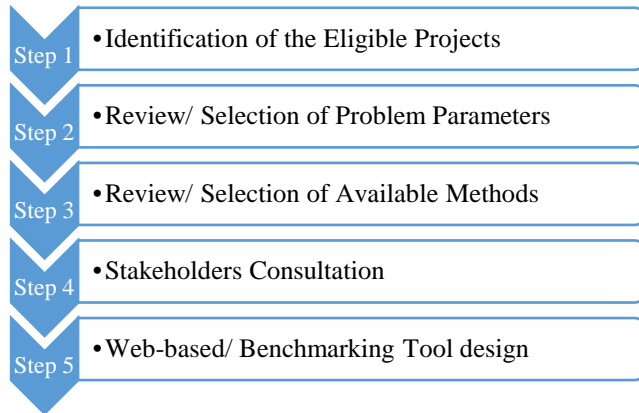
Figure 1: Main methodological steps

Step 3: Review/ Selection of Methods

In this step, a thorough review has taken place in order to run across the most suitable method that can incorporate and weight the selected criteria and benchmark candidate EE projects into predefined classes. Considering all the diverse and distinct factors that EE projects consist of while maintaining versatility, the review was primarily focused on Multi-Criteria Decision Analysis methods (MCDA). Methods such as the PROMETHEE UTADIS and the descendants of the ELECTRE-Tri method, such as the ELECTRE-Tri-B, ELECTRE-Tri-C, ELECTRE-Tri-nC, and ELECTRE-Tri-nB, have been taken into consideration [23].

Step 4: Stakeholders Consultation

The proposed methodology and the Tools developed are oriented towards EE companies and investors interested in investing in these kinds of projects. In order to properly assess and cover their needs and develop Tools that will provide added value to the EE sector, rigorous stakeholder consultation has been realised. The consultation activities included bilateral contacts, discussions, structured interviews, webinars, and workshops.

The consultation process has been organised into two distinct phases. The first phase took place before developing the respective methodology and Tools and is focused on identifying the design parameters. The second part was released after the initial development of the Tools in order to receive feedback and fine-tune the benchmarking process and methods. The stakeholder consultation steps are summarised in the following diagram.

The second part of the consultation included demonstrations and testings, which have been conducted in bilateral meetings with EE companies, project developers, and financers interested in sustainable financing. In addition, feedback on the methodology and the developed Tools has been received from 198 stakeholders by the dissemination of dedicated questionnaires.

Step 5: Web-based/ Benchmarking Tool design

The final step consists of the design and the technical development of the benchmarking Tools, which will assist energy efficiency stakeholders, such as investors seeking to finance sustainable projects and developers that design and implement such projects. The Tools will perform a complete series of actions corresponding to the value-chain of energy efficiency financing. First, the risks and maturity of the inserted investment ideas together with their EU Taxonomy compliance are being evaluated, and then the projects that pass the pre-evaluation are classified into predetermined categories to identify the Triple-A ones. Finally, the last step of Tools matches the investments with state-of-the-art green financing schemes.
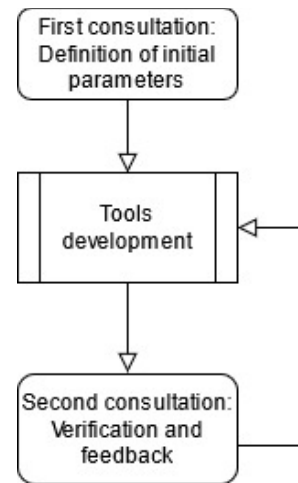
Figure 2: Stakeholder consolation steps

The Tools have been developed in Python 3.0 programming language, taking into advantage the versatility and power framework that it provides. In addition, the Python Django framework has been used to combine the individual scripts for the risk assessment, MCDA method, and criteria calculations into an integrated web-based environment with easy to use front-end and back-end functionalities. The back-end service utilises the SQLite database, while the app is hosted on the National Technical University of Athens' servers.

The Tools' design, functionalities, input-output data and the respective procedures, along with flowcharts, are presented and analysed in depth in Section 4: Web-based Benchmarking Toolbox.

3. Methods and Data

The proposed methodology is based on the actual needs of the EE financing sector, while the design parameters and requirements of the problem have been identified and verified through stakeholder consultation. All the respective processes, such as the identification, review, stakeholder consultation and development, have been realised within H2020 Triple-A project activities and are of vital importance for the assessment and mainstreaming of EE investments. For example, the ELECTRE Tri multi-criteria analysis method has been utilised for the necessary classification of the investments, the project risk elicitation methodology has been crucial for the identification of risky projects which could be immediately rejected from the Triple-A Assess Tool, and the Key Performance Indicators (KPIs) serve as the most suitable criteria

concerning the financial prosperity of the investments. Therefore, before presenting the Triple-A Tools step-by-step, it is essential to deliver the underlying methodological basis, referring to the tools' components.

3.1. Identification of the eligible projects.

The potential investments in EE measures covered by the presented methodology should be pertaining to the following five (5) sectors (i) Buildings, (ii) Manufacturing, (iii) Transportation, (iv) District energy networks, and (v) Outdoor lighting.

The project sectors have been identified in order to form a consistent typology that will include the majority of EE measures in such a way as to simplify the user experience of the developed Tools and batch them in a manner that will benefit the calculations of the respective benchmarking. The eight identified sectors are analysed in the following paragraphs:

- **Building sector**: Building sector projects ideas include any renovation compliant with the energy performance standards set in the applicable building regulations for major renovations transposing the Energy Performance of Buildings Directive (EPBD) or renovations that achieve energy savings of at least 30% in comparison to the baseline performance of the building before the renovation. The baseline performance and predicted improvement shall be based on a specialised building survey and be validated by an accredited energy auditor. Individual renovation measures and installation of renewables on-site that contribute to climate change mitigation by reducing GHG emissions and professional, scientific and technical activities necessary to support and validate buildings' renovation are also included in the Building sector [24]. Also, the construction of new buildings can be included if they meet at least the national requirements for nZEB and have a level of energy performance equivalent to the EPC rating of B (or above).

- **Manufacturing sector:** Under the Manufacturing sector are included projects of EE measures in industries in order to manufacture products with high EE, compared to relevant products already in the market, of manufacture products, key components and machinery that are essential for implementing projects that achieve high EE. More generally, any intervention in industries that aim to reduce energy losses and to improve the energy performance of equipment, machinery or install renewable energy systems can be included. Equally, improvement of existing industrial production lines and processes to achieve significant improvement of their EE and reduction of GHG emissions per unit of product are covered.

- **Transportation sector**: The Transportation sector includes measures in public transport vehicles, which can achieve substantial GHG emission reduction in terms of CO2e emissions per passenger- kilometre (gCO2e/km). The potential project should be referred to zero direct emissions land transport solutions, such as light rail transit, metro, tram, trolleybus, bus and rail, or fleets with direct emissions below 50 gCO2e/pkm until 2025 (non-eligible thereafter). Additionally, projects to finance the replacement of corporate fleet with new, more efficient passenger cars are included. The

vehicles can be light commercial vehicles and category L vehicles (2- and 3-wheel vehicles and quadricycles).

- **District Energy Networks**: As District Energy Networks are considered the construction projects, retrofit, expansion and operation of pipelines and associated infrastructure for distributing heating and cooling. For a project to be eligible, the system should meet the definition of efficient district heat/cool systems in the EU Energy Efficiency Directive and EIB energy lending policy regarding phasing out fossil fuels ("Supporting the energy transformation"). In detail, "efficient district heating and cooling" is defined as a district heating or cooling system using at least 50% renewable energy, 50% waste heat, 75% cogenerated heat or 50% of a combination of such energy and heat.

- **Outdoor lighting:** The Outdoor lightning sector includes any project focused on the renovation or renovation and expansion of existing outdoor lighting installation (e.g. street lighting, lighting of public squares) that lead to a high EE performance. All projects should foresee an appropriate control and measurement system.

3.2. Review/ Selection of Problem Parameters

The problem parameters are the key elements that form the criteria used in the methodology. The criteria, in turn, are defined and quantified by relevant indicators that can reflect EE project characteristics. In the presented methodology, the indicators for the benchmarking include data regarding the EU Taxonomy compliance of projects, risk estimations per country, sector, economic and financial KPIs based on project's characteristics, and the multi-benefits of projects as depicted by the potential contribution of EE measured by the Sustainable Development Goals. The typology and formation of the criteria have been inspired by the Final Standardised Tools Deliverable [25] published within the context of the Triple-A H2020 project. In the following paragraphs, each indicator type will be presented and explained.

- **EU Taxonomy compliance:** The methodology has pointed out the EU Taxonomy Technical screening criteria for each the identified project's sector. The criteria include thresholds regarding the technical characteristics of the proposed EE projects, which vary from one sector and type of activity to another. The technical screening criteria of the EU Taxonomy have been used to all the sectors, except the Outdoor Lighting sector, for which a dedicated set of criteria has been used, based on the outcomes of the PREMIUM LIGHT PRO project [26].

- **Risk Assessment:** The risk assessment of the potential investments in EE focuses on estimating in a qualitative and quantitative approach the total investment risk caused by a variety of factors. These factors are linked to the nature and principles of EE or country and sector individual characteristics [27]. The factors have been adequately identified and grouped into five (5) distinct risk categories. The categorisation aims to group the identified risk factors based on their conceptual characteristics and to facilitate the quantification procedure of the investment's risk. The aggregation of the risk parameters into a total investment risk

has been realised by applying a weighted average formula, as analysed in Section 3. The risk assessment methodology has been inspired by the Final Report on Risks of Energy Efficiency Financing and Mitigation Strategies Typology published within the context of the Triple-A H2020 project [28].

- **Economic KPIs:** The economic parameters could not be missing from a benchmarking methodology of potential investments. A selection of classic financial indicators, broadly used by investors, has been carried out, while the calculation methods have been parameterised to fit the cashflows and particularities of EE measures.

- The indicator's calculation has been based on EU Directives and Regulations on Cost-Benefit Analysis of Investment Projects, while data from EU official statistics have been used to provide a standardised, unbiased result. The indicators included in the benchmarking procedure are the Net Present Value (NPV), the Discounted Payback Period, the Internal Rate of Return (IRR), and the Cost-Effectiveness (or Avoidance Cost) [25].

- **Sustainable Development Goals:** An aggregated indicator has been included as a criterion in the benchmarking methodology, reflecting the necessity to improve EE in the case study countries and the identified sectors. The indicators used are official Eurostat statistics linked with the progress of the UN's Sustainable Development Goals. These indicators reflect the current situation of EE per country and sector, the situation regarding energy poverty, environmental pollution and greenhouse gas emissions, and energy security - sufficiency. Overall, a total of 11 sustainability indicators have been included, linked with the multi-benefits of EE and aggregated into a single criterion.

3.3. Review/ Selection of available methods

The review and selection of the most appropriate method is a crucial element enabling the benchmarking of the EE investment ideas to be performed in an affluent and efficient way. The problem parameters and requirements, such as the benchmarking of EE project ideas, in predefined classes based on several criteria, emphasised the need for a multi-criteria approach.

ELECTRE Tri was considered the most appropriate methodology to address the problem, as it accumulates key advantages that match the needs of the problem in question. The determinant factor that led to select the ELECTRE-Tri Multiple Criteria Decision Analysis as the benchmarking method is that it allows the classification of the candidate projects into predefined categories (alternatives) by constructing outranking relations and comparing each alternative to a limit [29].

Also, one of the strong points is that ELECTRE-TRI handles both qualitative and quantitative data, meaning that it can deal with the imperfect nature of knowledge [30]. It is a multi-criteria decision-making method which is mainly utilised for classification problems, providing labelling of the alternatives based on a set of conflicting and non-overlapping criteria.

Not to mention the discrimination thresholds that the selected method incorporates, which eliminate the serious drawbacks of

imprecision, ill-determination, and uncertainty that can undergo with the data available and used. The theoretical background, steps and processes performed in the ELECTRE Tri method are thoroughly presented in the relevant bibliography in [31], [32].

3.4. Calculation of the EE investments risk

In the context of the proposed methodology, an important pillar is calculating the investment's total risk, which was determined by combining both a qualitative and a quantitative approach. The major factors that affect the security of the implementation of investment that mainly consist of EE measures have been classified into the following five categories. Each risk category consists of risk factors that are estimated based on formulas or linguistic values, according to their nature.

- <u>Financial risk</u> is related to the creditworthiness of the applicant for the loan/financing. It consists of the risk factor "creditworthiness of the borrower". This factor indicates the financial capacity of the borrower to pay off his debt. It is considered a critical factor from the perspective of a financing institution or bank when considering giving a loan [33,34]. The risk factor is qualitative, and it is estimated by assigning linguistic values out of the scale ("Low", "Medium", "High", "Unknown") for evaluating the creditworthiness of the applicant for the loan.

- <u>Behavioural risk</u> is related to the rebound effect that may exist in the context of the inspected EE investment. It consists of the risk factor "rebound effect", which describes a specific behavioural bias in sites that have been implemented EE measures. This bias affects the end-user, and it usually emerges when the implementation of an EE project leads to lower costs for energy services combined with an increase in the demand for such services. It is a qualitative risk factor, and it is being quantified by assigning linguistic values out of the scale ("Insignificant ", "Low ", "Medium", "High", "Very High") to each type of project based on literature review and questions replied by the applicant of the project idea, as inspired by the Industrial Energy Accelerator project by UNIDO [35] .

- <u>Energy Market & Regulatory risk</u> consists of factors related to the country's energy prices and energy taxes volatility (and in which the investment will be implemented) and the requests for issuing work permits that may exist in the context of the inspected project. Energy prices and taxes volatility is associated with the price risk in EE investments, and it is a quantitative factor calculated based on the coefficient of variation (CV) of monthly values of Consumer Price Index (CPI) for the last fifteen years (01/2005-12/2019), as reported by OECD. The risk factor "request for issuing project permits" signifies the legislative complexity for the completion of a project and consists of a qualitative analysis considering the amount and type of project permits needed for the implementation of the project.

- <u>Economic risk</u> is differentiated from financial risk, and it is related to the country's economic environment that the investment takes place. The risk factor that facilitates the quantification of this project category is the "weak economic environment", which is connected to, among others, the country's interest rates, inflation, availability of finance, etc. [36]. It is a qualitative factor, quantified by the credit ratings

provided by Standard & Poor's (S&Ps) for the case-study countries.

- <u>Technological, Planning and Operational risk</u> is related to the technical complexity, the initial savings assessment, the implemented equipment, the project design, and the Operation & Maintenance of the inspected project. In more detail, the risk factors that are included in this Risk Category, are qualitative and they are quantified by assigning linguistic values out of the scale ("Insignificant", "Medium", "Very high") to each type of project based on questions answered by the applicant of the project. Specifically, the "Low quality of initial savings assessment" risk factor considers how the energy savings assessment and the baseline definition have been conducted. The "Implementation of low-quality equipment or poor project design" considers the existence of proof about the quality of the equipment to be installed, the experience of the team for planning and implementing the project and who conducts the technical implementation of the project. The "Inadequate Operation & Maintainance" considers the experience of the end-user in using and operating the proposed equipment, the existence of a maintenance plan, product warranties and Measurement & Verification protocols and standards.

Following the determination of the risk factors' values, the risk categories' values are then calculated by averaging the values of the risk factors of which each category is composed. The risk is being calculated by deploying the method of the weighted average

The total project risk value is the weighted arithmetic mean of the risk categories' values and is calculated as follows:

$$\textbf{\textit{Project risk value}} = \sum_{i=1}^{n} \textbf{\textit{w}}_i \times \textbf{\textit{R}}_i \qquad (1)$$

w_i : *i = 1, …, 5 are each risk category's weight*

R_i : *i = 1, …, 5 are each risk category's value*

The weights used in the equation emerge from the stakeholder consultation process that has been realised within the Triple-A project's activities. The weights values have been calculated through the respective answers in linguistic variable questions regarding the investors' preferences and their perception of energy efficiency projects' risk, and project developers' risk prioritisation according to the estimated impact on the EE projects implementation.

3.5. Final formation of the KPIs serving as classification criteria

The application of the ELECTRE Tri method, which was selected to follow through with the problem in question, is based on the elicitation of a series of criteria. These criteria emerge from the identified project parameters, as presented in Paragraph 3.2, and should follow the principles and restrictions implied by the ELECTRE-Tri methodological background. The final formation should produce non-overlapping criteria and which must affect directly the evaluation of the examined alternatives (EE investments). The selected criteria reflect the economic prosperity, the risk and the sustainability of the potential investment, based on specific KPIs.

The first two criteria consist of economic KPIs broadly used in the classic benchmarking of potential investments in the financing sector. The KPIs are calculated taking into account the expected economic savings of the potential EE investments, while the KPIs

parameters are based on the EU Directives and Regulations on Cost-Benefit Analysis of Investment Projects or reflected directly from EU official statistics to provide standardised unbiased results. The origin of the applied KPIs aspires to build confidence among investors and facilitate financing bodies and EE funds to detect and finance projects that meet the necessary criteria rapidly.

The **first criterion** is the economic indicator of <u>Cost-Effectiveness</u>. Cost-Effectiveness refers to the ratio of dis-counted benefits over the discounted costs. In its simplest form is a measure of whether an investment's benefits exceed its costs [37]. In the proposed methodology, the Cost-Effectiveness is calculated based on the project cost per kWh saved

The **second criterion** is also an economic indicator selected by the respective stakeholder of each potential investment. The selection can vary between the KPIs listed below:

- Net Present Value (NPV), reflecting the risk and cashflows discount by quantising it through the discount rate, the profitability of the investment by involving in the calculations the yearly income, the operational costs and the initial investment [38].

- Discount Payback Period, which quantifies the number of years necessary to recover the investment's budget while accounting for the time value of money [39].

- Internal Rate of Return (IRR) is a rate of return used in capital budgeting to measure and compare the profitability of investments.

- Cost-Effectiveness. Which refers to the ratio of dis-counted benefits over the discounted costs. In its simplest form is a measure of whether an investment's benefits exceed its costs [37]. In the proposed methodology, Cost-Effectiveness is calculated based on the project cost per kWh saved

The **third criterion** comprises the aggregated total investment risk, reflecting the likelihood of the unwanted case that the potential investment will not achieve the proposed savings in energy and thus fail to deliver the expected economic benefits. The aggregated total risk is calculated through the procedure described in detail in Paragraph 3.4.

The **fourth criterion** refers to the sustainability aspects of the proposed investment. Expect the financial and risk criteria, sustainability criteria are considered essential and have been incorporated into the methodology, as EE investments provide multi-benefits in social and environmental aspects, except energy savings. For example, they contribute to environmental and social factors while reducing energy consumption, leading to lower GHG emissions, improving comfort in living spaces, and industry productivity.

The sustainability criterion for the proposed methodology has been based on the United Nations' Sustainable Development Goals indicators have been selected based on quantitative analysis, as recorded by Eurostat's statistical indicators. These indicators reflect the current situation of EE, energy poverty, and environmental pollution. The identified indicators are directly linked with the United Nations' Sustainable Development Goals (SDG) Agenda [40], mainly to those related to the energy sector and environmental protection. The criterion emerges from the aggregation of statistical data per country or per country and sector, depending on the dataset's nature.

The utilised SDG factors that were aggregated in order to construct the SDG criterion for the ELECTRE Tri method are presented in Table 1 Selected SDG indicators.

Table 1: Selected SDG indicators.

Name	Criterion
Arrears on utility bills	It reflects the share of (sub)population (%) having arrears on utility bills, based on the question "In the last twelve months, has the household been in arrears, i.e. has been unable to pay on time due to financial difficulties for utility bills (heating, electricity, gas, water, etc.) for the main dwelling?"
Total population living in a dwelling with a leaking roof, damp walls, floors or foundation, or rot in window frames or floor	It indicates the share (%) of the population experiencing at least one of the following basic deficits in their housing condition: a leaking roof, damp walls, floors or foundation, or rot in window frames or floor.
Population unable to keep home adequately warm by poverty status	It indicates the share (%) of the population, who are unable to keep home adequately warm. Data for this indicator are being collected as part of the EU Statistics on Income and Living Conditions (EU-SILC) to monitor the development of poverty and social inclusion in the EU.
Primary energy consumption	It quantifies the Gross Inland Consumption in toe, excluding all non-energy use of energy carriers (e.g. natural gas used not for combustion but for producing chemicals).
Energy import dependency	The criterion shows the share (%) of the total energy needs of a country met by imports from other countries. It is calculated as net imports divided by the gross available energy
Final energy consumption in the industry sector	It includes all the energy supplied to the industry sector in toe; excluding deliveries to the energy transformation sector and the energy industries themselves).
Final energy consumption in the transportation sector	It measures the energy consumption of the transportation sector in toe, excluding deliveries to the energy transformation sector and the energy industries themselves.
Final energy consumption in other sectors or commercial and public services	It indicates the energy supplied to non-categorised sectors, commercial and public services in toe.
Final energy consumption in households per capita	The indicator measures how much electricity and heat every citizen consumes at home (kgoe/capita), excluding energy used for transportation. Since the indicator refers to final energy consumption, only energy used by end consumers is considered.
GHG emissions from energy consumption	The data are based on measures of the European Environmental Energy Agency and represent the GHG emissions caused by the energy sector in ktn $CO_{2\text{-eq}}$.
GHG emissions from the industrial sector	Similar to C_{10}, the C_{11} criterion contains the GHG emissions (in ktn $CO_{2\text{-eq}}$) caused by the industrial sector, as reported by the European Environmental Energy Agency.

The four criteria mentioned above are calculated automatically by the Triple-A Web-based Benchmarking Toolbox presented in Section 4. The calculation is being performed in the back-end of the respective Tool, while the user only needs to provide the data required for the calculations and reply to the requested questions. As described thoroughly in Paragraph 3.7, the four criteria feed a MultiCriteria Decision Analysis Method, which in turn performs the final benchmarking of the Web-based Benhcmaking Toolbox.

4. Web-based Benchmarking Toolbox

The methodology, typology, and expected results analysed in the previous paragraphs have come into existence by the implementation of web-based Tools [1]. The Tools form part of an integrated Toolbox, namely the Triple-A Toolbox, which has been implemented in the Python 3.0 programming language.

The Toolbox aims to perform an initial filtering of the EE project ideas, benchmark them into predefined classes, and connect them with investors via concrete financing methods. The examined projects, along with their attributes (project country, project sector, aggregated project risk etc.), are saved into a database in order to be identifiable and open for financing requests.

The Toolbox consists of three distinct Tools, namely the Assess Tool, the Agree Tool and the Assign Tool. The project architecture is designed in a way that each Tool can be used separately, thus offering interoperability among the different modules. The steps for each module of the Triple-A Toolbox are presented in Figure 3.
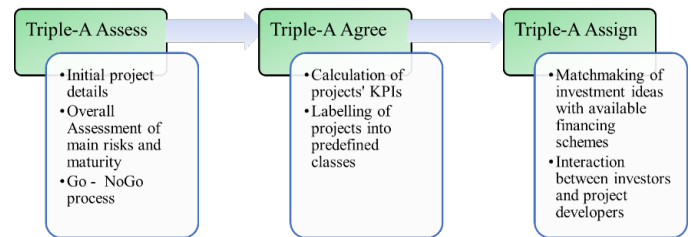
Figure 3: Web-based Tools functionalities

4.1. Assess Tool

The Assess Tool aims to perform initial filtering and screening of the EE project ideas. The main idea of the Assess Tool is the overall assessment of the main risks and the maturity, the candidate EE investments, and their EU Taxonomy compliance. To evaluate the risk of each EE investment, its main characteristics are considered, such as the country of implementation and the technology implemented within its context. On the one hand, the considered risks are related to the region where the project is implemented, the proposed technologies that are used, and the project management. On the other hand, the maturity of the investments is related to their readiness for implementation. In this respect, several key factors that could possibly affect the project's success and its access to sustainable financing have been considered (e.g., risk level, size of the investment, type of EE projects, EE measures eligible, EU Taxonomy compliance etc.), serving as eligibility criteria, aiming to assure the security and trust of the potential investments, which is a necessary action in order to proceed to further analysis.

The flowchart of the Assess Tool methodology is presented in Figure 4:
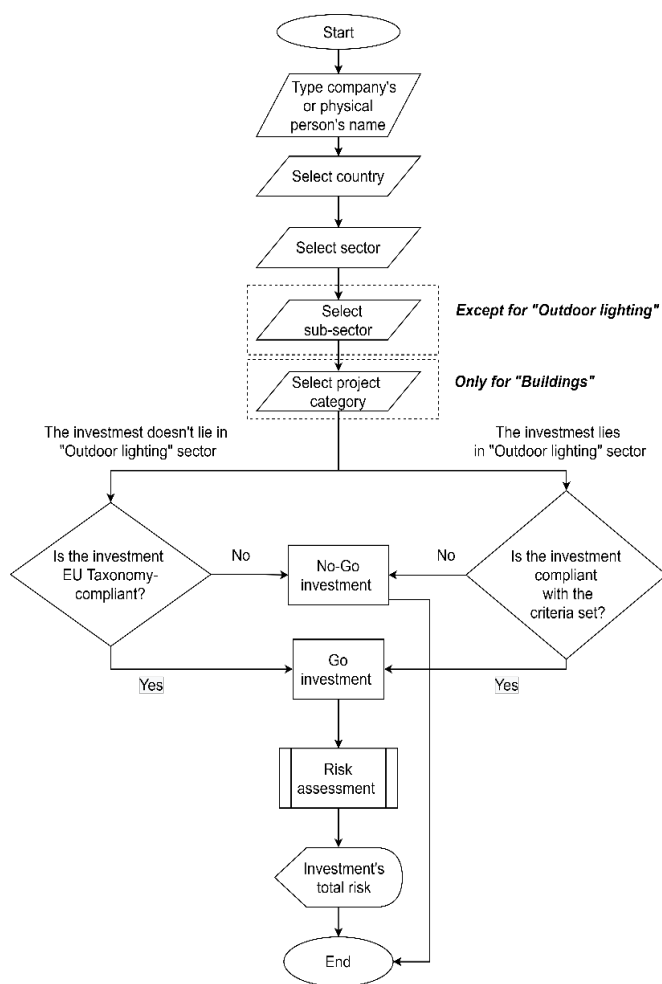
Figure 4: Assess Tool Flowchart

From the user's perspective, this Tool includes the following steps:

Step 1: Inserting the basic information of the candidate EE project (Company's or Physical person's name, country, sector).

Step 2: Selecting the project's subsector (except for the "Outdoor Lighting" sector). The project sectors and subsectors are presented in **Table 2**.

Step 3: Selecting the project category (Only for the building's sector; Table 4). In case that the investment concerns a building, select the project category, as presented in Table 3.

Table 2: Assess Tool Sectors

A/A	Sectors	Sub-sectors
S₁	Buildings	Residential
		Non-Residential
S₂	Manufacturing	Hydrogen
		Iron and Steel
		Aluminum
		Cement
		Low carbon technologies
		Fertilisers and Nitrogen
		Other organic basic chemicals
		Other inorganic basic chemicals
S₃	Transportation	Public Transport

		Passenger cars and light commercial vehicles
S₄	District Energy Networks	District Heating / Cooling Distribution
		Installation and operation of electric heat pumps
		Cogeneration of Heating / Cooling and Power
		Production of Heating / Cooling
S₅	Outdoor Lighting	-

Table 3: Assess Tool project categories.

Sectors	A/A	Project categories
Buildings	P₁	Building envelope retrofits
	P₂	HVAC&R retrofits
	P₃	Lighting appliances' retrofits
	P₄	Automatic control retrofits
	P₅	RES installations
	P₆	Construction of new buildings
Manufacturing	P₇	Manufacturing-specific retrofits
Transportation	P₈	Purchase of new vehicles
District Energy Networks	P₉	District Energy Networks retrofits/ expansion
Outdoor Lighting	P₁₀	Outdoor lighting retrofits

Step 4: After having imported all the project-specific information, select whether the investment is EU taxonomy eligible. Depending on the project sector, as well as the additional information is given (Project Type, Project Category, etc.), an EU taxonomy list appears, including the thresholds that need to be met by the investment and some explanations on the requirements, such as the metrics of each threshold. Then, the user should carefully check if the investment fulfils all the thresholds. In case that all the requirements are met, then the Yes checkbox should be selected, stating that the investment is taxonomy compliant. Finally, the user must fill in a checklist with taxonomy criteria, according to whether the investment is compliant with the criteria set in each case or not.

Step 5: Message notification to inform the user whether the project is Go or No-Go. If the answer to the EU Taxonomy compliance questions sets a project as No-Go, the user is notified immediately, without being able to complete the remaining steps of the Tool.

Step 6: Answer a series of project-specific questions, which are taken into consideration for the aggregated investment risk calculation.

Step 7: Calculation of the project's aggregated risk, according to the methodology described in **Paragraph 3.4**.

Step 8: Analytic presentation of the project's risk calculations, including the calculation for each distinct risk category, as described in Paragraph 3.4 and the total calculated risk of the investment.

4.2. Agree Tool

The Agree Tool implements the second pillar of the proposed benchmarking scheme for the EE project identification and labelling. The Agree Tool can only be applied to EE project ideas

that have successfully passed the Assess Tool and have been identified as "Go" projects. The Tool supports and facilitates the identification of the so-called Triple-A investments [12], e.g. the investments that have a high performance in the selected KPIs, as described in Paragraph 3.5. In other words, Triple-A EE investments are defined as the ones with extremely strong potential to meet their energy-saving targets, already from their conceptual phase. The candidate projects that are inserted into the Agree Tool are classified into one of the following categories: "Triple-A", "Reserved", or "Rejected", according to their performance on the evaluation criteria.

The flowchart of the Triple-A Agree Tool methodology is presented in Figure 5.
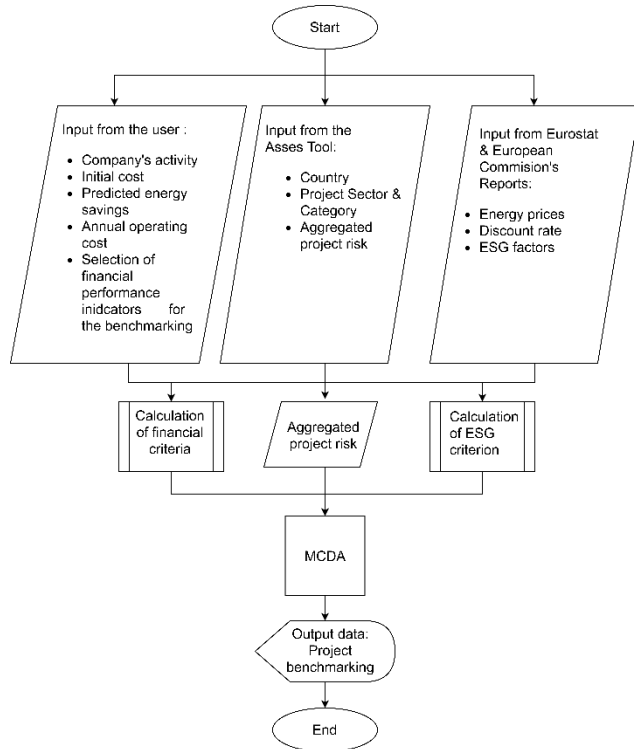
Figure 5: Agree Tool flow chart.

The steps for the Agree Tool from the user's (investor, banker, financer) point of view are described in detail as follows:

Step 1: Definition of project country and project sector in order to set the ELECTRE Tri algorithm's thresholds to the customised values.

Step 2: Select the criteria which will be used for the evaluation of the project and also insert the required input for the calculation of the criteria. The ELECTRE Tri algorithm is applied on the basis of 4 evaluation criteria, as described in Paragraph 3.5. There are two financial criteria (K1 & K2) combined with one aggregated risk criterion (K3) and one SDG criterion (K4). The criteria of ELECTRE Tri must comply with the restriction of consistency. Therefore, two financial KPIs should be applied to the algorithm. The aggregated risk criterion is fed to the model from the Assess Tool's risk calculation process. Finally, the SDG criterion stems from the methodology discussed in Paragraph 3.5.

Step 3: Weight adjustment of the ELECTRE Tri criteria according to the importance of each factor and the user's preferences. The

established weights reflect the effect of each criterion on the final benchmarking. Extreme weight values (0% or 100%) are rarely assigned to a criterion. At the same time, the sum of all criterion weights should be 100%. The Agree Tool supports default values for weights which can be easily changed by the user from the Tool's interface. The criteria thresholds of the benchmarking classes have been set according to outcomes of similar web-based Tools and databases that accumulate data and knowledge from implemented EE projects, such as the DEEP platform [41], as well as from input provided by experts via the stakeholder consultation. Until March 2021, a number of almost 100 users (project partners, Triple-A project Advisory Board Members and external stakeholders) have been signed up and tested the developed Tools, providing valuable feedback. In addition, the Triple-A Tools have been tested with real projects that have been collected in an EE project pipeline within the Triple-A project activities.

Step 4: After the execution of the multi-criteria algorithm analysis, the candidate EE project is classified into one of the predefined categories: Triple-A, Reserved, and Rejected. The three categories are explained in the paragraphs below:

Triple-A Projects: This category includes projects which could be considered as the best investment opportunities and, thus, merit attention by the funding organisations. These are investments that demonstrate extremely strong potential to meet their energy-saving targets, from their primitive phase, when they are still considered as project fiches from the funding institutes. The above characterisation comes in accordance with the definition of the respective Triple-A investment grade, which refers to investments with a strong capacity to fulfil their financial commitments by achieving the expected performance targets.

Reserved Projects: This category includes a series of projects which are profitable, but they do not guarantee that they will fulfil the predefined performance targets. Such projects have achieved a good but not outstanding performance according to the evaluation criteria. Even though they have not managed to achieve the top score in the evaluation methods, they are capable of repaying the initial capital invested and of contributing significantly to the energy savings.

Rejected Projects: This category includes projects which have a poor total performance according to the evaluation criteria. Their risk is higher than the predefined threshold, and they seem incapable of recovering the total investment. Also, projects which were labelled as "No-Go" during the Triple-A Assess phase are also characterised as Rejected.

4.3. Assign Tool

The Assign Tool implements the final step of the integrated methodological framework. The main functionality of this Tool is the matchmaking of investment ideas, which have successfully passed the Assess and Agree tools, with a list of potential financing schemes. Thus, it aspires to support all involved parties to achieve project delivery, such as contracts, underwriting procedures, etc.

The Assign Tool includes numerous user interfaces in order to facilitate all beneficiaries and EE financing instruments. The platform's main target groups are, on the one hand, financing bodies and, on the other hand, companies and project developers. From the aspect of financing bodies, commercial or green investment banks are included, as well as investment funds, developers and managers of financial products, Unit Investment Trust and several types of financers that target financing Triple-A

opportunities to include in their portfolio. From the aspect of project developers, the Tool is oriented towards EE companies, credit professionals, ESCOs, construction companies and management investment companies that try to attract financing in order to implement their projects.

The flowchart of the Assign Tool methodology is presented in Figure 6.

The Assign Tool provides three different types of financing schemes, aiming to be as flexible as possible and to provide several options for both financing bodies and project developers. The three supported types of financing are (a) *Green Loans and Mortgages* and (b) *Green Bonds,* and (c) Energy Efficiency Auctions.

Loans are one of the most traditional financing mechanisms, although their characteristics may vary significantly for different financing cases. During the last decades, financing institutes have prioritised environmental issues and objectives, and Green financing products have emerged as a new trend. [42]. The developed Tool has classified the financing cases to Green Loans and Green Mortgages. The need for such classification has emerged because of the popularity of such financial products, even though mortgages are a form of a loan.

The Assign Tool supports the procedure of financing proposals between investors and project developers. Investors are enabled to send financing proposals for projects, having the flexibility to set the preferred loan parameters. Any type of financing proposal is acceptable, e.g., investors are not obliged to cover the total cost of a project. They could partially finance the project and select the preferred percentage of leverage. Project developers can assess the proposal and accept or reject it by using their personalised interface of the Tool.

The second type of financing, which is supported by the Triple-A Assign Tool is Green Bonds. A bond is a type of investment that represents a loan between a borrower and a lender, being considered a debt investment. Green Bonds are any type of bond instrument where the proceeds will be exclusively applied to finance or re-finance, partly or fully, new, or existing eligible Green Projects [43]. The Assign Tool introduces a new methodological approach, enabling an investor to come across companies and project developers to aggregate their projects into a portfolio. EE projects are often small and fragmented in nature in comparison with large energy infrastructure projects such as power plants [44]. Project aggregation expedites EE financing, as bundling small and medium-size projects into a larger build-up reduce the overall risk and is more appealing to investors.

The Assign Tool creates the ideal circumstances to create links between primary entities that are involved in the procedure of bond issuance. Based on the "Proposal for an EU Green Bond Standard" [45] published in June 2020 by the EU Technical Expert Group on Sustainable Finance, the Assign Tool supports Green Bond issuers to be compatible with the EU Green Bonds Standard by providing all necessary functionalities and information.

The third and final financing method covered by the Assign Tool is the EE auctions. The EE auctions is a policy instrument aiming to achieve energy savings at highly attractive prices and avoiding deadweight effects associated with financial support [46]. The procedure resembles the renewable energy auctions already implemented in a number of countries. Lessons learned and previous experience of EE bidding and auctions in Switzerland have proved that tendering could possibly help to save 26% of

public funds as compared to more traditional subsidise programmes [47].
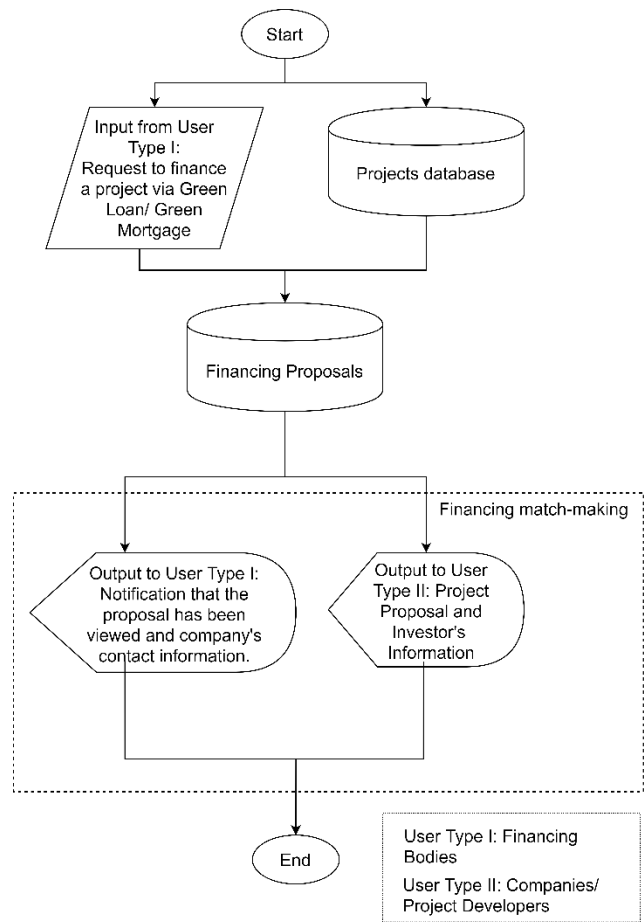
Figure 6: Assign Tool flowchart

The Assign Tool supports the publication of EE auctions in the web-based Tool, enabling interested key actors to get informed about the available auctions. In addition, with the support of the database of projects, followed by the respective information about each project characteristic, stakeholders can filter projects and identify the ones that perfectly match the parameters of the auction. As well, the creation of a printable portfolio of projects is obtainable, and the contact details of each company or project developer are accessible so as to proceed to further actions.

The Triple-A Tools are accessible through the Triple-A Standardised Toolbox platform (https://aaa-h2020.eu/tools), which was implemented in a user-friendly environment, rendering the whole benchmarking procedure flexible and easy to follow, while also offering a communication canal between project developers and financing bodies through flexible options such as green loans and green bonds and energy efficiency auctions.

5. Conclusions

It is widely accepted that many investments in EE are well below the required level to meet the established long-term targets, both in economic and energy savings levels. The present manuscript introduces both an integrated **methodology** to facilitate the financing of energy efficiency projects, as well as a fully-

featured **Web-based Tool**, namely the **Triple-A Standardised Tools,** that incarnates that methodology.

The Triple-A Standardised Tools aspire to set the way for **identifying** and **financing** Triple-A investments, based on a solid methodological procedure that takes into consideration many different aspects of the examined investment. The Triple-A Tools are designed with an integrated process, constituting them orientated towards different types of stakeholders in EE financing.

In other words, the developed Tools facilitate project developers to compare and label their projects in a standardised way through the Assess & Agree Tools, while also providing easy access to financiers, bankers, and investors to finance bankable green projects through the Assign Tool. The Triple-A Standardised Tools have been practically presented to project developers and investors, receiving positive feedback and raising interest in their functionalities through targeted stakeholder consultation activities.

The **methodology** described in the present paper facilitates EE financing in a practical way, differentiating from existing methodologies, by providing integrated standardized procedures. Also, the methodology:

- Offers a concrete typology of sectors and project types, risk categories, KPIs, benchmarking, and financing Tools widely accepted and known by stakeholders and key actors that do business in the EE sector.

- Incorporate state-of-the-art elements of sustainable financing, such as the EU Taxonomy and the UN's Sustainable Development Goals,

- Ensures interoperability with similar databases and Tools, providing an integrated experience and standardisation to the users.

- Supports innovative financing methods, which are still in the rollout stage but are expected to change the scenery of EE financing, such as the EE auctions.

The **Tools** developed based on the described methodology provide an integrated approach in a variety of aspects of EE financing:

- the **Assess Tool** addresses the need for evaluation of the risks and maturity of the examined EE investment ideas along with their EU Taxonomy compliance.

- the **Agree Tool** provides a solid classification mechanism for the projects that successfully pass the Assess Tool, identifying the remarkable project as the Triple-A ones. Finally,

- the **Assign Tool** supports the matchmaking of the potential investments with state-of-the-art green financing schemes and creates a community of EE stakeholders (Companies, Project Developers, Investors, Financing Bodies).

The Triple-A Tools provide a one-stop-shop approach to connect project developers with investors, by assessing the maturity and the EU Taxonomy compliance of potential EE projects, while benchmarking them in a standardised way. The respective tools have been already used and tested by stakeholders within the consultation activities of the H2020 Triple-A project [48]. Valuable feedback has been provided and considered and updates have been realised in the Web-based Benchmarking Tool.

In addition, in order to confirm that the developed Tools provide added value to energy efficiency professionals, respective questions have been added to the stakeholder consultation questionnaires and the 53% have replied that they find the Triple-A Web-based benchmarking Tool useful to their profession.

Furthermore, engineering and energy efficiency companies have demonstrated interest in learning more about the Tools and their functionalities, while the Tools' developing team was asked to make a live presentation in the Corporate Social Responsibility School 2021 [49] in July 2021. The Tools will also be demonstrated in the Regional Workshop organised by the Triple-A project. In the events mentioned above participate green and sustainable financing investors, engineering companies, project developers that develop energy efficiency projects and policymakers.

Until August 2021, 627 users have visited the Triple-A Tools online platform, while almost 120 stakeholders have been signed up. In addition, 113 projects have been inserted into the Tools in order to be benchmarked and connected with financing methods. Furthermore, the Triple-A Tools have reached 336 participants in the Triple-A events (workshops and Capacity Building Webinars), 72 participants in 59 bilateral meetings and 17 meetings with the Triple-A Advisory Board Members and 198 replies in the respective questionnaires.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

References

[1] A. Papapostolou, F.D. Mexis, E. Sarmas, C. Karakosta, J. Psarras, "Web-based Application for Screening Energy Efficiency Investments: A MCDA Approach," 11th International Conference on Information, Intelligence, Systems and Applications, IISA 2020, 2020, doi:10.1109/IISA50023.2020.9284403.

[2] International Energy Agency (IEA), Net Zero by 2050, Paris, 2021.

[3] Council of Europe Development Bank, Energy Poverty in Europe: How energy efficiency and Renewables can help, 2019.

[4] OECD/IEA, Energy Efficiency 2018: Analysis and outlooks to 2040, 2018.

[5] M. Andrei, P. Thollander, I. Pierre, B. Gindroz, P. Rohdin, "Decarbonization of industry: Guidelines towards a harmonized energy efficiency policy program impact evaluation methodology," Energy Reports, **7**, 1385–1395, 2021, doi:10.1016/J.EGYR.2021.02.067.

[6] A. Papapostolou, C. Karakosta, K.-A. Kourti, H. Doukas, J. Psarras, "Supporting Europe's Energy Policy Towards a Decarbonised Energy System: A Comparative Assessment," Sustainability, **11**(15), 4010, 2019, doi:10.3390/su11154010.

[7] Efficiency Financial Institutions Group (EEFIG), Energy efficiency – the first fuel for the EU economy, 2014.

[8] A.B. Jaffe, R.N. Stavins, "The energy-efficiency gap," Energy Policy1, **22**(10), 804–810, 944.

[9] Horizon H2020 Triple-A project, Updated Web-Based Database on Energy Efficiency Financing and Supporting, , Deliverable 3.5, 2021.

[10] C. Karakosta, A. Papapostolou, Transformation Pathways towards a Clean,

Secure and Efficient European Energy System: A MCDA Approach, Nova Science Publishers, Inc., 2019.

[11] Financing Energy Efficiency | EEFIG Underwriting Toolkit, May 2021.

[12] H. Doukas, "On the appraisal of 'Triple-A' energy efficiency investments," Energy Sources, Part B: Economics, Planning, and Policy, **13**(7), 320–327, 2018, doi:10.1080/15567249.2018.1494763.

[13] T. Fawcett, G. Killip, "Re-thinking energy efficiency in European policy: Practitioners' use of 'multiple benefits' arguments," Journal of Cleaner Production, **210**, 1171–1179, 2019, doi:10.1016/J.JCLEPRO.2018.11.026.

[14] C. Karakosta, K. Papapostolou, G. Vasileiou, J. Psarras, Financial Schemes for Energy Efficiency Projects: Lessons Learnt from In-Country Demonstrations, Academic Press, 2020.

[15] T. Tsoutsos, S. Tournaki, E. Farmaki, P. Sonvilla, P. Lensing, J. Bartnicki, A. Cobos, M. Biscan, "Benchmarking Framework to Encourage Energy Efficiency Investments in South Europe. The Trust EPC South Approach," Procedia Environmental Sciences, **38**, 413–419, 2017, doi:10.1016/j.proenv.2017.03.125.

[16] N. Carbonara, R. Pellegrino, "Public-private partnerships for energy efficiency projects: A win-win model to choose the energy performance contracting structure," Journal of Cleaner Production, **170**, 1064–1075, 2018, doi:10.1016/j.jclepro.2017.09.151.

[17] LAUNCH H2020 | Sustainable Energy Assets in Europe, May 2021.

[18] RenOnBill | About, May 2021.

[19] The Project - EEnvest, May 2021.

[20] E2DRIVER, May 2021.

[21] X-tendo Toolbox, May 2021.

[22] C. Karakosta, A. Papapostolou, Linking Stakeholder Engagement to Multiple Future Policies in the European Energy Sector, 383–392, 2020, doi:10.4018/978-1-7998-1196-1.ch022.

[23] A. Ishizaka, P. Nemery, Multi-Criteria Decision Analysis: Methods and Software, Wiley, West Sussex, 2013.

[24] C. Karakosta, A. Papapostolou, J. Psarras, Concrete Steps towards the Promotion of Renewable Energy Deployment in an Effort to Tackle Climate Change, Nova Science Pub Inc, 2016.

[25] F.D. Mexis, K. Touloumis, A. Papapostolou, C. Karakosta, Deliverable 4.2, Final Triple-A Standardised Tools, Triple-A Project: Enhancing at an Early Stage the Investment Value Chain of Energy Efficiency Projects funded by European Union's Horizon 2020 Grant Agreement No 846569, Athens, 2021.

[26] Austrian Energy Agency, LED Street Lighting Procurement & Design Guidelines, developed for "Premium Light Pro" Project, funded by European Union's Horizon 2020, Grant Agreement No 695931; Road lighting - Part 5: Energy performance indicators, CSN EN 13201-5, 2015, 2017.

[27] T. Loureiro, M. Gil, R. Desmaris, A. Andaloro, C. Karakosta, S. Plesser, "De-Risking Energy Efficiency Investments through Innovation," Proceedings, **65**(1), 3, 2020, doi:10.3390/proceedings2020065003.

[28] Triple-A Final Report on Risks of Energy Efficiency Financing and Mitigation Strategies Typology, Athens, 2020.

[29] F. José, S. Greco, M. Ehrgott, Multiple Criteria Decision Analysis: State of the Art Surveys, Springer Nature, 2006.

[30] B. Roy, J.R. Figueira, J. Almeida-Dias, "Discriminating thresholds as a tool to cope with imperfect knowledge in multiple criteria decision aiding: Theoretical results and practical issues," Omega, **43**, 9–20, 2014, doi:10.1016/j.omega.2013.05.003.

[31] W. Yu, Aide multicritère à la décision dans le cadre de la problématique du tri: concepts, méthodes et applications, Diss. Université Paris IX-Dauphine, 1992.

[32] V. Mousseau, R. Slowinski, P. Zielniewicz, ELECTRE TRI 2.0 a Methodological Guide and User's Manual, Laboratoire d'Analyse et de Modélisation de Systèmes pour l'Aide à la Décision (LAMSADE), Paris, 1999.

[33] E. Mills, "Risk transfer via energy-savings insurance," Energy Policy, **31**(3), 273–281, 2003, doi:10.1016/S0301-4215(02)00040-X.

[34] L. Wang, J. juan Peng, J. qiang Wang, "A multi-criteria decision-making framework for risk ranking of energy performance contracting project under picture fuzzy environment," Journal of Cleaner Production, **191**, 105–118, 2018, doi:10.1016/j.jclepro.2018.04.169.

[35] Industrial Energy Accelerator (2020): Energy Efficiency Risk Management - A Report for Financial Intermediaries in Brazil. Report prepared by adelphi (Germany) and SITAWI (Brazil) under the UNIDO lead IEA initiative. January 2020.

[36] T. Pettinger, Factors affecting investment, Economics Help, 2019.

[37] U.S. Environmental Protection Agency, "Understanding Cost-Effectiveness of Energy Efficiency Programs," 2008.

[38] EBF, Sustainable Finance.

[39] Bankers Almanac, Standard & Poor's Definitions, 2018.

[40] United Nations, SDG Indicators, Jul. 2021.

[41] Energy Efficiency Financial Institutions Groups (EFFIG), De-risking Energy Efficiency Platform, Apr. 2020.

[42] D. Kahlenborn, C. Annica, I. Georgiev, F. Eisinger, D. Hogg, Defining 'green' in the context of green finance, Brussels, Belgium, 2017.

[43] Internatioanl Capital Market Association (ICMA), Green Bond Principles, May 2020.

[44] Aggregation and bundling of loans for Energy Efficiency – the holy grail for the European Investment Bank: An opportunity for Housing Federations? | Build Up, Dec. 2020.

[45] Technical Expert Group on Sustainable Finance, "Teg Report Proposal for an Eu Green Bond Standard," (June), 1–79, 2019.

[46] P. Radgen, K. Bisang, I. Koenig, "Competitive tenders for energy efficiency – lessons learnt in Switzerland," in ECEEE INDUSTRIAL SUMMER STUDY PROCEEDINGS, 81–89, 2016.

[47] Swiss Office for energy, Energy efficiency bidding and auctions: Ten years of experience in Switzerland, 2019.

[48] H2020 Triple-A Project, Aug. 2021.

[49] CSR School - Corporate Sustainability & Responsibility, Aug. 2021.

A Task-based Paradigm for Promoting an Alternative Thinking Style in Teaching Mathematics

Mikhail Rodionov[1], Zhanna Dedovets[*,2], Irina Akimova[2]

[1]Department of Informatics and Methods of Teaching Informatics and Mathematics, Penza State University, Penza, 440026, Russia

[2]Faculty of Humanities and Education, The University of the West Indies, St. Augustine Campus, Trinidad and Tobago

A B S T R A C T

The article identifies an alternative style of thinking as one of the important components of human intellectual development. It is shown that it can be effectively implemented in mathematics lessons at school. The purpose of this study is to develop and substantiate a strategy for the formation of an alternative style of thinking among students in mathematics lessons. The innovation in this research lies in the development of a new approach to the formation of an alternative style of thinking of students, involving the progression of schoolchildren up the "ladder of levels", purposefully correlating task structures with their "alternative analogs". There are essential research findings. The levels of formation of the alternative style of thinking of schoolchildren are defined and their multiple characteristics are given. It is shown that as the main means of actualizing an alternative style of thinking, it is advisable to set tasks that provide alternative options for analyzing the elements of their content area. The stage-by-stage work of students as they move up the "ladder of levels" is presented. Methodological recommendations for teachers of mathematics have been created and partially tested. These were presented at several seminars / training sessions and were successfully applied in practice throughout the year. Statistical processing using Pearson's criterion $\chi 2$ was applied at the end of the year to the results of the performance of special tasks by 52 students of grades 3-4 of one of the schools in Penza. Some students applied the traditional method (27 students), and the remainder applied the methodology proposed by the authors (25 students). Analysis of the results revealed a higher level of mastery of the alternative style of thinking among students in the experimental group.

1. Introduction

Being ready to perceive novel and unusual feelings, contradicting ideas initially estimated as the only correct ones, is an indicator of human intellectual and personal development. According to Vladimir Zinchenko, a person develops when he "rises above the available types of activity, chooses some, or builds a new one" [1].

The indicator under consideration, being an alternative thinking style, presupposes openness to all possible aspects of experience, including an unusual one; freedom from personal bias, stereotypes, prejudices and other external constraints. It admits a solution to problems opposite to generally accepted ones, based on the willingness of a person to "contradict himself", being the basis for the formation of tolerance of any dissenting opinion and action in the future [2–12].

This readiness is one of the factors in overcoming stereotypes. Provided an individual considers a particular situation from different angles, takes into account various points of view on the essential relationships between concepts established in a situation, he will be able to at least partially decentralize his thinking, and overcome his initially unilateral perspectives.

Despite numerous studies somehow affecting the role and place of an alternative style of thinking in the intellectual development of a person [2, 3, 7, 8, 10, 13, 14], the patterns and possibilities of the formation of such a style in the educational space at school has not been considered as an object of special research. In particular, issues of the main conditions and possibilities for the formation of the considered phenomenon in school education, the stages and levels of such formation, the diagnostic means to achieve these levels, the nature of educational goals in various academic subjects targeting the actualization of

249

alternative forms of educational and research activity, have not been considered by the scientists [14, 15].

As a result, purposeful formation of an alternative thinking style remains outside the attention of teachers and methodologists, leading to passive thinking in school students. In particular, they are not ready to transit from standard methods for performing educational and research activities to alternative methods. In other words, these standard methods for performing activities are fixed in their minds as rigid algorithms that do not allow any structural transformation.

Thus, the above conditions determine the relevance of the proposed research.

2.   Problem Statement

An alternative style of thinking can be effectively realized within the framework of mathematical content, since the presence of deep and often hidden connections between different branches of mathematics, different "mathematical languages" and general methods provide a possibility to correlate different perspectives on various mathematical concepts, facts and patterns. These possibilities are described in the works of many well-known mathematicians and methodologists [1, 4, 8, 15–20, etc.].

In particular, the actualization of the alternative style of thinking was very vividly illustrated in [15]. The great scientist correlated numbers and algebraic quantities: "Calculations are made either using numbers, as in ordinary arithmetic, or using species as in algebra. Both methods are based on the same principles and lead to one goal, but arithmetic using a specific and particular way, and algebra applying an indefinite and general way" [15]. However, "in arithmetic, when solving problems, one only passes from the given quantities to the sought ones; in algebra one works backwards, passing from the sought quantities, being considered as data, to the given values, being considered as the sought ones, in order to come to the conclusion or the equation, from which it would be possible to find the sought quantity" [15].

Newton implemented these considerations in finding a binominal series expansion of the function $(1 + x)^{\propto} (\forall \propto)$,, and in calculating the area bounded by the curves $(1 - x^2)^{\frac{n}{2}}$ for various natural numbers n [21]. Newton was well aware that, according to the logic of his construction, the obtained expansion was hypothetical, and not a proven fact. Therefore, he proposes other ways to confirm the correctness of the obtained expansion, which are of interest in the context under consideration.

In particular, Newton applied Vièta's method for square root extraction for numbers to polynomials [21]. Meanwhile, he formulated the rule for "square root extraction" from a polynomial, and he explained this rule with examples, in particular, an example of obtaining a binomial series for [21]

$$(1 + x)^{\frac{1}{2}} = \sqrt{1 + x}$$

$$\sqrt{1 + x} = 1.$$

Despite the fact that there are tasks for actualizing an alternative style of thinking in modern textbooks in mathematics, preparation materials for the final certification of school students,

and materials for various Olympiads, this orientation has not been fully implemented in mass practice of math education.

According to our observations and numerous publications, many school students realize the possibility of an alternative consideration of a particular problem situation only as reasoning that must be remembered and applied when solving problems of a specific type, rather than as a universal approach to its analysis [9–11, 22–24]. To prevent this, an adequate pedagogical format, considering the purposeful actualization of an alternative style of thinking as one of the main tasks of school mathematics education, is urgent.

Considering the above, the purpose of this study is to define theoretical foundations for the formation of an alternative thinking style for school students in the educational process, and to develop an appropriate methodological structure to ensure such thinking goes beyond the range of stereotyped patterns and ideas.

3.   Literature Review

Various aspects of an alternative thinking style, being an adequate indicator of personal intellectual development, have been considered by many scientists who associated it with the categories of decentralized, lateral, tolerant, or divergent thinking. The absence of this style is evident in the unilateral and tendentious thinking of a person using mainly stereotyped patterns and algorithms [2, 3, 8, 10, 17]. Therein, his ability for creative activity is poorly developed, which, in turn, negatively affects his ability to solve educational problems.

In this context, a special role is played by the mathematical content, which is characterized by a high degree of generalization, universality, and alternativeness of research and proof methods used. For example, one can mention the duality principle, which plays an important role in projective geometry. According to this, the projective space can be isomorphically mapped onto itself with the "transformation" of points in the plane, and vice versa. Obviously, the above regularity significantly helps to formulate statements of the mentioned theory, since any proof can automatically be applied to two different meaningful alternatives.

Accordingly, the targeted process of mathematics education creates the greatest opportunity for teaching generalized methods of mental activity to students, providing an opportunity for a multilateral analysis of the problem situations under consideration. It is convenient to organize such an analysis by changing the semantic status of the components of these situations. In this case, the initially key element of the task can be considered as an auxiliary element, and the initially auxiliary element as being a key one.

This approach appears in many psychological and pedagogical studies in various modifications. For example, the founder of Gestalt therapy, introduced thought experiments that contribute to the personal growth of a person ("feeling of opposing forces", "attention and concentration", etc.) [25]. The same procedures can be used in teaching mathematics. In particular, when solving geometrical problems, school students often consistently analyze the corresponding geometric configurations, highlighting the most promising one for further exploration, realizing the feeling of "background and figure" [8, 11, 22]. Having reconstructed the

problem situation, students get the opportunity to actualize the "feeling of opposing forces" [1,11, 22, 26, 27].

The role of the considered characteristic of thinking in the development of scientific ideas was emphasized by many famous scientists [1, 18]. In particular, it is noted that the concepts and facts under consideration should be not investigated in the "normal", correct, and usual form, but "in a critical situation, in fever and passion..." [28]. The authors also suggested paying special attention to characteristics that do not correspond to "the usual intuitive ideas, and allow us to look at well-known facts from a new angle, to see new and unexpected content therein" when conducting research [18]. Similar considerations in favor of the implementation of a "pathological" approach to the research subject are expressed by the famous mathematician in [1].

In a didactic context, updating the considered approach can be viewed as a kind of heuristic method used at the stage of problem solving [2, 3, 11, 22, 29]. In particular, this technique is the base of many non-standard math tasks. Mastering it, along with solving purely didactic problems, allows for the forming of a multilateral and "multidimensional" approach to the subject of study, "decentralizing" the viewpoint on the considered mathematical content and, thus, contributing to the formation of a sufficiently flexible and "variable" style of thinking of students.

From the foregoing, it follows that it is expedient to develop a special methodological structure that ensures the development of an alternative thinking style when teaching mathematics at school. The appropriate orientation should be a constant component of the educational process, one of the methods of cognition that allows for effectively organizing an educational heuristic search in the development of new subject content.

Based on the analysis of a number of studies in the context of the problem under consideration [2–4, 8, 11, 22, 25, 28, 30–33], and our own pedagogical experience, the following methodological provisions can be taken as the basis for such work:

- It is advisable to actualize and form an alternative thinking style for school students on the basis of special tasks implying an alternative consideration of a certain mathematical model and its components. These tasks should be interrelated with the general context of mathematical material, ensuring the effective achievement of the educational goals planned by the teacher.

- In accordance with active learning strategies, the implementation of the previous condition in the learning process presupposes an organization of adequate independent activity of students, during which certain connections and relations are established between the considered alternative positions, and a qualitatively new integrative knowledge arises. The highest developmental effect is present when the idea of applying some heuristic procedure comes directly from students in the form of conjecture and suppositions. George Pólya writes, "A conjecturing student associates himself therewith; his prestige and self-esteem depend on the outcome of the case to some extent, and he is eager to find out whether his guess turns out to be correct or not; he will be actively interested in his task and the work of class, he will not fall asleep, and will not be distracted by other tasks" [8].

- Actualization of various semantic positions, being complementary and mutually exclusive, presupposes interaction and cooperation of students' efforts. According to Rada Granovskaya and Yuliya Krizhanovskaya, "during joint activities, the initial attitudes of some participants on the inability to solve the problem are removed. Interstimulation of partners ... reduces self-criticism and prevents the displacement of original ideas into the subconscious" [3]. In particular, the so-called "effect of a teasing interlocutor" [33], which defends the position opposite to that of the author is very effective in the considered perspective. This approach makes it possible to impart a subjective-personal character to the educational exploratory process, which ensures the effective mastering of the corresponding heuristic techniques.

- The role of a teasing interlocutor can be effectively performed by a modern virtual educational environment, which allows for quickly moving and changing elements of a particular mathematical model when expedient. During the process it becomes possible to actualize alternative key elements for a relatively short period of time, and determine further prospects for their use when considering this specific task situation, or, to select the optimal alternatives [33].

The use of virtual educational and developmental environments can be implemented locally when organizing a heuristic search for a solution to a given specific problem, or it can become a means for full-fledged creative reflection on the meaningful links between its elements, due to their optimal dynamization and visualization. In the course of such work, it is given an initially boring (task) situation, after "having started to formulate and think it over" students see that "it acquires components, details, and new storylines, sometimes completely unexpected from the viewpoint of its initial perception" [33]. This approach is mentioned in a number of our publications [6, 30, 31, 34, 35].

4. Materials and Methods

The development of this research concept was carried out in line with the main trends in the system of mathematical education, in particular, problem-based and student-centered approaches, taking into account the specifics of the mathematical content [3, 4, 8, 9, 11, 22, 28, 34, 36–38]. The main instrument to implement these approaches is a system of special math tasks, leading to actualization of an alternative thinking style in school students regulated by a teacher. The study involved three clusters of materials.

Firstly, we have analyzed normative and program documents, educational standards of general secondary education, programs of mathematical disciplines, scientific and methodical works of leading domestic and foreign experts in the field of developmental psychology, didactics, theory and methods of mathematical education, materials on the history of mathematics, current textbooks, teaching aids, methodical recommendations, and educational software [1–3, 18, 21, 28, 33, 39]. The study of these materials indicates that the problem of the formation of alternative thinking in school students currently has not been singled out as a research subject for specialists in the field of mathematical education.

Secondly, we have conducted a longitudinal study of teaching mathematics in a number of educational institutions in Penza (Russia). The background competence of school students was revealed in terms of their readiness to manifest an alternative thinking style when solving exploratory math tasks. It was shown that about 80% of the school students surveyed in a number of Penza schools were not ready for consideration of variable alternatives to a given mathematical models, limiting themselves to analyzing the only possible alternative. Insufficient attention of teachers to the described component of educational and exploratory mathematical activity seems to be the main reason for such a problem.

Finally, the third cluster of materials comprised the results of an exploratory pedagogical experiment to determine the possibilities of introducing the developed mathematical content and make an assessment of its effectiveness in terms of creating the readiness of school students to apply an alternative thinking style when solving exploratory math tasks. The exploratory experiment was carried out in the course of a two-year mathematical training of school students at Lyceum No. 3 for gifted students of the Penza region, and a number of educational institutions in Penza. In total, the experiment included about 80 school students at the basic and senior levels of education.

The content of the materials being tested is partially presented in a number of our textbooks and articles [7, 11, 22, 30, 31].

The data collection technique was implemented using special educational and exploratory tasks. The results of their solving by school students were subjected to qualitative and quantitative analysis on the basis of a special quantitative scale at the background and productive stages. The following indicators were considered as the main characteristics of the alternative style of thinking: readiness to actualize unusual associative links; the ability to perceive the problem in an unusual way; readiness to build initially unobvious alternative constructions (an openness to an extraordinary combination of elements ("game-research"); the ability to alter actions; the ability to re-construct the acquired information in accordance with the structure of the problem domain [25]).

The data collection technique was implemented using special diagnostic approaches based on the use of two-level tasks. Each diagnostic task included two parts. In the first part, it was proposed to solve a task that involves finding a solution based on the actualization of an alternative task structure with a similar solution method. The second part included questions for which the students did not have clear guidelines. The purpose of posing these questions was to give students a certain "hint" about the possibility of developing the initial task. Attempts to find such answers testifies to the presence of a certain orientation of the students towards the search for alternatives. At the same time, students were required to set goals independently, (in particular, when they formulated tasks); focus on finding the most general way of action; effectively "translate" the information contained in the original problem into an alternative mathematical language; engage in full-fledged reflection on the revealed meaningful interrelationships.

The data collection technique was implemented using a special diagnostic intervention based on the use of two-level educational search tasks. Each diagnostic task included two parts. In the first part, it was proposed to solve a task that involves finding a solution based on the actualization of an alternative task structure with a similar solution method. The second part included questions for which the students did not have clear guidelines. The very purpose of posing these questions was to give schoolchildren a certain "hint" about the possibility of developing the initial task situation.

Already the presence of attempts to find such answers testifies to the presence of a certain orientation of the senior pupils towards the search for alternatives. At the same time, the ability to engage in relatively independent goal-setting was required (in particular, when drawing up tasks); focus on finding the most general approach; effective "translation" of the information contained in the original problem into an alternative mathematical language; full-fledged reflection on the revealed meaningful interrelationships (for example, when completing the definition of the task situation). The following are examples of such tasks for consideration.

Task 1:

- Let D be the point of the AC side in the ABC triangle; r, r1 and r2 the radiuses of the inscribed circles of the triangles ABC, ABD and BDC, accordingly. Prove that r < r1 + r2.
- Is it possible to extend the proved inequality to a larger number of inscribed circles? (You can, for example, divide the AC side into more parts, or take point D inside the ABC triangle).
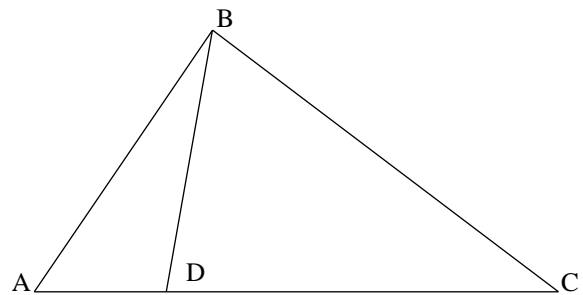
Figure 1: Illustration of the task

Task 2:

- Prove that the bisector of the exterior angle at the vertex of an isosceles triangle is parallel to the base.
- How many ways of proof can you suggest? (The number of proposed and successfully implemented methods is considered).

To establish an ordinal measurement scale, a preliminary assessment of possible answers was carried out and a weighted assessment in points was made. Completion of only the first part of the assignment was assessed by one point, a purposeful attempt to answer the question of the second assignment added one more point, while the performance of the second part of the assignment was assessed by two points. If, when completing the second part of the tasks, the student offered at least three answer options, then one more point was added to the total number of points. Thus, the student could receive a maximum of four or five points for one task. The total grade that could be obtained for completing the entire work, including four tasks, did not exceed 18 points. Three categories of students were identified. A student who received from 0 to 6 points corresponds to the initial level of the formation

of an alternative style of thinking; students who received from 6 to 12 points were assigned to the intermediate level, and from 13 to 18 - to the high level.

Comparison of the results of the tasks performed by the students of the experimental and control groups showed that the purposeful use of the methodology for devising meaningful alternatives in the study of mathematical material significantly influenced the increase in the readiness of students to successfully solve mathematical search problems.

5. Research Results

5.1. Levels of formation of an alternative thinking style

With regard to mathematics, the characteristics defined above can be actualized in solving math tasks, the solution of which involves implementation of a certain set of intellectual procedures. These procedures are reflected in the exploratory activity of school students, corresponding to the levels of its development [16]. However, their status varies at different levels (Table 1).

Table 1: Levels of Formation

Levels of formation of an alternative thinking style	Object of the analysis	Dominant way for actualizing an alternative style of thinking	Selection criteria for an alternative search	Nature of the intellectual procedures involved
Initial	A particular task	External (based on strict external requirement)	Desire to "function" strictly within the prescribed range of choice	Random chaotic or systematic search for alternatives
Medium	General task situation	Mixed (result-oriented)	The choice is determined by different model representations of the task situation	Selection of a technique by analogy with the solution of "related" tasks, accompanied by its experimental verification
High	System of related tasks	Internal (search-oriented)	Independent and reasonable variation of the conditions and requirements of the task	Selection of a heuristic procedure generated by purposeful modification of the components of the task domain

This scale of levels allows us to outline a strategy for the consistent formation of an alternative style of thinking, based on the use of systems of special tasks which are presented below.

5.2. Substantive and Methodological Tools

5.2.1. Enumeration of Alternatives

We now consider the main methodological features of the formation of an alternative thinking style of school students during their moving up the "ladder of levels".

At the initial level, alternativeness is directly related to the selection of possible task options, and its transformation. It is important that the expediency of such a choice should acquire a

subjective personal meaning for a school student. According to Zinchenko, it means that the subject has a special type of "awareness of ignorance of some knowledge", or "an emotional experience of ignorance of some knowledge", as stated by Konstantin Pigrov. Johann Amos Comenius qualifies such awareness as the source of the thirst for knowledge.

Let us consider an example of actualizing an alternative thinking style of students by enumerating alternatives presented in a particular task situation.

There are three criteria for congruent triangles with different combinations of angles and sides of a triangle studied in school course of planimetry. If we enumerate possible combinations, we can formulate some more statements (with two angles and a side opposite to one of the angles; with two sides and an angle opposite to one of the sides; and with three angles).
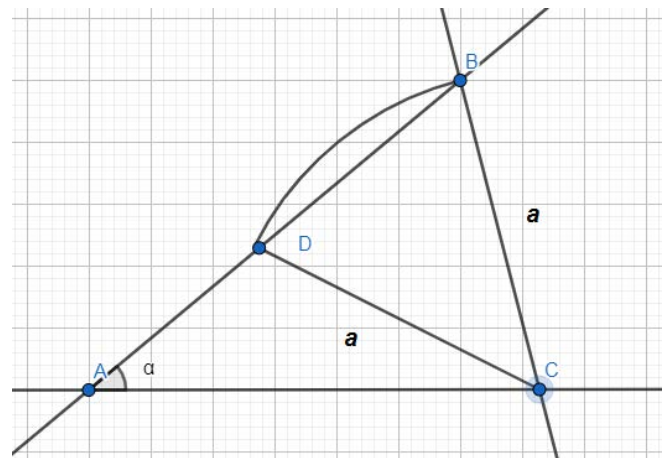
Figure 2: Illustration of the task

The first criterion formulated by students can serve as a starting point for studying the sum of angles of a triangle.

In the second case, we consider the fourth criterion for congruent triangles with two possible solutions for constructing a triangle from the corresponding elements based thereon (Figure 2). The unambiguity of the solution can be provided by the additional condition of uniformity of the angles at the vertex B of one triangle, and the corresponding vertex D of another triangle.
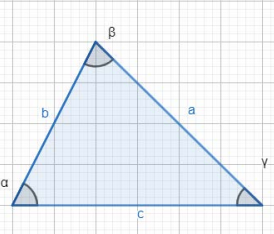
Finally, the obvious inequity of a criterion for congruent triangles by three angles naturally raises the question of congruence of triangles.

The described alternatives matrix can be clearly presented for school students on the basis of a directive that includes all the selected cases (Table 2).

This example shows the possibility of actualizing an alternative thinking style of a student by enumerating various specific models of a common task, aimed at identifying all combinations with a minimum number of triangle elements, unambiguously determining its construction when considering standard mathematical material.

To conduct further exploratory activity, we should move away from traditional stereotypes and apply well-known heuristic procedures [2, 3, 8, 9, 11, 22, 26].

Table 2: An Example of a Directive for School Students

Criteria for congruent triangles	Constructing triangles given three elements	Solving triangles
1. Two triangles are congruent if two corresponding angles and the side in between are equal (a – side; α and γ – angles).	 (Construction is possible and unambiguous at $\gamma + \beta < 180^0$)	Angle α is found according to triangle angle sum theorem; side c is calculated by the law of sines; side b is found according to the law of cosines.

For example, we consider the replacement of structural components of the task domain. With regard to school mathematical content, the mentioned procedure can be actualized with the help of a specific teaching technique, namely, "he (she, it) does not know what he (she, it) really is."

We illustrate this statement with the following artificially constructed example. Consider the usual quadratic equation $x^2 - 13x + 36 = 0$. This equation obviously involves numbers and letters. Now we make an assumption that "numbers do not know that they are numbers, and letters do not know that they are letters." In other words, "the number 6 does not know that it is a known number, the letter x does not know that it is an unknown number." Let us find the number 6:

$$6^2 - \left(\frac{13x}{6}\right) \times 6 + x^2 = 0$$

$$D = \frac{169x^2}{36} - 4x = \frac{25x^2}{36}$$

$$6 = \frac{\frac{13x}{6} - \frac{5x}{6}}{2} = \frac{2x}{3} => 2x = 18 \quad x = 9$$

$$6 = \frac{\frac{13x}{6} + \frac{5x}{6}}{2} = \frac{3x}{2} => 3x = 12 \quad x = 4$$

This example illustrates the potential for an alternative search for thought, being opposite to the standard algorithm. According to Zinchenko, it is "at the junction of opposite processes of value comprehension and signification of meanings that new images are born, conveying certain meanings and making the meaning visible" [1].

The specificity of this approach is constant correlation of the original design for task solution with its alternatives forming the basis of this solution in a new semantic paradigm. The logic of an exploratory process using the adopted approach is shown in Figure 3.
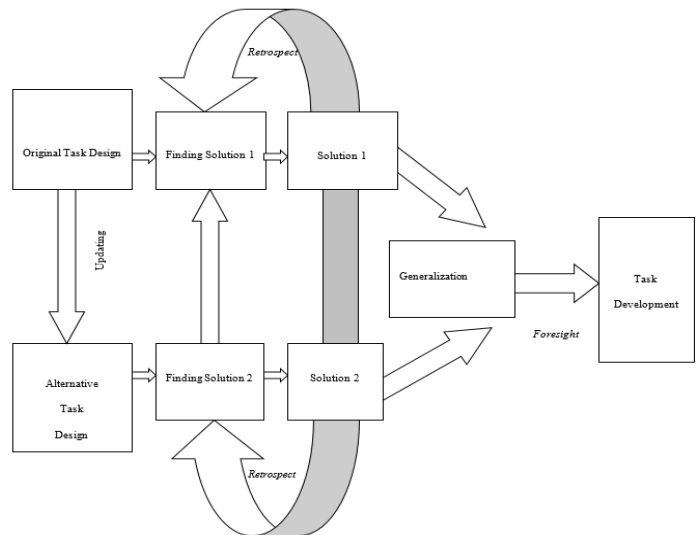
Figure 3: Schematic for an exploratory process

This schematic demonstrates updating of possible alternative problem-solving in the process of sequential reformulation of the task situation, ensuring effectiveness of the exploratory process being implemented.

Here are some examples of such updating when solving non-standard tasks:

"The number does not know that it is a letter"

Calculate: $\frac{254 \times 399 - 145}{254 + 399 \times 253}$

Solution: Let a=254, b=399. Then $\frac{ab - (b - a)}{a + b(a - 1)} = 1$

(Soros Olympiad, 1994/95, Round 1, Grade 10)

Calculate: $\left(\frac{1 - 2(17.3)^3}{1 + (17.3)^3}\right)^3 + \left(\frac{17.3 \cdot (2 - (17.3)^3}{1 + (17.3)^3}\right)^3 + (17.3)^3$

Solution: Let$(17.3)^3$=a. Then $\frac{(1 - 2a)^3 + a(2 - a)^3}{(1 + a)^3} + a = \cdots = 1$

(Republican Mathematics Olympiad, Kiev, Ukraine)

Which number is greater?

$\frac{10^{1967} + 1}{10^{1968} + 1}$ or $\frac{10^{1968} + 1}{10^{1969} + 1}$;

Solution: $\frac{a + 1}{10a + 1}$ v $\frac{10a + 1}{100a + 1}$;

$100a^2 + 101a + 1$ v $100a^2 + 20a + 1$

$101 > 20$

Prove that $1993 \times 1995^3 - 1994 \times 1992^3$ is an integer cube.

Solution: $1994 = a$;

$(a-1) \times (a + 1)^3 - a(a - 2)^3 = \cdots = (2a - 1)^3$

"Algebraic or numerical expressions do not know that they are letters"

(Intellectual Marathon, 1992)

Find the sum: $\cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{3 + \cfrac{1}{4 + \cfrac{1}{5 + \cdots + \frac{1}{n}}}}}} + \cfrac{1}{2 + \cfrac{1}{3 + \cfrac{1}{4 + \cfrac{1}{5 + \cdots + \frac{1}{n}}}}}$.

Solution: $\frac{1}{1 + \frac{1}{A - 1}} + \frac{1}{A} = \frac{A - 1}{A} + \frac{1}{A} = 1$.

Solve the equation $\frac{x^2 + x + 2}{3x^2 + 5x - 14} = \frac{x^2 + x + 6}{3x^2 + 5x - 10}$.

Solution: $\frac{a}{b} = \frac{a + 4}{b + 4} => \frac{a}{b} = 1$, $3x^2 + 5x - 14 = x^2 + x + 2 \ldots$

Answer: $x=2$, $x= -4$.

It is clear that the described principle of transition to meaningful alternatives can also be used when considering geometric configurations: "The angle does not know that it is exterior"; "The circle does not know that it is escribed," etc.

Let us consider an example of the exploratory process to analyze one of configurations.

Example. Prove that the sum of angles DBE and CAM does not depend on the choice of the line l (Figure 4).
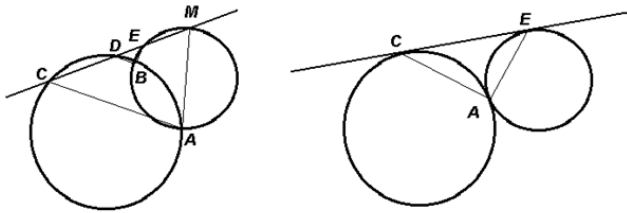
Figure 4: Illustration of the task

It is proposed to consider the line *CM* as a dynamic image that can move relative to circles, and turn into a tangent to both of them ("*a secant does not know that it is a tangent*"). Similarly, it's possible to move from an intersection to an external contact of circles ("*intersecting circles do not know that they are just tangent to each other*").

We obtain a chain of alternative configurations, which are degenerate cases of the initial one (Figure 3). The latter naturally leads to a derivative task situation.

Example. Find the angle CAE, where point A is the point of external contact of two given circles, and CE is their common tangent.

Solving the task, it is possible to assume that the angle CAE, considered as the sum of two angles, obtained at the intersection of the line *CE* and common tangent to two circles, is equal to 90°. As for the original geometric configuration, we consider the secant lines *CM* and *AB* as analogs of common tangents in a degenerate alternative construction. As in the degenerate case, the solution consists of expressing the angles forming the angle *CAM*, via the angles of the triangle *BDE*.

$\angle CAB = 1/2 \cup CDB = \frac{1}{2}(360° - \cup CAB) = \frac{1}{2}(360° - 2\angle CDB) = 180° - \angle CDB = \angle BDE$.

Likewise, $\angle BAM = \angle BED$. Then, $\angle CAM = \angle CAB + \angle BAM = \angle BDE + \angle BED = 180° - \angle DBE$.

Finally, we get: $\angle CAM + \angle DBE = 180°$.

Thus, we have shown that by highlighting the key link (determining the successful search) based on a meaningful analysis of the alternative configuration, one can successfully update the relevant considerations when examining the original one.

5.2.3. Structuring the Task

Since the use of an alternative style of reasoning is regarded by the student mainly as a local intellectual act in understanding the task situation at previous levels of mathematical development, then it is associated with the possibility of free and informed choice from a combination of implicitly presented meaningful alternatives at a higher level. It is quite obvious that in this case the heuristic potential for choosing alternatives increases significantly. Thenceforward, the larger set of heuristic procedures the student actualizes, and the more systemic they are, the more efficient a person's intellectual development is.

Successful procedure for an informed alternative choice becomes one of the main instruments for maintaining an active exploratory activity in schoolchildren. One of the most effective methods for creating an opportunity for such a choice is a step-by-step transformation of components of a task situation as applied to mathematics education, leading to the construction of a sequence of new task structures, somehow related to the original one, which is accompanied by their comparative analysis. The advantage of such work stems from the fact that the initial intellectual activity of a student does not stop when the solution of the original problem is completed. Moreover, it receives an additional heuristic impulse during its transfer to new substantive content, ensuring full realization of his individual trajectory of development.

Transformation of task structures can be potentially implemented in many alternative formulations, determined by the choice of the involved techniques, or their expedient combinations. As a result of specially organized joint work of subjects of the educational process, one can obtain a whole cycle of alternative derivative tasks based on some initial mindset. The main methodological problem in this case is to ensure effective anticipation of the most promising and available areas of exploratory activity for implementation at the particular stage of training.

We consider one of such complexes, obtained by successive generalization of the following original situation.

Original Task. Construct a circle that is tangent to the lines *m* and *n* (*m*//*n*), and to the circle (*O*, *r*) lying between these lines.
Having changed the location of this circle ("*the circle lying between two given lines does not know that in fact it does not lie between them*"), we get the task:
Task 2. Construct a circle that is tangent to two different parallel lines, and a given circle.
If in Task 2 we consider that the lines coincide ("*the lines do not know that they coincide*"), then we get the following task:
Task 3. Construct a circle that is tangent to a given line at a given point *M* thereon, and a given circle that is tangent to this line.
Having ignored the condition of tangency of a given circle and line ("*the circle that is tangent to the line does not know that it actually is not tangent to it*"), we get the following task:
Task 4. Construct a circle that is tangent to a given line at a given point *M* thereon, and a given circle.
Having ignored the condition of tangency at a given point ("*the circle does not know at what point it is tangent to the line*"), we get the following task:
Task 5. Construct a circle that is tangent to a given line and a given circle.
Having replaced a line with a circle in Task 5 ("the circle does not know that it is the line"), we get the following task:
Task 6. Construct a circle that is tangent to two given circles, and to one of them at a given point.
Having replaced a circle with lines in Task 6 ("the circle does not know that it is two lines"), we get the following task:

Task 7. Construct a circle that is tangent to the lines *m* and *n* (*m∥n*), and a line *c* intersecting the lines *m* and *n* at points *K* and *L*, *K* ∈ *m*, *L* ∈ n.

Using alternative task structures in the educational and exploratory work, ideologically related to the original task situation, a teacher gets the opportunity to stimulate an exploratory activity of school students, while enriching and improving their intellectual and personal potential.

It is important to note that similar procedures can be visualized quite effectively using a modern software environment. By selecting, swapping, and moving the initially actualized and background elements of considered mathematical structures with a manipulator, one can obtain alternative constructions that allow "seeing" a potentially successful area of an exploratory activity [6, 31, 32, 35].

6. Approbation

A unified, generally recognized methodological concept of the formation of an alternative style of thinking by students in mathematics lessons presumably is not provided in real school practice. The test of this hypothesis was carried out within the framework of the first stage of the experiment, the main task of which was: to study the nature of the spontaneous influence of the process of "natural" development of the content of school mathematics courses and the personality traits of the teacher on the effective actualization and formation of an alternative style of thinking in schoolchildren.

To confirm this fact, the results of the tests performed by 80 students of grades 3-4 of a number of schools in Penza (Russia) were analyzed. Control work included tasks that imply elements of the study of task structures and the feasibility of enumerating alternative options for the search process. Evaluation of their implementation in points took into account the characteristics of the levels of development of the alternative style of thinking presented above. The data obtained indicate that more than half of the surveyed schoolchildren can be assigned to the initial level of such development (52%), 39% to the average level, and only 9% to the high level. These data indicate that teaching mathematics in itself in accordance with the current curriculum does not yet provide a high level of development of the alternative style of thinking of schoolchildren.

Approbation of the proposed methodological solutions was carried out in two segments, differing from each other in the number of students involved and the nature of the experimental work.

1. A "free" experiment was carried out in several schools, gymnasiums and lyceums in Penza (№№ 3, 6, 11, 12, 42, 47), covering about 200 students. It was preceded by a lot of preparatory work. In the course of its implementation, we read a series of lectures for teachers in the field of advanced training, as well as mathematicians of the final year of Penza State University (as part of elective courses), in which the listeners were familiarized with the main theoretical provisions of the developed concept and a number of teaching materials.

Additional group and individual consultations were held with teachers who expressed their willingness to take part in the experiment. A special role in the implementation of this strategy was given to the initial and current diagnostics of relatively stable indicators of the development of an alternative style of thinking in

order to clarify the individual educational trajectory of a particular student.

The experimental work was carried out for two years in accordance with the above recommendations, taking into account the achieved level of development of objective motivation. Throughout this period, regular observations of the course of the educational process were carried out, open lessons were organized, laboratory experiments and conversations with students were conducted.

Without purporting to be an exhaustive categorical conclusion, we consider it possible to note certain positive shifts in the development of the alternative style of schoolchildren's thinking, which were found in a significant number of the schoolchildren who participated in the experiment. These schoolchildren have significantly increased their interest in solving tasks of a creative and research nature. Moreover, when performing such tasks, they in most cases independently transformed their conditions on the basis of implementing alternative search strategies possible in a given situation. In the course of studying the theory, the students involved in the experiment tried to "get ahead" of the teacher's plan, "to see the goal from afar." At the same time, a certain independence of judgment became more characteristic of them, expressed in a more critical and reflexive attitude to the decision made, the desire to supplement and improve it in relation to both the analyzed situation and a wider class of mathematical concepts.

2. The "local" experiment was carried out in one of the lyceums of the city of Penza, where one of the authors of the article worked for 10 years as a methodologist and teacher of mathematics. Purposeful experimental work in line with the studied issues was carried out at the Lyceum for three years from 2018 to 2020. At the beginning of this work, the teachers of mathematics of the experimental classes were given explanations of its goals and content, discussed the possibilities of implementation, and proposed educational materials that included systems of tasks of a special type corresponding to different levels of formation of an alternative style of thinking. Ongoing consultations were offered to teachers in the course of attending lessons, analyzing the results of control and cross-sectional works.

To partially confirm the effectiveness of the proposed tools, the results of 52 students in grades 3-4 were analyzed separately, some of whom worked according to the traditional method (27 students), and the other part according to the methodology proposed by the authors (25 students). At the first stage of the study, a number of introductory tests were carried out, consisting of tasks that imply elements of the study of task structures and the feasibility of enumerating alternative options for the search process. The distribution according to the levels of formation of alternative thinking of schoolchildren was carried out in accordance with the criteria reflected in the section "Materials and Methods".

Table 3: The results before the experiment

Levels of formation of ATS	Before the experiment			
	Experimental group		Control group	
	Quantity	%	Quantity	%
Initial	12	48	12	44,4
Medium	8	32	10	37
High	5	20	5	18,6

Table 3 and Figure 5 show that the level of formation of the alternative style of thinking in both groups is rather low. About half of all subjects are at the elementary level, and only about 20% of students have reached a high level.
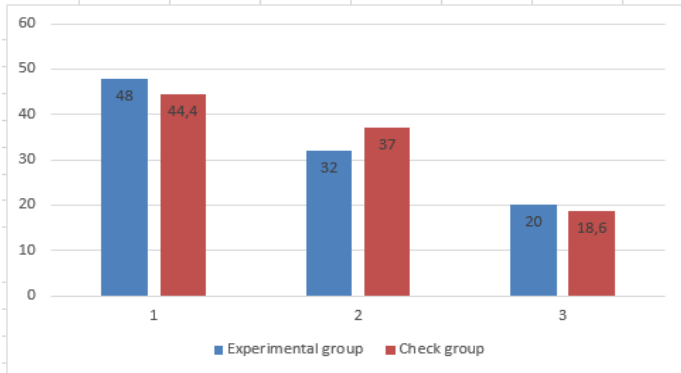
Figure 5: The results before the experiment

The data obtained were statistically processed using Pearson's chi square test (goodness of fit). The processing results are shown in Table 4.

Table 4: The results before the experiment

Levels off ATS	Q$_{emp}$		Q$_1$-Q$_2$	(Q$_1$-Q$_2$)^2	Q$_1$+Q$_2$	$\frac{(Q_1-Q_2)^2}{Q_1+Q_2}$
	Experimental group	Control group				
Initial	12	12	0	0	24	0,00
Medium	8	10	-2	4	18	0,22
High	5	5	0	0	10	0,00
Σ	25	27				0,22

The data in the table indicate the absence of significant differences in the level of development of the alternative style of thinking among students in both groups (empirical value of Pearson's criterion $X^2 = 0,22$). In other words, the nature of the distribution of schoolchildren according to the levels of development of this style was initially approximately the same.

Experimental work within the framework of this study was carried out for two years. In the course of teaching, the pupils of the experimental classes were offered the tasks described above, corresponding to their level of formation of an alternative style of thinking. For the initial level, emphasis was placed on an alternative enumeration of mathematical structures, to which tasks were added to replace the structural components of these structures. At the second and third levels, schoolchildren purposefully worked on a relatively independent compilation and analysis of task systems in accordance with the concept presented in the article.

As a result of the experimental work control measures were again carried out.

Comparison of the dynamics of changes in motivational characteristics in the selected groups of students (Table 5) showed

that over the same time interval in the experimental classes, about 40% of students moved to a higher level of development of an alternative style of thinking, while the number and distribution of pupils at each of the levels in the second group changed insignificantly. A few students in this group even showed a slight decrease in interest in mathematics, explained by their future professional choice (not related to mathematical activity).

Table 5: The results after the experiment

Levels of formation of ATS	After the experiment			
	Experimental group		Control group	
	Quantity	%	Quantity	%
Initial	3	12	10	37
Medium	9	36	13	48,2
High	13	52	4	14,8

The data obtained were again subjected to statistical processing using Pearson's chi square test (goodness of fit). The processed results are shown on Figure 6 and in Table 5.
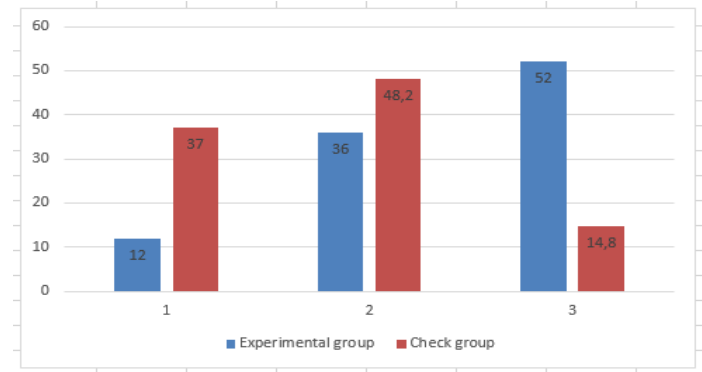
Figure 6 The results after the experiment

The results of statistical processing of the data obtained indicate the appearance of significant differences in the analyzed indicator between the schoolchildren of the control and experimental groups (Pearson's criterion $X^2 = 9,26$)

Table 6: The results after the experiment

Levels off ATS	Qemp		Q$_1$-Q$_2$	(Q$_1$-Q$_2$)^2	Q$_1$+Q$_2$	$\frac{(Q_1-Q_2)^2}{Q_1+Q_2}$
	Experimental group	Control group				
Initial	3	10	-7	49	13	3,77
Medium	9	13	-4	16	22	0,73
High	13	4	9	81	17	4,76
Σ	25	27				9,26

In other words, we found a significant difference between the distribution of characteristics of the formation of the alternative style of thinking of schoolchildren in the studied groups of schoolchildren. This difference indicates that the developed methodological solutions are more effective than the traditional ones in terms of the formation of the specified style.

In general, it can be stated that, as evidenced by the results of approbation, the described level-by-level strategy for the formation of an alternative style of thinking of schoolchildren in the process of teaching mathematics proved to be quite effective both in terms of the gradual formation of students' readiness to overcome their initially one-sided attitudes, and in terms of the development of their cognitive activity generally.

7. Conclusions and Suggestions

In the course of the study, the author's concept of the level-by-level formation of an alternative style of thinking of schoolchildren in mathematics lessons was developed. As the main means of such formation, it is advisable to consider special tasks that provide for the presence of alternative options for analyzing the elements of its subject area. Within the framework of this concept, the following scientific and methodological results were obtained:

- The important role of the alternative style of thinking is revealed, which underlies the realization of human intellectual activity and which ensures effective interaction with the environment and the ability to tolerantly respond to its diversity.
- It is shown that in the framework of mass school education, insufficient attention is paid to the formation of such a style. As a result, as many publications testify, many schoolchildren master the methods and techniques of educational and search activity as rigid algorithmic structures that hardly allow variation of semantic positions with respect to the concepts under consideration and the relations between them.
- A mathematics course has great potential in terms of the formation of an alternative style of thinking, as it is characterized by the presence of a variety of problem material, the mastering of which is possible from various positions and using various methods. For the full realization of this potential, an adequate pedagogical format is required, within which an alternative style of thinking can be actualized. As such a format, it is advisable to consider the joint activity of a teacher and students in solving special tasks that provide for the availability of alternative options for analyzing the elements of their subject area.
- When planning work on the development of an alternative style of thinking for schoolchildren in the process of teaching mathematics, it is advisable to single out a number of levels of formation of this style, the multifaceted characteristics of which are presented in the text of the article. This characteristic formed the basis for the development of diagnostic tools that ensure the assignment of a student to the appropriate level.
- The content of the work revealed the transition of schoolchildren up the "ladder of levels". At the initial stage, such work is directly related to a basic enumeration of possible options for the state of a particular task structure and its possible transformation. At the next stage, a purposeful search for opportunities to connect the problem situation under consideration with a more familiar alternative problem structure by reformulating its condition is carried out, as well as identifying how promising this structure is in terms of choosing a possible search direction. Finally, at the third stage of the work, the alternative style of thinking is actualized as a result of a conscious free choice from a number of implicitly presented meaningful alternatives. Such a choice can be provided on the basis of a step-by-step transformation of the components of the task situation, leading to the construction of a sequence of new alternative task structures, which is accompanied by their comparative analysis.
- On the basis of the selected theoretical positions, teaching materials were developed, which were successfully tested in a number of schools in Penza (Russia) over a number of years. As evidenced by the results of approbation, subjected to statistical processing, the level-by-level strategy, as described by the author, for the formation of an alternative style of thinking for schoolchildren in the process of learning mathematics proved to be, on the whole, quite effective both in terms of forming students' readiness to overcome their initially one-sided attitudes, and in terms of developing their cognitive activity on the whole.

At the next stage of the research, we plan to organize a wider pedagogical experiment to determine the optimal means of implementation of the proposed recommendations into a meaningful compilation of various programs in mathematics for different countries and regions.

References

[1] M. Uzunoglu, M.S. Alam, "Dynamic modeling, design and simulation of a P 1. V.P. Zinchenko, Psychological Foundations of Pedagogy (Psychological and Pedagogical Bases for the Construction of Elkonin-Davydov System of Developmental Education), Moscow, Gardariki, 2002.

[2] E. de Bono, Lateral Thinking: A Textbook of Creativity, London, Penguin Books, 1977.

[3] R.M. Granovskaya, Yu.S. Krizhanovskaya, Creativity and Overcoming Stereotypes, St. Petersburg, OMS Publ., 1994.

[4] M.A. Kholodnaya, Cognitive Psychology. Cognitive Styles, Moscow, Yurait Publ., 2018.

[5] V.G. Thomas, B.A. Parsons, "Culturally responsive evaluation meets systems-oriented evaluation," American Journal of Evaluation, **38**(1), 7–28, 2016, doi: 10.1177/1098214016644069.

[6] N.N. Khramova, M.M. Rodionov, "The development of thinking variability of schoolstudents in mathematics using 1C: Math designer," Computer Science and Education, **7**(256), 15–21, 2014.

[7] W. Lake, M. Wallin, G. Woolcott, W. Boyd, A. Foster, C. Markopoulos, W. Boyd, "Applying an alternative mathematics pedagogy for students with weak mathematics: meta-analysis of alternative pedagogies," International Journal of Mathematical Education in Science and Technology, **48**(2), 215–228, 2017, doi:10.1080/0020739X.2016.1245876.

[8] D. Pólya, Mathematical Discovery: On Understanding, Learning and Teaching Problem Solving, Moscow, Nauka, 1976.

[9] M. Rodionov, Z. Dedovets, "The development of students' intellectual tolerance in the process of teaching mathematics at secondary level," London of Research in Humanities and Social Sciences, **18**(2), 1–11, 2018, https://journalspress.com/LJRHSS_Volume18/397_The-Development-of-Students-Intellectual-Tolerance-in-the-Process-of-Teaching-Mathematics-at-Secondary-Level.pdf.

[10] M. Rodionov, Z. Dedovets, "The forming of students' intellectual tolerance: theoretical study and implementation," International Journal for Cross-Disciplinary Subjects in Education, **4**(1), 775–781, 2014, doi: 10.20533 / ijcdse.2042.6364.2014.0259.

[11] M.A. Rodionov, E.V. Marina, Formation of Variable Thinking of School Students in Solving Construction Problems: A Manual, Penza, PSPU Publ., 2006.

[12] A.G. Stepanov, "The style of thinking in the formation system mythologemes paintings of socio-historical reality," Bulletin of Chelyabinsk State University, **8**(430), 29–34, 2019, doi: 10.24411/1994-2796-2019-10805.

[13] M.A. Gadzhimuradov, Z.D. Gadzhieva, "Features of mathematical thinking

and its development when learning geometry," The World of Science, Culture and Education, **2**(69), 189–191, 2018.

[14] K.G. Kozhabaev, R.S. Gabdullin, "About pupils thinking development in the process of mathematics training," Modern High Technologies, **5**(3), 554–559, 2016, http://top-technologies.ru/ru/article/view?id=35952.

[15] E.Kh. Galyamova, "Methodological peculiarities of the process of geometry teaching considering the cognitive styles," Prepodavatel XXI Vek, 1(1), 153–158, 2019.

[16] S.N. Lipatova, O.A. Milinis, "Psychological and pedagogical support of the development of students's giftedness in the system of additional education: innovative model," Otechestvennaya i Zarubezhnaya Pedagogika, **2**(49), 200–210, 2018.

[17] C. Plumpton, Generalization and structure. In Process of Learning Mathematics, L.R. Chapman, Ed., Oxford, Pergamon Press, 87–121, 1972.

[18] H. Poincaré, The Foundations of Science (Science and Hypothesis. The Value of Science. Science and Method), Moscow, Nauka, 1990.

[19] T.L. Killpack, L.C. Melón, "Toward inclusive STEM classrooms: What personal role do faculty play?" CBE Life Sciences Education, **15**(3), 2016, doi:10.1187/cbe.16-01-0020.

[20] E.Yu Ilaltdinova, V.V. Kisova, "Teaching talent and teaching intellect: theoretical and applied approaches design," Vestnik of Minin University, **4**(25), 2018, doi:10.26795/2307-1281-2018-6-4-9.

[21]. I. Newton, Mathematical Works. Transl. by D.D. Mordecai-Boltovsky, Moscow, Leningrad, ONTI Publ., 1937.

[22] M.A. Rodionov, Formation of Search Motivation in the Process of Teaching Mathematics: A Manual for Students and Teachers, Penza, PSPU Publ., 2001.

[23] M.K. Watfa, D. Audi, "Innovative virtual and collaborative teaching methodologies," Behaviour & Information Technology, **36**(7), 663–673, 2017, doi: 10.1080/0144929x.2016.1275806.

[24] M.M. Thomson, D. DiFrancesca, S. Carrier, C. Lee, "Teaching efficacy: exploring relationships between mathematics and science self-efficacy beliefs, PCK and domain knowledge among preservice teachers from the United States," Teacher Development, **21**(1), 1–20, 2017, doi:10.1080/13664530.2016.1204355.

[25] F. Perls, R. Hefferline, P. Goodman, Experiences in the Psychology of Self-Knowledge: A Workshop on Gestalt Therapy, Moscow, Gil-Estel Publ., 1993.

[26] M. Rodionov, Z. Dedovets, E. Pavlova, N. Sharapova, I. Akimova, "Design and implementation of adaptive technology for teaching mathematics to school students based on integrated diagnostic approach to subject preparation and competence development," Revista Amazonia Investiga, **9**(26), 458–472, 2020, doi:10.34069/AI/2020.26.02.53.

[27] M.A. Rodionov, E.V. Marina, Developmental Potential of Mathematical Tasks, and Possibilities for its Actualization in the Educational Process: A Manual, Penza, PSPU Publ., 2010.

[28] I. Lakatos, Proofs and Refutations: The Logic of Mathematical Discovery, Moscow, Nauka, 1967.

[29] J. Engelbrecht, A. Harding, "Interventions to improve teaching and learning in first year mathematics courses," International Journal of Mathematical Education in Science and Technology, **46**(7), 1046–1060, 2015, doi:10.1080/0020739X.2015.1070441.

[30] T.A. Chernetskaya, M.A. Rodionov, "Interactive creative environments as a means for forming elements of mathematical research activity in schoolstudents," Computer Science and Education, 3, 36–41, 2014.

[31] V. Mkrttchian, G. Vishnevskaya, M. Rodionov, Avatar-Based Learning and Teaching in Modern Educational Environments: Emerging Research and Opportunities, Hershey, PA, IGI Global, Information Science Reference, 2018.

[32] P.C. Abrami, R.M. Bernard, E. Borokhovski, D.I. Waddington, C.A. Wade, T. Persson, "Strategies for teaching students to think critically: a meta-analysis," Review of Educational Research, **85**(2), 275–314, 2015, doi:10.3102/0034654314551063.

[33] H. Heckhausen, Motivation and Activity, Vol. 1, Moscow, Pedagogica, 1986.

[34] M.A. Rodionov, I.V. Akimova, T.A. Chernetskaya, "Creation of the constructive creative environment on the basis of realization of interactive dynamic models," In the World of Scientific Discoveries, **9**(57), 21–34, 2014, doi:10.12731/wsd-2014-9-2.

[35] M. Rodionov, Z. Dedovets, E. Pavlova, N. Sharapova, I. Akimova, "Design and implementation of adaptive technology for teaching mathematics to school students based on integrated diagnostic approach to subject preparation and competence development," Revista Amazonia Investiga, 9(26), 458–472, 2020, doi:10.34069/AI/2020.26.02.53.

[36] V.A. Mazilov, Yu.N. Slepko, "Pedagogical giftedness as a key prerequisite for efficient modern educational system," Integration of Education, **23**(1), 37–49, 2019, doi:10.15507/1991-9468.094.023.201901.037-049.

[37] S. Abramovich, A.Z. Grinshpan, D.L. Milligan, "Teaching mathematics through concept motivation and action learning," Education Research International, 7, 1–13, 2019, doi:10.1155/2019/3745406.

[38] G.T. Klinkov, M.A. Rodionov, O.E. Kozlova, E.V. Veszetiu, E.V. Vovk, "Implementation of discussion technologies in a personality-centered professional education," Revista Amazonia Investiga, **9**(26), 82–87, 2020, doi: 10.34069/AI/2020.26.02.9.

[39] K.W. Remijan, "Project-based learning and design-focused projects to motivate secondary mathematics students," Interdisciplinary Journal of Problem-Based Learning, **11**(1), 2017, doi:10.7771/1541-5015.1520.

VoIP Codec Performance Evaluation on GRE with IPsec over IPv4 and IPv6

Oluwaseun Ayokanmi Alausa, Samson Afolabi Arekete[*], Mba Obasi Odim, Abosede Oyenike Oguntunde, Adewale Opeoluwa Ogunde

Department of Computer Science, Redeemer's University, Nigeria

A B S T R A C T

Scientists succeeded in implementing conventional public switch telephone network (PSTN) into internet protocol by launching H.323 IP telephony. The irrelevant and unknown captions in H.323, computer scientists have replaced H.323 by Session Initiation Protocol (SIP) for Voice-over-IP (VoIP). However, the security of voice communication over IP is still a major concern. Besides, security and performance contradict features. VoIP exhibits a quality-of-service requirements that are sensitive to time. Example of such QoS requirements are delay, jitter, and packet loss. Integrating Internet Protocol Security (IPsec) with Generic Routing Encapsulation (GRE) encrypts and authenticate packets from the sender to receiver, but that raises the question of performance as VoIP is time sensitive. Consequently, three codecs were evaluated to determine the efficiency of each on GRE and IPsec implementation on Internet Protocol version 4 and Internet Protocol version 6 (IPv4 and IPv6), respectively. The topology design and device configuration in this study adopted Graphic Network Simulator 3 (GNS3) and Distributed Internet Traffic Generator (D-ITG) to generate VoIP traffic. The evaluation revealed that the G.723.1 codec achieved better results on IPv4 and IPv6 over GRE with IPsec than other codecs used in the experiment. Furthermore, the codec of choice is a major factor in IPsec VoIP deployment, as also revealed in this study.

1. Introduction

There is a budding interest among network providers in deploying a converged IP network that is more cost-effective and more comfortable to manage [1]. Voice communication over IP leverages the packet-switched network architecture, allowing both voice and data packet to negotiate the IP infrastructure simultaneously. The IP architecture supports the carriage of any traffic (voice, video, and text) using best effort carriage devices. As reasonably priced as the IP network is, it poses some threat and challenges for Voice communication [2]. Tunnelling protocol such as Generic Routing Encapsulation Protocol provides packet-level encapsulation only with support for IPv4 and IPv6. Network packets are more susceptible to attack and sniffing using only GRE. Hence, the need for Internet Protocol Security (IPsec); a standard responsible for packet authentication and encryption at layer 3 with support for IPv4 and IPv6 [3, 4].

The use of IPsec is ubiquitous in Virtual Private Networks today for a two-way authentication between hosts at the start of a session and crypto keys for negotiation during an established

session [5, 6]. IPsec can secure packet flows between host-to-host, network gateways, and network to host communication [4]. Implementing VoIP over such a reliable and secure protocol ensures a high level of security and privacy.

GRE is a packet encapsulation protocol that supports multicast traffic [6]. GRE provides support for both IP protocol and non-IP protocols. GRE requires IPsec to provide reliable security attributes like encryption and authentication as it is not a secure protocol on its own [7].

Encryption converts data to a cypher-text format that makes data unreadable to sniffers. The appropriate algorithm and key are required to access the cypher-text content. Examples of cryptographic algorithms include Triple Encryption Standard (3DES) and Advanced Encryption Standard (AES) [8].

However, VoIP has time-sensitive traffic, which exhibits a bounded quality of service requirement, such as delay, jitter, and packet loss [9]. It is crucial to verify that adding an extra layer of security on the VoIP packet does not degrade the QoS requirements.

[*]Corresponding Author: Samson Afolabi Arekete, Email: areketes@run.edu.ng

Related works [5, 10–12] reviewed have investigated GRE and IPsec over IPv4 with little concentration on IPv6, thus a Motivation to understand better the effect of IPsec and GRE on VoIP codec over IPv4 and IPv6.

2. Objective of the study

The objective of this study includes:

1. Evaluate the performance of three VoIP codec on GRE with IPsec considering their effect on delay, jitter, and packet loss.
2. Compare the performance of each codec in (1) above over IPv4 and IPv6.

3. Significance of the study

This study's findings helped verify the appropriate codec suitable for GRE with IPsec network, staying within the limits of ITU-T recommended VoIP quality standard [13]. Through its experiment, the study has established that the choice of voice codec plays a significant role in VoIP IPsec deployment. With the right voice codec, IPsec does not in any way degrade the quality of voice communication.

4. IP Protocols and Voice Codecs

Two IP protocols are discussed in this section, namely IP version 4 and IP version 6. Three codecs developed by the ITU's Telecommunication unit for audio compression and decompression were highlighted: G.711, G.723.1, and G.729.3 codecs.

IPv4 – Internet Protocol version 4

Currently, the Internet Protocol version 4 is mainly used to communicate over the Internet, although deployment of a successor protocol, IPv6, is ongoing. Usually, an IP address holds two different data types, such as the host address and the network address. IPv4 address was structured based on 32-bit values and typically expressed in dotted decimal notation with four octets separated by decimals, for example, 192.168.120.80. IPv4 addresses were divided into five different classes; however, classes A, B and C are generally used. Class A provides the highest amount of IP addresses, while class B provides less than class A and class C offers the least amount of IP addresses [14,15]

IPv6 – Internet Protocol version 6

IPv4 has been used in the Internet world for over two decades. Unfortunately, this is approaching the limit of its capacity of hosting, which is 2^{32} bit addresses. To improve capacity, IPv6 was planned and structured with enhanced features to provide better services than IPv4. With its improved capacity, IPv6 is capable of delivering 2^{128} bit addresses. In addition, IPv6 eliminates the use of Network Address Translation (NAT) and Variable Length Subnet Mask (VLSM) since it has enough IP addresses for all the users around the globe [16].

G.711 Codec

Hypothetically, G.711 codec delivers an exceptional class of voice service that requires higher processing and bandwidth [17].

According to [18], G.711 codec uses 80-bytes of the frame for voice encoding over a 10ms interval. The G.711 codec uses about 64kbps for a one direction call and 128kbps for two-way communication.

G.723 Codec

For voice and multimedia, the ITU-T developed another codec called the G.723 codec, which is an extension of the G.721 codec. In theory, G.723 codec is not suitable for sounds because of its lower-quality output [19]. In [20] the author noted that the G.723 codec was especially fashioned for voice encoding at low transmission capacity. The G.723 codec can operate at 6.4kbps and 5.3kbps with 24-bytes and 20-bytes, respectively.

G.729 Codec

The G.729 codec ITU-T standard engages the Conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP) algorithm to compress a payload for a low bit rate. In theory, G.729 codec delivers reasonably high speech performance [21]. The author in [22] noted that G.729.1 codec has an in-built scalable design set up as an extension of an existing specification. The G.729 codec can interoperate at 8kbps.

In Table 1, the VoIP codec specifications illustrate and compare three ITU-T codecs showing their features, namely, bandwidth, sample period, frame size, and rate.

Table 1: VoIP codec specifications [13]

Codec	G.711	G.723.1	G.729
Bandwidth (Kbps)	64	5.3/6.4	8
Sample period (ms)	20	30	20
Frame size (payload)	160	20/24	20
Rate (Packets /s)	50	33	50

Generic Routing Encapsulation Protocol

GRE, a tunnelling protocol developed by Cisco, which was later standardized by the Internet Engineering Task Force (IETF) [23,24], works by encapsulating layer packet within another IP datagram. It supports broadcast, multicast and other non-IP protocol traffic. GRE tunnels are not secured as data payload are not encrypted and verified. Hence, in real-time, IPsec is integrated with GRE to guarantee the security and integrity of packets.

Internet Protocol Security

In 1998, the IETF drafted the Internet Protocol Security in Request for comments 2411 [25], which was obsoleted by RFC 6071 [26]. IPsec ensures the security of layer three packets for both IPv4 and IPv6. IPsec provides level peer verification, data source validation, data privacy, and data integrity [27–31]. It supports encryption protocol such as DES (Data encryption standard), 3DES (Triple data encryption standard), AES (Advanced encryption standard), authentication protocol such as MD5 (Message Digest 5) and SHA (Secure hash algorithm) [32].

The codecs and IP protocols explained in this section were used in this study. The following section discussed the performance metrics used for evaluation in this study, such as delay, jitter and packet loss.

5. Traffic Generating Tool

D – ITG (Distributed Internet Traffic Generator)

A packet-level traffic generator developed by [33] allows the simultaneous generation of multiple flows of traffic. It can generate realistic traffic patterns from Internet protocols like TCP, UDP, ICMP, and VoIP. Delay, Jitter, Packet Loss are metrics supported by D-ITG. It is a multi-platform application as it works on Linux and Windows. D-ITG can send traffic via UDP or TCP, supporting various voice codecs [1]. According to [34] distributed Internet traffic generator can generate IPv4 and IPv6 traffic that precisely imitates the forms explicated by the Inter Departure time (IDT) and the packet size (PS) stochastic process. Embedded in D-ITG are some statistical models proposed to replicate traffic related to Voice activity detection, DNS, Telnet and VoIP. D-ITG's architecture and components are depicted in Figure 1. The Sender and Receiver are used to generate traffic. The traffic report is held by the ITGLog and can be decoded using the ITGDec, which is the analyzer. The controller controls the activities of each component of the D-ITG.
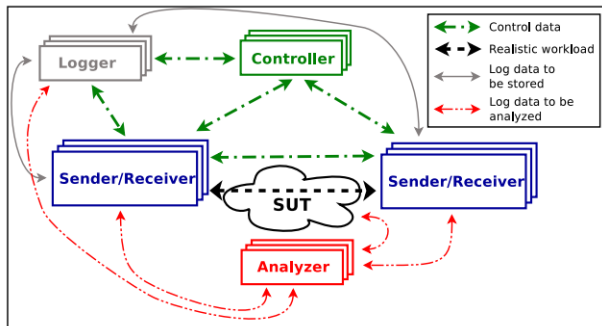
Figure 1: Architecture of D-ITG and its components [33]

6. Performance Metrics

Delay

The time elapsed for a block of data to transit from the sender through the receiver's network infrastructure is regarded as delay. Voice packets can be degraded if it takes a long time to deliver packets from end-to-end [35]

D_t, the total voice packet delay, is calculated thus:

$$D_t = D_n + D_e + D_d + D_c + D_{de} \qquad (1)$$

where the components are delays due to:

D_n = Network

D_e = Encoding

D_d = Decoding

D_c = Compression

D_{de} = Decompression

Jitter

Delay variation, often referred to as jitter, is a vital QoS metric in voice communication. It is the measurement of the time difference between packets sent and packets arriving. According to [36,37] a jitter of above 30ms will adversely affect call quality. Jitter can be expressed as follows:

$$J_{(i)} = J_{(i-1)} + \frac{\left| D_{(i-1,i)} \right| - J_{(i-1)}}{16} \qquad (2)$$

where:

$J_{(i)}$ = jitter of the ith packet

$J_{(i-1)}$ = jitter of the (i-1) packet, and

$D_{(i-1,i)}$ = delay between the packet (i) and (i-1)

Packet Loss

When a packet in transit between two or more hosts on a network fails to reach their destination, packet loss is said to have occurred. It occurs due to an error in data transmission or congestion in the network [38]. ITU-T recommends a maximum of 3% packet loss for VoIP [39]. The Packet Loss rate is computed using the number of RTP packets anticipated and acknowledged from each source. The number of packets is counted as they arrive. The receiver computes the number of packets expected using the difference between the highest segment received and the first segment [40]. Packet loss can be expressed as follows:

$$P_l = P_e - P_r \qquad (3)$$

where:

P_l = Packet loss

P_e = Packet expected

P_r = Packet received

In Table 2, the VoIP QoS requirement as recommended by the ITU is given as follows:

Table 2: VoIP Quality Requirement [13]

QoS requirement	Good	Acceptable	Poor
Delay (ms)	0 – 150	150 – 300	>300
Jitters (ms)	0 – 20	20 – 50	>50
Packet Loss (%)	0 – 1	1 – 3	>3

7. Literature Review

A review of related studies is carried out in this section. A comparison of the performance VoIP over BGP/MPLS, VPN and MPLS network was carried out in [5]. The study was simulated under OPNET Modeler using the G.711, G.723.1, and G.729A codecs. The study submitted that the G.729A codec gave a

superior performance over the BGP/MPLS VPN network. The author in [41] noted that encryption is not a standard service in BGP/MPLS VPN, thus not immune to DoS attacks or intrusion.

Furthermore, Performance analysis of IPsec VPN over VoIP network was investigated in [42]. The experimental result showed a rational decline in performance due to the encryption and authentication process carried out by IPsec. Besides, no VoIP codec was evaluated in the study.

The performance of VoIP over GRE tunnel, which provides no extra layer of security, was simulated in [11]. The study submitted that GRE did not lead to a substantial rise in the QoS parameters such as delay and call setup time. GRE being a transparent tunnelling protocol, is not suitable for conveying sensitive traffic such as VoIP. Traffic going over the GRE tunnel is very prone to DoS attack and packet sniffing.

Evaluation of VoIP over multiple Multiprotocol Label Switching (MPLS) tunnelling was carried out in [43]. The study adopted various MPLS architecture with IPsec Secure Hash Algorithm (SHA) and AES for encryption. According to the study, giving the extra layer added by IPsec, VoIP becomes unusable and results in poor speech quality.

Furthermore, an analytical evaluation of layer three tunnelling protocols using VoIP traffic was conducted in [12]. The study attributed the stability of delay variation for non-IPsec scenarios to the non-verification of layer three packets by IPsec encryption and integrity controls. The study concluded that the high delay value reported was a result of IPsec encryption and authentication.

IPv4 to IPv6 and IPv6 to IPv4 transition mechanism was evaluated in [44] with and without virtual private network like IPsec. The codecs were evaluated over Point-to-Point Protocol and IPsec over 4to6 and 6to4 transition mechanism. The metric assessed for the VoIP codec is throughput. The study reported that G.711.2 and G.723.1 codec had the highest throughput.

In [10] the author adopted the riverbed modeller to evaluate SIP performance first with VPN and without VPN. The study evaluated the G.711 codec. The study results revealed that VPN does not lead to higher call set up time as opposed to the submission of [45]. QoS metric like delay, jitter, and packet loss was not evaluated in the study.

The authors in [46] evaluated the performance of the G.711 codec using Open Shortest Path First (OSPF) and Routing Information Protocol (RIP) on IPv4 and IPv6. According to the study, the quality of voice communication in an IP network is dependent on variables such as codec, delay, jitter, packet loss, and routing protocols to reduce convergence time in the case of a link failure. OSPFv3 and RIP next generation (RIPng) performed better in terms of packet delay disparity, end-to-end delay and jitter when compared with OSPFv2 and RIPv2. No tunnelling protocol like GRE was used in the study.

The performance of VoIP over IPv4 and IPv6 was simulated in [47] using the OPNET modeller. The experiment was conducted using the G.711 codec with no tunnelling protocol employed. The study submitted that VoIP on IPv6 had superior performance compared to IPv4.

The study in [48] evaluated three MPLS architecture, IP MPLS, MPLS VPN and MPLS IPsec. IP Service Level Agreement (IP SLA) was used to generate VoIP traffic with delay, jitter, loss rate used as evaluation parameters. The study adopted GNS3 as the network simulator, with G.711 codec being the only codec evaluated over the three MPLS architecture. The study reported that MPLS IPsec resulted in the degradation of voice communication while IP MPLS and MPLS VPN gave acceptable latency and jitter results.

This present study seeks to expand the scope to include IPv4 and IPv6 based on the literature reviewed. Three codecs were used (G.711.1, G.723.1, G.729.3) in this study. The methodology is discussed in the next section.

8. Methodology

This study was carried out in an emulation environment using GNS3. Distributed Internet Traffic Generator (D-ITG) installed on Ubuntu Linux 16.04LTS was used to generate VoIP traffic. The G.711.1, G.723.1, G.729.3 codec were evaluated on the GRE with IPsec tunnel over IPv4 and IPv6.

In this study, a site-to-site VPN topology was used, as depicted in Figure 2. The study considered the performance evaluation of VoIP codec on GRE with IPsec encryption and authentication features over IPv4 and IPv6 Protocol.
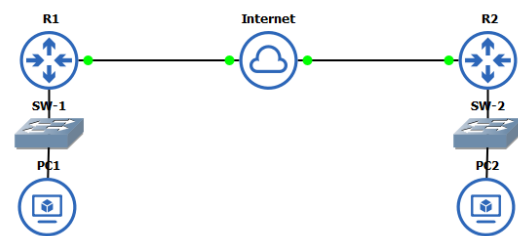
Figure 2: Network topology for the experimental testbed

The study considered three QoS parameters: delay, jitter, and packet loss. Table 3 shows the VoIP parameters used in the study, while Table 4 shows the IPsec configuration Profile.

Table 3: VoIP Parameters

Traffic	VoIP
Codec	G.711.1; G.723.1; G.729.3
Number of packets	100 – 1000 p/s

Table 4: IPsec Configuration Profile

Encryption	3DES
Packet integrity	MD5
Authentication	Pre-Share

Lifetime	86400s
Cisco IOS	c7200

9. Result and Discussion

The result of this study is subdivided into sections A, B, C. Section A discusses the results for the delay on GRE with IPsec over IPv4 and IPv6. Section B discusses the result for Jitter on GRE with IPsec over IPv4 and IPv6. Section C discusses Packet Loss on GRE with IPsec over IPv4 and IPv6. Table 5, 6, and 7 presents the summary of results for the delay, jitter, and packet loss.

Table 5: Delay (ms) performance summary for GRE with IPsec over IPv4 and IPv6

Codec	G.711.1	G.723.1	G.729.3
GRE with IPsec (IPv4)	196.785	69.186	64.583
GRE with IPsec (IPv6)	389.306	78.056	95.333

From the result in Table 5, the G.723.1 codec gave superior performance compared to the other two codecs. On GRE with IPsec for IPv6, the G.723.1 codec gave a superior performance for delay at 78.056ms compared to G.711.1 at 389.306ms and G.729.3 at 95.333ms. High delay often leads to low speech quality in a VoIP network. The lower the delay, the better the voice quality. According to [13], the ITU-T recommended that VoIP delay should not exceed 300ms. The result in Table 5 clearly shows that the G.711.1 codec exceeds the ITU requirement for the delay on IPv6. Of the three codecs evaluated, the G.723.1 and the G.729.3 codec stood out on IPv4 and IPv6.

Table 6: Jitter (ms) performance summary for GRE with IPsec over IPv4 and IPv6

Codec	G.711.1	G.723.1	G.729.3
GRE with IPsec (IPv4)	18.038	4.561	4.968
GRE with IPsec (IPv6)	48.31	4.588	11.018

In Table 6, the result for jitter is presented for GRE with IPsec over IPv4 and IPv6. The performance of the three codecs on the GRE with IPsec over IPv4 was superior to IPv6. In Table 2, the acceptable QoS requirement for jitter is 50ms. In Table 6, the G.711.1 codec reported a high delay variation on GRE with IP sec over IPv6 at 48.31ms, making the codec not suitable for IPv6 GRE with IPsec deployment. However, the G.711.1 codec fared better on GRE with IPsec over IPv4 with a reported value of 18.038ms. On the other hand, the G.723.1 codec and the G.729.3 codec gave a superior performance over IPv4 GRE with IPsec deployment. The reported value for both codecs were 4.561ms and 4.968ms, respectively. On the GRE with IPsec over IPv6, the G.723.1 codec gave superior performance compared to the two

other codecs. The G.723.1 codec reported a performance value of 4.588ms on the GRE with IPsec over the IPv6 protocol. Comparing both the IPv4 and IPv6 results, it is safe to submit that the G.723.1 is preferred to the other two codecs.

Table 7: Packet loss (%) performance summary for GRE with IPsec over IPv4 and IPv6

Codec	G.711.1	G.723.1	G.729.3
GRE with IPsec (IPv4)	67.53	0.1	1.34
GRE with IPsec (IPv6)	86.06	0.45	22.56

For packet loss, the acceptable range is 3% maximum as depicted in Table 2. From Table 7, the G.711.1 had the highest packet loss at 67.53% on IPv4 and 86.06% on IPv6. Hence, the G.711.1 codec becomes unusable on both IP Protocols. Furthermore, on the GRE with IPsec over IPv4, the G.723.1 codec and G.729.3 codec reported a value of 0.1% and 1.34%, respectively, which fall within the acceptable range of the ITU recommendations as depicted in Table 2. For the G.723.1 codec and G.729.3 codec, it is safe to submit that both codecs are suitable for the GRE with IPsec over IPv4 deployment.

On the other hand, the G.729.3 codec exceeded the ITU requirement for packet loss on the GRE with IPsec IPv6 deployment with a reported value of 22.56%. Hence the G.729.3 codec is not fit for use on the IPv6 GRE with IPsec deployment. On the GRE with IPsec over IPv6 deployment, the G.723.1 codec gave a superior performance at 0.45%, which is within the acceptable limits of the ITU recommendation as depicted in Table 2. For the packet loss metric, this study submits that the G.723.1 codec is suitable for IPv4 and IPv6 deployment of GRE with IPsec while G.729.3 codec is suitable for IPv4 GRE with IPsec deployment.

10. Conclusion

This study has provided quantitative evidence by evaluating the performance of three VoIP codecs on GRE with IPsec over IPv4 and IPv6. VoIP codec has a vital role in IPsec enabled network architecture. Delay, jitter, and packet loss were the three QoS parameters evaluated in the study. Since these QoS metrics cannot independently determine the quality of voice communication over IP networks, then the codec that gave a superior performance across the three QoS metric would be considered an efficient codec for GRE with IPsec implementation.

The only codec in this study that efficiently balances acceptable delay, jitter, and packet loss across the IPv4 and IPv6 deployment is the G.723.1 codec. The G.723.1 codec has proved to be usable for IPsec implementations on GRE with IPsec. This study agrees with [49] that regardless of IPsec, the codec of choice influences the voice communication quality output over the IP network. Low bandwidth codec such as G.723.1 and G.729.3 are better choices for IPsec enabled VoIP implementations.

Specific limitation exists in the present study which the methodology could not accommodate. Codecs like Speex, Silk, iLBC, OPUS, and CELT were not evaluated as the traffic generating tool (D-ITG) does not support them. Other IPsec encryption algorithm like AES can also be evaluated in further studies alongside SHA. The present study did not consider any routing protocol like OSPF, EIGRP, and RIP in IPv4 and IPv6 environment. These areas would be addressed in further studies.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

We would like to appreciate the Department of Computer Science, Redeemer's University, for the privilege to conduct this research.

References

[1] H. Sathu, M.A. Shah, "Performance comparison of VoIP codecs on multiple operating systems using IPv4 and IPv6," International Journal of E-Education, e-Business, e-Management and e-Learning, **2**(2), 122, 2012.

[2] S. V. Subramanian, R. Dutta, Measuring SIP proxy server performance, Springer International Publishing, 2013, doi:10.1007/978-3-319-00990-2.

[3] S. Kent, R. Atkinson, RFC2402: IP authentication header, 1998.

[4] R. Atkinson, "IP Encapsulating Security Payload (ESP)," RFC 4303, (1827), 1–44, 1995, doi:10.17487/RFC4303 M4 - Citavi.

[5] I.S.I. Alsukayti, T.J. Dennis, C. Edwards, "Performance analysis of VoIP over BGP-MPLS VPN technology," in PGNet conference, 2011.

[6] K.A. Ogudo, "Analyzing generic routing encapsulation (GRE) and IP Security (IPSec) tunneling protocols for secured communication over public networks," in icABCD 2019 - 2nd International Conference on Advances in Big Data, Computing and Data Communication Systems, Institute of Electrical and Electronics Engineers Inc.: 1–9, 2019, doi:10.1109/ICABCD.2019.8851004.

[7] S. Jahan, M.S. Rahman, S. Saha, "Application specific tunneling protocol selection for Virtual Private Networks," in Proceedings of 2017 International Conference on Networking, Systems and Security, NSysS 2017, Institute of Electrical and Electronics Engineers Inc.: 39–44, 2017, doi:10.1109/NSysS.2017.7885799.

[8] J.G.M. Srivastava, R. Sheeja, "Design and Implementation of Crypto based Water Marking Techniques for EHR Security," Test Engineering and Management, **82**, 10788–10792, 2020.

[9] K. Salah, Deploying VoIP in Existing IP Networks, CRC Press: 19–40, 2018, doi:10.1201/9781420070217.pt1.

[10] A.A. Eskandar, M.R. Syed, M.B. Zarei, "SIP over IP VPN : Performance Analysis," in Proceedings on the International Conference on Internet Computing (ICOMP), 1, 2014.

[11] A.A. Eskandar, M.R. Syed, Z.M. Bahareh, "Performance analysis of VOIP over GRE tunnel," International Journal of Computer Network and Information Security, **7**(12), 1, 2015.

[12] F. Bensalah, N. El Kamoun, A. Bahnasse, "Analytical performance and evaluation of the scalability of layer 3 tunneling protocols: case of voice traffic over IP," IJCNS International Journal of Computer Science and Network Security, **17**(4), 361–369, 2017.

[13] J. Yu, I. Al-Ajarmeh, "Call admission control and traffic engineering of VoIP," in 2007 Second International Conference on Digital Telecommunications (ICDT'07), 11, 2007, doi:https://dx.doi.org/10.1109/ICDT.2007.44.

[14] Y. Rekhter, B.M. Chrysler, D. Karrenber, G.J. de Grooe, E. Lear, Address Allocation for Private Internets, Journal of Chemical Information and Modeling, **5**(9), 1689–1699, 2013.

[15] J. Davies, A. Northrup, Windows Server 2008 networking and network access protection (NAP), Microsoft Press, 2008.

[16] S. Deering, R. Hinden, others, Internet protocol, version 6 (IPv6) specification, 1998.

[17] L. Roychoudhuri, E. Al-Shaer, H. Hamed, G.B. Brewster, "Audio transmission over the Internet: Experiments and observations," in IEEE International Conference on Communications, 2003. ICC'03., 552–556,

2003.

[18] Y. Jeong, S. Kakumanu, C.L. Tsao, R. Sivakumar, "VoIP over Wi-Fi networks: Performance analysis and acceleration algorithms," Mobile Networks and Applications, **14**(4), 523–538, 2009, doi:10.1007/s11036-009-0157-6.

[19] H. Cui, K. Tang, T. Cheng, "Audio as a support to low bit rate multimedia communication," in ICCT'98. 1998 International Conference on Communication Technology. Proceedings (IEEE Cat. No. 98EX243), 544–547, 1998.

[20] M. Menth, A. Binzenhöfer, S. Mühleck, "Source models for speech traffic revisited," IEEE/ACM Transactions on Networking, **17**(4), 1042–1051, 2009, doi:10.1109/TNET.2008.2006222.

[21] L. Ding, R.A. Goubran, "Assessment of effects of packet loss on speech quality in VoIP," in Proceedings - 2nd IEEE International Workshop on Haptic, Audio and Visual Environments and their Applications, HAVE 2003, Institute of Electrical and Electronics Engineers Inc.: 49–54, 2003, doi:10.1109/HAVE.2003.1244724.

[22] I. Varga, S. Proust, H. Taddei, ITU-T G.729.1 scalable codec for new wideband services, IEEE Communications Magazine, **47**(10), 131–137, 2009, doi:10.1109/MCOM.2009.5273820.

[23] D. Farinacci, T. Li, S. Hanks, D. Meyer, P. Traina, "Generic Routing Encapsulation (GRE)," Request for Comments, 1–9, 2000, doi:http://dx.doi.org/10.17487/RFC2784.

[24] G. Dommety, Key and Sequence Number Extensions to GRE, Request for Comments, 1–7, 2000, doi:http://dx.doi.org/10.17487/RFC2890.

[25] R. Thayer, N. Doraswamy, R. Glenn, {IP} Security Document Roadmap, 1998.

[26] S. Frankel, S. Krishnan, "RFC6071: IP Security (IPsec) and Internet Key Exchange (IKE) Document Roadmap," Request for Comments, **6071**, 1–63, 2011, doi:10.17487/rfc6071.

[27] S. Ariga, K. Nagahashi, M. Minami, H. Esaki, J. Murai, "Performance evaluation of data transmission using IPSec over IPv6 networks," in Proc. INET, 18–21, 2000.

[28] N. Ferguson, B. Schneier, A cryptographic evaluation of IPsec, 1999.

[29] R. Yuan, W.T. Strayer, Virtual private networks: technologies and solutions, Addison-Wesley Longman Publishing Co., Inc., 2001.

[30] W.B. Diab, S. Tohme, C. Bassil, "Critical VPN security analysis and new approach for securing VoIP communications over VPN networks," in WMuNeP'07: Proceedings of the Third ACM Workshop on Wireless Multimedia Networking and Performance Modeling, 92–96, 2007, doi:10.1145/1298216.1298238.

[31] J. Wu, "Implementation of virtual private network based on IPSec protocol," in 2009 ETP International Conference on Future Computer and Communication, IEEE: 138–141, 2009, doi:https://dx.doi.org/10.1109/FCC.2009.16.

[32] R. Weerawarna, TCP/UDP network performance evaluation of various IPSEC algorithms: an empirical test-bed analysis of a virtual private network protocol., 2013.

[33] A. Botta, A. Dainotti, A. Pescapé, "A tool for the generation of realistic network workload for emerging networking scenarios," Computer Networks, **56**(15), 3531–3547, 2012, doi:10.1016/j.comnet.2012.02.019.

[34] A. Botta, A. Dainotti, A. Pescapè, "Multi-protocol and multi-platform traffic generation and measurement," INFOCOM 2007 Demo Session, **2010**, 4–5, 2007.

[35] K. Kim, Y.J. Choi, "Performance comparison of various VoIP codecs in wireless environments," in Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication, ICUIMC 2011, 2011, doi:10.1145/1968613.1968718.

[36] A. Culleton, "Evaluation of Voip Technologies As a Replacement for Traditional Pstn Based Pbx Systems," 2006.

[37] M. Ahmed, M.A. Suhaimi, Q.S. Md Faisal, S. Haseeb, "Evaluating QoS performance of streaming video on both IPv4 and IPv6 protocols," in Proceedings of the 2007 spring simulaiton multiconference-Volume 1, 109–116, 2007.

[38] Y. Ennaji, M. Boulmaif, C. Alaoui, "Experimental analysis of video performance over wireless local area networks," in International Conference on Multimedia Computing and Systems -Proceedings, 488–494, 2009, doi:10.1109/MMCS.2009.5256645.

[39] B. Goode, "Voice over Internet Protocol (VoIP)," Proceedings of the IEEE, **90**(9), 1495–1517, 2002, doi:10.1109/JPROC.2002.802005.

[40] A.-V.T.W. Group, H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, others, RFC1889: RTP: A transport protocol for real-time applications, RFC Editor, 1996.

[41] M. Behringer, "Analysis of the Security of BGP/MPLS IP Virtual Private Networks (VPNs) Status," RFC 4381, 1–22, 2006.

[42] M. Babu, "Performance Analysis of IPSec VPN over VoIP Networks Using OPNET," International Journal of Advanced Research in Computer Science and Software Engineering, **2**(9), 38–44, 2012.

[43] F. Bensalah, N. El Kamoun, A. Bahnasse, "Scalability evaluation of VOIP over various MPLS tunneling under OPNET modeler," Indian Journal of Science and Technology, **10**(29), 1–8, 2017.

[44] S. Narayan, S. Ishrar, A. Kumar, R. Gupta, Z. Khan, "Performance analysis of 4to6 and 6to4 transition mechanisms over point to point and IPSec VPN protocols," in IFIP International Conference on Wireless and Optical Communications Networks, WOCN, IEEE Computer Society: 1–7, 2016, doi:10.1109/WOCN.2016.7759027.

[45] G. Salama, M. Shehab, A. Hafez, M. Zaki, "Performance Analysis of Transmitting Voice over Communication Links Implementing IPsec," International Conference on Aerospace Sciences and Aviation Technology, **13**(AEROSPACE SCIENCES), 1–12, 2009, doi:10.21608/asat.2009.23489.

[46] M. Syed, I.Y. Ambore, "Performance evaluation of OSPF and RIP on IPv4 & IPv6 technology using G.711 codec," International Journal of Computer Networks and Communications, **8**(6), 1–15, 2016, doi:10.5121/ijcnc.2016.8601.

[47] M. Ahmed, A.T. Litchfield, S. Ahmed, A. Mahmood, M.E.H. Meazi, "VoIP performance analysis over IPv4 and IPv6," International Journal of Computer Network and Information Security, **6**(11), 43, 2014.

[48] F. Bensalah, N. El Kamoun, A. Bahnasse, "Evaluation of tunnel layer impact on VOIP performances (IP-MPLS-MPLS VPN-MPLS VPN IPsec)," International Journal of Computer Science and Network Security (IJCSNS), **17**(3), 87, 2017.

[49] E. Antwi-Boasiako, E. Kuada, K. Boakye-Boateng, "Role of codec selection on the performance of IPsec secured VoIP," in 2016 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2016, Institute of Electrical and Electronics Engineers Inc.: 2508–2514, 2016, doi:10.1109/ICACCI.2016.7732434.

.

Low-Power Primary Cell with Water-Based Electrolyte for Powering of Wireless Sensors

Dmitry Petrov[*], Ulrich Hilleringmann

Sensor Technology Group, Paderborn University, Paderborn, 33098, Germany

ARTICLE INFO	ABSTRACT
	In this work we discuss a special technique for powering of sensor systems, based on a low-power primary cell consisting of two electrodes, made from different metals, with water (lake, rain or tap water) used as an electrolyte. Once placed into an aqueous solution, the primary cell generates a small electric current, which may be utilized for powering of sensor systems. The generated electrical energy is fed into an energy storage (capacitor). After transformation of the voltage by a step-up converter, it is used for supplying the electrical sensor circuit. The expected output power of the developed circuit is 10-15 mA by 2 V output voltage during 0.2-0.5 second. The improved voltage converter topology with implemented maximal power point techniques allows significant reduction of the energy storage's size in the second revision of the circuit and thus reduction of the resulting size of the board. The implemented sensor board with discussed powering technique, assembled in Paderborn University was already tested in different practical scenarios.

1. Introduction

This paper is an extension of work originally presented in 2020 IEEE Sensors [1]. Energy harvesting technologies are often used for powering of wireless sensors, especially for industrial monitoring applications [2]. Powering techniques like photovoltaic and thermoelectric generators are widely presented on the market, but there is always a need for new ideas for energy powering techniques [3]. The Sensor Technology Group at Paderborn University is directly developing the power supply system for wireless sensors based on a galvanic cell used for generation of small electrical currents with a water-based electrolyte (lake, rain or tap water) [1, 4].

The wireless sensors powered by the developed electrochemical cell can be used e.g., for construction monitoring and detection of water leakages in roofs or wall structures of industrial and private buildings. Such sensors integrated in building structures can get activated by contact with water, caused by a water leakage. This water serves as an electrolyte for the galvanic cell, which provides the electrical energy for the sensor electronics. The main advantage of this approach is a long sensor lifetime in the not activated state, longer than most batteries' durations that are currently available on the market (usually up to 10 years).

2. Implementation of a galvanic cell

The implemented galvanic cell with two metal electrodes mounted on the non-conductive substrate (composite epoxy material) is depicted in Figure 1, left. To perform the measurements, the developed cell is placed inside the vessel with tap water serving as an electrolyte (Figure 1, right).
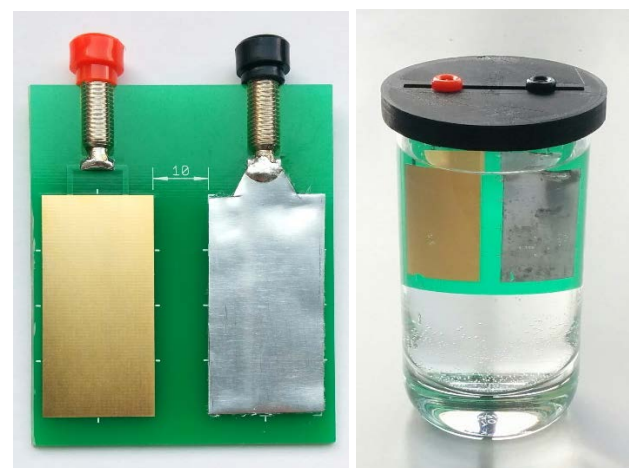

Figure 1: The implemented setup of a galvanic cell with two metal electrodes (left), and galvanic cell placed in water, serving as electrolyte (right)

[*]Corresponding Author: Dmitry Petrov, Email: petrov@sensorik.uni-paderborn.de

The difference of electric potential between two electrodes results from the standard potential of the chosen electrode metals [5-7].

During the experimental measurements, multiple combinations of electrode materials have been compared and tested. The practically established parameters of the galvanic cells (short-circuit current and open-circuit voltage) are depicted in Figure 2.
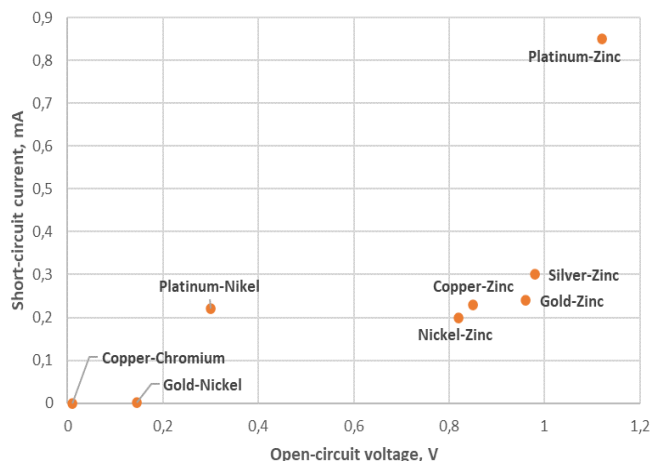
Figure 2: Established parameters for different metals, used for galvanic cell electrodes with electrode areas of 8 cm2 (4 x 2 cm), in tap water [1]

Table 1 shows the standard electrochemical potentials of commonly used electrode materials [5].

Table 1: Standard electrochemical potential of metals commonly used as electrode materials

Electrode material	Reaction pair	Standard potential, V
Gold	$Au^{3+} + 3e^-$	1.5
Gold	$Au^{2+} + 2e^-$	1.69
Platinum	$Pt^{2+} + 2e^-$	1.2
Silver	$Ag^+ + e^-$	0.80
Copper	$Cu^+ + e^-$	0.52
Zinc	$Zn^{2+} + 2e^-$	-0.76

Although the pair platinum and zinc results in the highest value of output power, a combination of zinc and gold metals has been selected for the developed galvanic cell [8], because gold finish is a standard coating technique in the production of printed circuit boards (PCB). A gold layer of 100-200 nm is typically used to protect the contact pads of the PCB against corrosion and simplifies the soldering process [9]. An adequate gold thickness allows the stable operation of the implemented cell, although, any defect in the gold layer (like fracture or porosity) reduces the generated power by reducing the open-circuit voltage.

Figure 3 shows the established performance of the implemented cell with gold and zinc electrodes of 8 cm^2 area (4 x 2 cm), loaded to a capacitance of 50 mF. The distance between the electrodes was 10 mm. The calculated energy saved in the capacitor achieved 9,8 mJ after 5 minutes and about 11 mJ in 10 minutes.
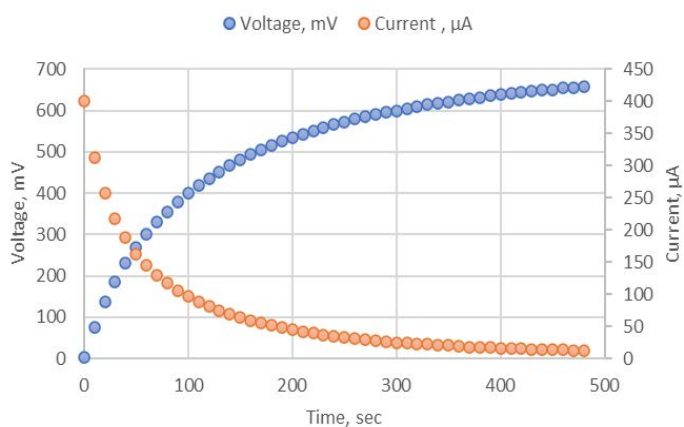
Figure 3: Current and voltage curves charging a capacitance of 50 mF [1]

Figure 4 depicts the load curve of the developed galvanic cell. The established open-circuit voltage is around 700 mV, the corresponding maximum power point (MPP) is about 250 mV. The calculated generated power achieves up to 44 µW, if the galvanic cell operates close to the MPP voltage.

The performance of the tested galvanic cell working on the capacitive load (shown in Figure 3) could be improved by connecting a voltage converter working with the established maximum power point between the output of the galvanic cell and charged capacitance. This design of the voltage converter will be discussed in detail in chapter 4.
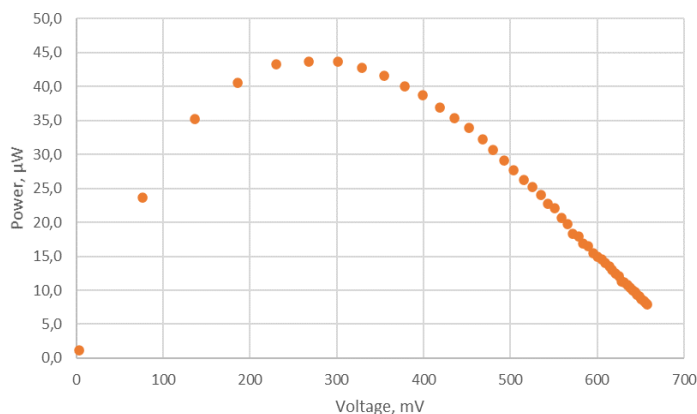
Figure 4: The load curve of the galvanic cell consisting of Au and Zn.

3. Sensor board design

The selected combination of gold and zinc electrodes connected by water ensures an open circuit voltage of about 700 mV. This voltage is lower than the operating voltage of typical electronic circuits (from 1.8 V to 2.0 V for common low-voltage microchips). The maximal output current of the developed galvanic cell is between 100 and 200 µA (short-circuit current), this is much lower than the typical current consumption of wireless transmitter ICs (10 to 15 mA). Therefore, a step-up voltage converter circuit was implemented to boost the output cell voltage. The harvested energy has to be integrated and must be stored temporarily in an energy storage circuit, e.g., a capacitor or super capacitor, until the energy amount is high enough to power the sensor circuit for the required time (0.2 to 0.5 seconds).

The first revision of the developed sensor board uses a super capacitor as an energy storage component. It is connected between the output of the galvanic cell and the input of the voltage boost circuit. The block diagram of this first revision is shown in Figure 5. After charging the super capacitor up to 700 mV (threshold voltage), the chosen step-up converter is starting up. It boosts up performs the voltage to 2.0 V to supply the electrical circuit of the sensor board. After complete discharging of the super capacitor, the charging cycle will be repeated. The low-voltage synchronous boost converter, TPS61201 from Texas Instruments [10], was implemented in the first revision of the developed sensor board.
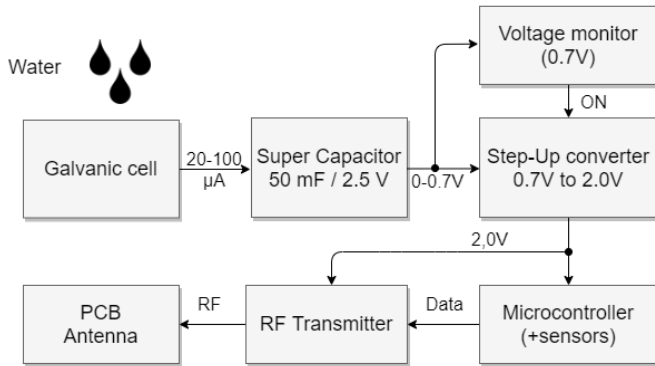
Figure 5: Block diagram of the developed circuit board with energy storage connected to the input of the voltage converter

The required capacitance of the super capacitor is between 20 and 50 mF, which is enough to deliver the estimated current of 10-15 mA over a time interval of 0.2 to 0.5 seconds. For the super capacitor, a model produced by AVX company with a capacity of 50 mF was used in the voltage converter circuit [11]. To supervise the charging process of the super capacitor, a discrete low-voltage detector by RICOH with 0.7 V threshold voltage is used [12].

The first revision of the sensor PCB (with a size of 27 mm x 100 mm) is depict in Figure 6. Two metal electrodes, each at a size of 8 cm^2, are located on the bottom side of the board. The electronic components are placed on the upper side of the PCB. The electrical contact between the gold coated positive electrode and the electrical circuit is performed by a through-hole connection (the black point at the top right corner of the left electrode). The negative (zinc) electrode is cut out of a zinc plate, the mechanical and electrical connections of the zinc electrode to the upper side of the PCB is realized by two flaps, inserted into board openings and connected to the contact pads on the top side. The openings are sealed with silicone to prevent the electrical contact of the sensor circuit with the electrolyte.

On the top side of the developed sensor board the supercapacitor is located on the left, the voltage converter and transmitter circuit in the middle of the board, and the RF antenna on the right side. A complete sealing of the electrical circuit (the complete top side of the board) could also be implemented for practical applications to protect the circuit from the moisture and dust.

The total component cost for the first revision of the developed sensor PCB is about 9.0 €(calculated for 1000 pieces).
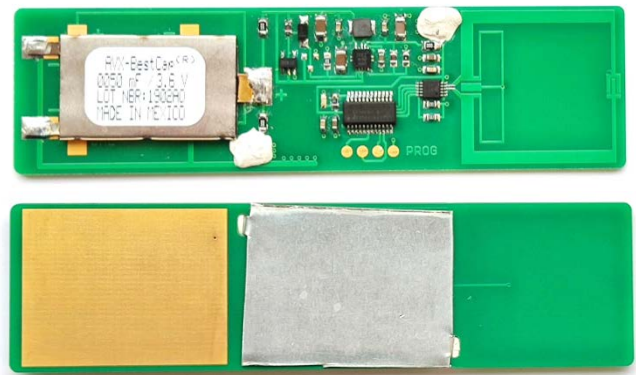
Figure 6: Design of the first revision of the sensor board (top and bottom side)

4. Optimized sensor board design

The minimal size of the first revision of the developed board is defined by the size of the super capacitor (energy storage), which occupies about 30% of the board area. The new improved topology of the power supply circuit which is currently implemented, uses the maximum power point (MPP) tracking technique. It allows an operation of the galvanic cell close to the point of maximal efficiency. In this topology the energy storage is placed between the output of the step-up converter and the sensor electronics. Here the storage capacitor can be charged to a higher voltage level (3.3 V instead of 0.7 V) giving the possibility to save considerably more electrical energy, thus reducing the required capacitance. This allows less expensive alternatives to the supercapacitor, like tantalum capacitors. By using the new topology, the sensor PCB area may be reduced by about 30%. The block diagram of the second revision of the sensor board with implemented MPP tracking is depict in Figure 7.
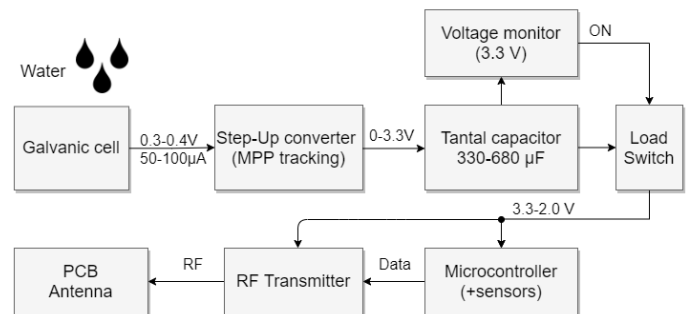
Figure 7: Block diagram of the sensor board with energy storage at the output of the voltage converter and implemented MPP tracking

Using the selected ADP5091 voltage converter with implemented MPP tracking by Analog Devices [13] increases the output power of the galvanic cell and significantly reduces the charging time (from 8-10 minutes to 3-4 minutes). Further optimisation of the electrode sizes and relation between gold and zinc area help to reduce the total size of the sensor board from 100 mm x 27 mm to 75 mm x 27 mm (reduction by 25%). The component costs in this version were also reduced by about 30%. The total costs for the second revision of the sensor PCB are about 6.5 €(from 1000 pieces).

The second revision of the sensor board is shown in Figure 8. The tantalum capacitors for storing the generated electrical energy (two orange blocks on the top side of the PCB) occupy only a small

part of the sensor board area. The voltage converter is located on the left and the RF antenna with the transmitter circuit on the right side of the sensor PCB.
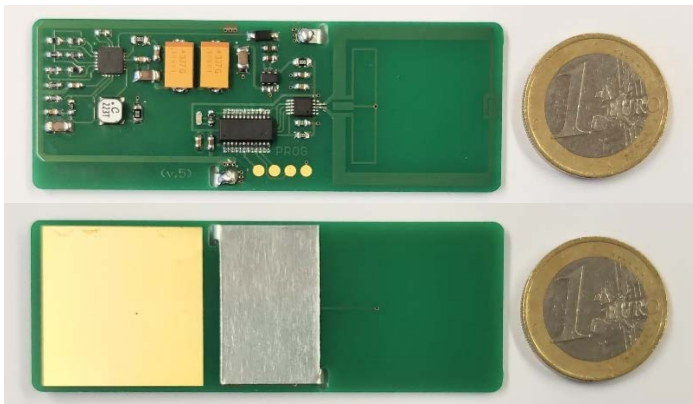
Figure 8: Design of the second revision of the sensor board (top and bottom side)

Both, gold and zinc electrodes are placed on the bottom side of the board. The optimized relation between gold and zinc areas (up to 1:2) allows to significantly reduce the electrode area. The size of the gold electrode was reduced from 25 x 32 mm to 25 x 25 mm (reduction by 22%) and the size of the zinc electrode was reduced from 25 x 32 mm to 25 x 14 mm (reduction by 56%).

5. Wireless data transmission

The amount of collected electrical energy should be high enough to perform the initialization of the EFM8 low-power MCU by Silicon Labs [14], the measurement tasks, and additionally the wireless transmission of the data packages via Si4012 RF transmitter [15]. The power consumption of the RF transmitter circuit is most significant compared to other parts of the circuit. Additionally, the current consumption changes very quickly, depending of the data transmission cycle. Knowing the current profile of the RF transmitter helps to design the power supply circuit and to calculate the required amount of electrical energy for the sensor operation. The measured real-time current consumption of the transmitter circuit is shown in Figure 9.
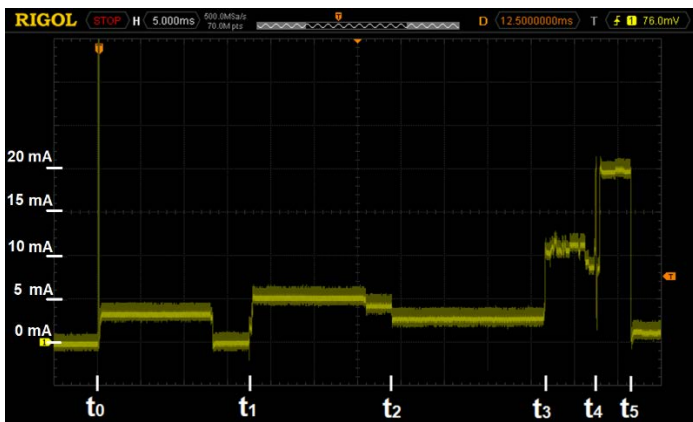
Figure 9: Current consumption of the RF transmitter with the time scale according to the Table 2

This measurement was realized by measuring the voltage drop over a shunt resistor, placed in the supply line of the transmitter circuit.

The important time intervals of the data transmission are shown in Table 2. The calculated energy value is 0.7 mJ for start-up of the transmitter module and the first transmitted data packet, and 0.3 mJ for following data packets. Between transmission of data packets, the transmitter IC goes in a power down mode, reducing the current consumption down to few microamperes.

Table 2: Measured Current consumption of RF transmitter

Time interval	Task	Current, mA	Power, mW
$t_0 - t_1$	Power-on	3.5	7.0
$t_1 - t_2$	Initialize	5.0	10.0
$t_2 - t_3$	Compose data package	4.0	8.0
$t_3 - t_4$	Initialize transmission	10.0	20.0
$t_4 - t_5$	Data transmission	19.0	38.0

Sending of multiple sequential packages reduces the probability of a data loss (e.g., by interfering of a single packet with electromagnetic noise spikes). However, the multiple data packets increase the energy budget of the sensor system. In the first revision of the sensor board, the number of repetitions counts was 4 to 6 data packages, depending on the charging state of the super capacitor.

The implemented randomized time intervals between the individual packages, shown in Figure 10, should additionally increase the safety of the data transmission.

In the second revision of the sensor board a number of 2 to 3 transmitted packets was implemented and successfully tested, which helps to reduce the required energy budget of the sensor circuit almost by 50% (considering the power consumption of the MCU, not affected by reducing the number of transmitted packets).
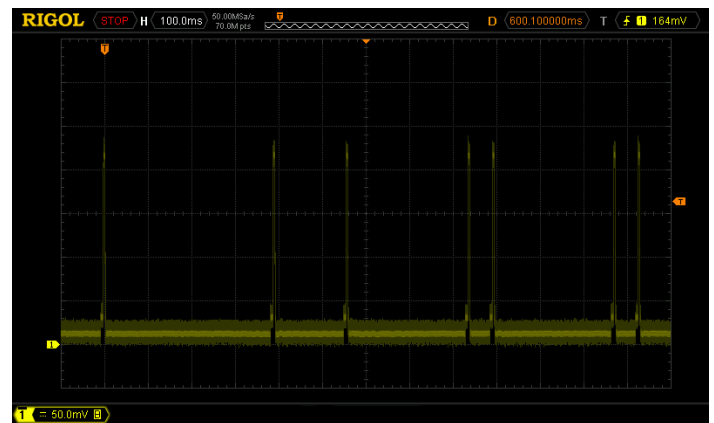
Figure 10: Multiple data packets with randomized time intervals

To keep the time of a single data transmission as short as possible, the length of the data package was also decreased to a minimum: the package consists out of two bytes preamble, four bytes of payload and additional two bytes for the control sum (CRC-16), leading to a total length of the data package of only 8 bytes (Table 3). Since the implemented sensors operate as simple water detectors with transmitting only their unique ID number, the payload bytes are currently holding the four bytes of the sensor ID. For other sensor implementations for measuring and transmitting of additional parameters the packet length needs to be adjust correspondingly.

Table 3: Transmitted data packet

Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7	Byte 8
Preamble	Preamble	ID-1	ID-2	ID-3	ID-4	CRC-16	CRC-16

6. Receiver board

To receive and evaluate the sensor data, a receiver board based on the Si4355 866 MHz RF receiver module by Silicon Labs [16] was developed. The block diagram of the receiver board is shown in Figure 11.
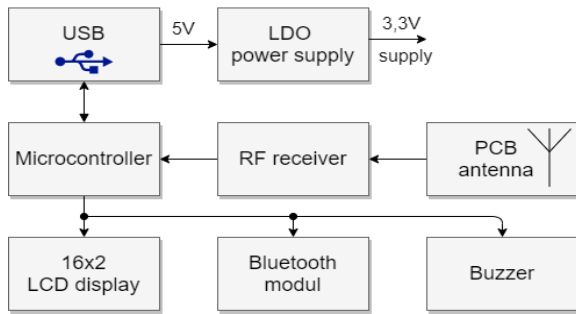
Figure 11: Block diagram of the receiver board [1]

The received data packets are decoded by the RF receiver module, the payload data are processed by the MCU and send via USB interface to the external system. The received data packets are also shown on the built-in LCD display, additionally the buzzer tone signalizes the receipt of a new packet. The received packets can also be transmitted wirelessly via Bluetooth interface to a tablet or a mobile phone. The developed receiver board is powered via USB interface from the host device.

The photo of the developed receiver board is depicted in Figure 12. The USB interface is located on the left side of the board, the receiver IC and the RF antenna are placed on the right side of the main PCB, separated from the main part of the electrical circuit and having their own ground area. The pin-header on the top side of the PCB is used to connect the external LCD display, in the middle the FT232RL USB to UART interface by FTDI Chip [17] and the power supply circuit are located.
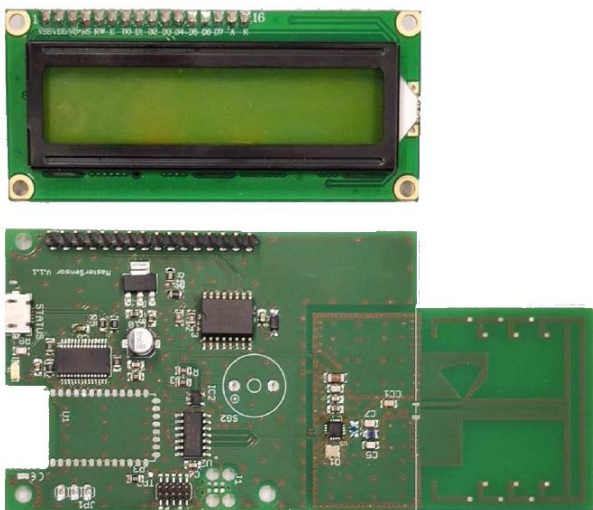
Figure 12: Receiver board with external LCD display module

The receiver board is controlled by an ATtiny1614 microcontroller by Microchip [18]. To address all the implemented interfaces an additional 16-bit port expander IC is used [19]. The Bluetooth module [20] is placed on the bottom left side of the PCB (not assembled in the current revision of the board).

The connected and powered receiver board is shown in Figure 13. The LCD display is mounted above the receiver circuit, covering most electronic components. The connected USB cable also powers the receiver board from the host PC. The LCD display shows the programmed firmware revision in the first line and the last received data packet (sensor ID number) in the second line.

Figure 13: Assembled and powered receiver board showing the last received data packet on the LCD display

The measured transmission range was up to 150 meters outdoors and 30 to 40 meters indoors. According to the orientation of transmitter and receiver antennas, the transmission range may deviate by 20-30%.

A bidirectional communication between sensors and receiver board has not been implemented in the current revision of the sensor system, caused by the small energy budget and limited by the size of the sensor board.
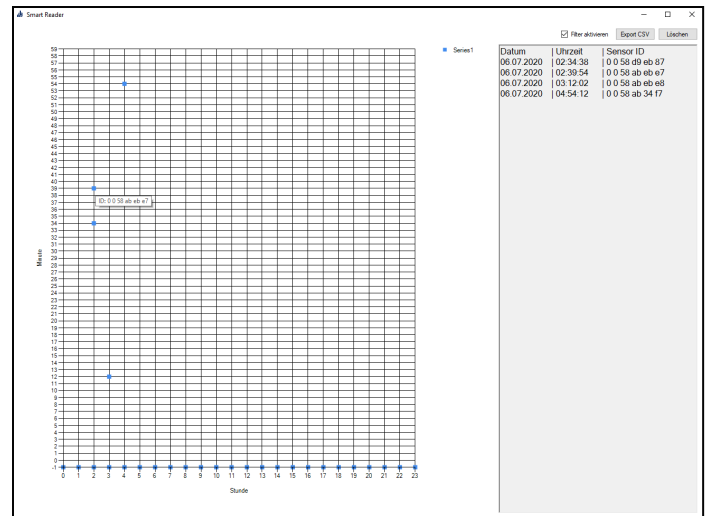
Figure 14: Developed sensor mapping software for time-based representation of activated sensors

7. Developed PC software

For some practical applications (e.g., construction monitoring and detection of water leakages), a graphical depiction of sensor mapping with a time-based representation of activated sensors may help to establish the starting point for water leakages inside the

monitored building. An example of the developed software with graphical representation of activated sensors is shown in Figure 14.

On the left side of the display the last activated sensors are shown with the blue points. The list of activated sensors including the sensor ID with a corresponding time stamp is shown on the right side of the window. The sensor mapping feature with graphical representation of the physical sensor placement inside the monitored area is not implemented in the current revision of the software; it will be developed in the next revision.

8. Conclusions

The developed system for powering of wireless sensor systems uses a build-in galvanic cell. It may be used in different scenarios, where the use of conventional batteries is not desired or not possible. Another advantage compared to battery powered systems is a significantly longer storage time of the galvanic cell.

After initiating the galvanic cell, that means both cell electrodes get in contact with water, the lifetime of the cell will be limited by corrosion of the negative (zinc) electrode. The confirmed lifetime can take some days to some weeks, depending on the application and the type of used electrolyte. The confirmed lifetime makes possible the implementation of different scenarios, like water detectors, one-way sensors for environment parameter control etc.

The first revision of the sensor PCB with a size of 27 x 100 mm assembled at Paderborn University was already tested in different practical scenarios.

The second revision of the energy supply circuit with implemented MPP tracking has a higher efficiency in combination with a smaller PCB size of 27 x 75 mm. It also allows a reduction of the resulting board dimension by 26%. The total components cost in the second revision were reduced by about 30%.

The energy supply circuit in combination with the galvanic cell generates enough energy for sending data packages every 3-5 minutes. By increasing the size of the electrodes, more complex sensor systems with higher energy budgets can be implemented.

The developed receiver board receives and evaluates the transmitted sensor data. The received data packages are also shown on the built-in LCD display and forwarded via USB interface to the host-PC. They can be transmitted wirelessly via Bluetooth interface to a tablet or a mobile phone.

The corresponding PC software with the time-based representation of activated sensors in a graphical form including the list of activated sensors with corresponding time stamp was also developed.

9. Outlook

Further optimisation of the sensor board could focus on the transmitter circuit including the transmitter antenna, which currently occupies about 30% of the board size. Another antenna topology with a smaller size should be investigated and implemented.

Furthermore, the operation characteristics of the developed cell under varying environment conditions (humidity and temperature) should be investigated.

Additionally, methods to increase the generated output power could be explored, e. g. by coating the electrodes with material influencing the conductivity of the water and therefore increasing the output power (e.g., natrium chloride, or citron acid).

The developed PC software can be improved by some additional features like a sensor mapping tool with a graphical representation of the physical sensor placement inside the monitored area. The option of combining the registered sensors to different sensor groups (e. g. for monitoring of different objects) is also possible.

References

[1] D. Petrov, U. Hilleringmann, "Water-based primary cell for powering of wireless sensors", IEEE Sensors 2020

[2] G. Tuna, V.C. Gungor, „Energy harvesting and battery technologies for powering wireless sensor networks", Industrial Wireless Sensor Networks, 2016

[3] F. Mazunga, A Nechibvute, "Ultra-low power techniques in energy harvesting wireless sensor networks: Recent advances and issues", Scientific African, Volume 11, 2021

[4] H. F. Rezaei, A. Kruger and C. Just, "An energy harvesting scheme for underwater sensor applications", Electro/Information Technology (EIT), 2012 IEEE International Conference on. IEEE, 2012

[5] R. A. Huggins, Advanced batteries: materials science aspects, Springer, 2009

[6] J. Marple, Primary Battery Design, Encyclopedia of Applied Electrochemistry. Springer, New York, 2014

[7] D. Linden, "Handbook of batteries", Fuel and energy abstracts, 1995

[8] K. B. Lee, L.Lin, "Electrolyte-based on-demand and disposable microbattery", Journal of Microelectromechanical Systems, **12**, 1-8, 2003.

[9] Multi-CB, Printed circuit board surface finish, [Online]. Available: https://www.multi-circuit-boards.eu/en/pcb-design-aid/surface/surfaces.html. Accessed on: Apr. 25, 2021.

[10] Texas Instruments, TPS6120x Low Input Voltage Synchronous Boost Converter, [Online]. Available: http://www.ti.com/lit/ds/symlink/tps61200.pdf. Accessed on: Apr. 25, 2021.

[11] AVX, SuperCapacitors, BZ Series, [Online]. Available: http://www.avx.com/products/supercapacitors/bestcapreg-bz-series. Accessed on: Apr. 25, 2021.

[12] RICOH, Low Voltage Detector, [Online]. Available: https://www.mouser.de/datasheet/2/792/r3116-ea-1770467.pdf. Accessed on: Apr. 25, 2021.

[13] Analog Devices, Ultralow Power Energy Harvester PMU with MPPT and Charge Management, [Online]. Available: https://www.analog.com/en/products/adp5091.html. Accessed on: Apr. 25, 2021.

[14] Silicon Labs, Ultra Low Power Microcontrollers, [Online]. Available: https://www.silabs.com/mcu/8-bit/efm8-sleepy-bee/device.efm8sb10f8g-qfn20. Accessed on: Apr. 25, 2021.

[15] Silicon Labs, Crystal-less RF Transmitter, [Online]. Available: https://www.silabs.com/documents/public/data-sheets/Si4012.pdf. Accessed on: Apr. 25, 2021.

[16] Silicon Labs, FSK Sub-GHz receiver, [Online]. Available: https://www.silabs.com/documents/public/data-sheets/Si4355.pdf. Accessed on: Apr. 25, 2021.

[17] FTDI Chip, USB Full Speed to Serial UART IC, [Online]. Available: https://ftdichip.com/products/ft232rl. Accessed on: Apr. 25, 2021

[18] Microchip, 8-bit AVR microcontroller, [Online]. Available: https://www.microchip.com/wwwproducts/en/ATTINY1614. Accessed on: Apr. 25, 2021

[19] NXP Semiconductors, Remote 8-bit I/O expander [Online]. Available: https://www.nxp.com/part/PCF8574T#/. Accessed on: Apr. 25, 2021

[20] DSD TECH Bluetooth 4.0 module, [Online]. Available: http://www.dsdtech-global.com/2017/08/hm-10.html. Accessed on: Apr. 25, 2021

Discover DaVinci: Blockchain, Art and New Ways of Digital Learning

Marko Suvajdzic[*,1], Dragana Stojanovic[2]

[1]*Digital Worlds Institute, University of Florida*, Gainesville, *FL, 32601, USA*

[2]*Faculty of Media and Communication*, *Singidunum University, Belgrade, 101801, Serbia*

A B S T R A C T

Discover DaVinci is a novel augmented reality system that incorporates blockchain technology with experiential learning to engage participants in an interactive discovery of Leonardo da Vinci's oeuvre. The software was created by Marko Suvajdzic, first author of this paper, and it was produced at the University of Florida Digital Worlds Institute. In the true spirit of this "Renaissance man", Discover DaVinci explores new ideas and technologies "ahead of their time", opening up questions about usage of blockchain system in the domain of art and technology. This paper discusses some of these questions, such as relation of art and technology, usefulness of blockchain system for digital art, and new materiality of art in digital and informational age. Proposed work of this manuscript is to present the field of digital learning through a general review and more specifically through a prism of Discover DaVinci project created by Digital worlds Insititute at the University of Florida.

1. Introduction: Art and Technology, the Unseparable Pair

Speaking about art in contemporary times, together with a still present idea about its alleged autonomy [1], we may tend to understand it as framed within a particular institutional, academic or actual scene [2]. If we concentrate on its manifestations on the stage, in galleries or museums, where we experience it as a certain kind of a screen given to the audience, it is easy to completely miss its deep, continual and historical relation to technology and experimentation. In the other words, as traditional art audience, we often regard art as an object to be watched, listened to, assessed, valued or sold/bought. In such a constellation, the art audience is left in a quite passive state, although art, as a communicational model of encoding and decoding (note the obvious relation to technology and programming techniques!), offers its audience much more [3].

One of the dimensions of art production and placement that is explored in the beginning of 21st century is certainly connected to contemporary communicational, digital and web technologies. It may seem as a new field to be questioned and explored, but the relation of art and technology goes a lot more back into history. It is enough to recall a philosophical term technē, which was used to refer to a making or a doing, but it was also used to designate a kind of an episteme, and it was a name for knowledge in the widest

sense [4], [5]. What is more important here is that it was used for any kind of artistic practice that involved making-doing, as well as gaining knowledge through technological experimentation. It was particularly used for crafts and mechanical art, but Larry Shiner argues that it might not have been strictly about art as such, but much more about human ability to make and perform, or, more directly, to use technology to improve the possibilities of making-doing, of understanding art, and of presenting it to others in a concrete historical context [6].

It seems that the current context in which matters of art and art production are just being developed and innovated belongs to digital reality, or, even more forward, to a technological and artificial intelligence (AI) environment. In these digital circumstances, the focus is transferred from an art-object to an art-information, and from human to machine agency, but it does not mean excluding human factor from the art worlds, institutions and economies – on the contrary. We may say that by introducing machine intelligence and transactional systems, their human users are getting chance to consider these changes as motivational and developmental factor not only in the field of technology, but also in the fields of social sciences and art.

One of the technological platforms, or, better to say, a technological system that is found right at the crosspoint of mathematics, sociology, economy, law and politics is blockchain

[*]Corresponding Author: Marko Suvajdzic, Email: marko@digitalworlds.ufl.edu

system. As this paper will propose and as we will show in further discussion using Discover DaVinci project as an example, the usefulness of blockchain system for the world of art is, although contested, quite prominent and more and more explored – not only within new forms of art production, presentation, protection, marketing and trading, but also within new forms of learning about art in contemporary times. Researches in this field might prove important not only for technology-oriented studies, but also for the education and learning studies.

2. Blockchain and Art: Is It All About the Money?

Looking back to beginnings of blockchain system application, it is obvious that it was created as an "electronic payment system based on cryptographic proof instead of trust…" [7]. This means that blockchain was first and foremost meant to be a more secured and complex system for monetary transactions. This system is based on cryptographically carved currencies of which the first one was Bitcoin, after which it was followed by many more. Blockchain system is thus a distributed database whose security, paradoxically enough, lies in an increase of parties included. So, instead of a trust in one mediator as a representative of a desired transaction, there are hundreds and thousands of nodes, mutually connected, that carry the information about the transaction. The security factor is ensured by peer-to-peer system (P2P), where the server is shared between all the computers and included parties – not only the parties involved directly in transaction, but also all the parties that embody this complex system of humans, codes and computational nodes. What is interesting here is that by increasing the number of involved participants we decrease the risk of a problematic transaction. This is a very specific cooperation of machines and humans, working together on the task of transferring information and ensuring that it is not copied or doubled illegally. The information is condensed into a block, and blocks are packed into chains (thus the name blockchain). This is done through a hash function, which is a form of cryptography that functions as a suture between blocks, and it is placed in such a way that it guarantees that the information within a block cannot be changed internally or externally without the complete change being detectable in all the chain nodes. The only, though theoretical way of manipulating blockchain system lies in an option that at least 51% of elements in the chain decide to manipulate the chain in the same way, and in an exact same time. From here we conclude that the more peers in the blockchain network, the less possibility of manipulating it in any kind of way. On the other side, some downsides of blockchain system include work efficiency problems and high resource consumption due to the process of mining, and that is currently in the focus of research of all those who work on blockchain system inclusion in any of the fields it may find its usage [8].

But is the blockchain system really all about money, and why is its decentralized mechanics so important for the world of art, and, more so, for the process of learning about art? First of all, blockchain system is a system – which means, a specifically operating platform of communication. In this way, any value we put in it, money, art or otherwise, gets included into a specific, mechanical, human-machine built world, a network-like system in itself. Although some can argue that blockchain system is well outside of traditional institutions, and those arguments are given in a very strong way [9], it seems that blockchain system in itself forms a specific set of communicational and transactional rules

that inscribes itself in the digital entities it works with. In the sense of previous understanding of art institutions and art worlds [10], it definitely skips the logic of material art exhibition and valuation, but it creates a new world for art existence and presentation, or, more precisely, it adds a new, digital, immaterial layer to previously known logics of art creation, trading and consumption. Thus we cannot really speak about the new, potentially foreign system to the world of art, but of a new technological option for art development and market.

Speaking of that, blockchain system is not only about the art market, although it is a big part of innovation it brings. Of all the other benefits it may bring, besides better possibilities to protect digital artworks and authors' rights, blockchain system offers many options for solving the issues of copying digital works [11]. It also helps promoting yet unestablished authors, and it can resolve the issues of returning material artworks to the countries where they originated from, or to the individuals that have rights to its ownership – in the same time not disturbing work of museums and galleries that currently exhibit these works. Of course, and on the top of it all, blockchain system offers new possibilities of learning about art, especially through gamification strategies. But, before we delve deeper into a concrete example of this possibility that will be described through the analysis of Discover DaVinci project of the University of Florida Digital Worlds Institute, there is still enough space to think through another perspective that the blockchain system brings to the world of art, and that is the issue of (a new) materiality of art in the domain of digital art.

3. Digital Age and New Materiality of Art

In her text "Collections Management on the Blockchain: A return to the Principles of the Museum" Helen Kaplinsky mentions posthuman museums in the context of human-machine, or human-algorithm dialogical or network interaction [12]. Mentioning posthuman museums (and maybe we can extend this into proposing a thought of posthuman classrooms?), she poses the question which can be condensed into the following: do we still need bodies in the age of digital and blockchain platforms prevalence? And, how it will influence the world of art? Or, more broadly said, do we still need materiality of the artworld? As Martin Zeilinger says, "Digital art can thus help us rethink questions related both to the materiality of the digital and to the immateriality of the art object as an abstract container of value" [13]. But how is digital art linked to a material, and an art object to immaterial in this quotation? It seems that a digital contemporaneity, and, especially blockchain technology is contributing further to the complexity of this issue, becoming a relevant and interesting topic not only for art itself, but also for the theoretical discourses around art.

Looking briefly into the nature of digital art, we may easily conclude that digital art is a perfect example of an ephemeral, immaterial art practice – so much that it is actually hard to pinpoint the storing "place", or even a "source" of a digital artwork – be it because of its many copies and shares, or due to its (non)physical, digital qualities. It is important to note that digital art is not the first attempt to immaterialize the artwork and

art practices itself – that attempt can be traced back to conceptual art, as an institutional critique towards capitalist ideologies, as a way of disintegration of traditional art norms, and also as an experimentational practice and exploration of art in a dematerialized, "post-aesthetic" condition [14]. It was in 1973 when Lucy Lippard hoped that "'conceptual art' would be able to avoid the general commercialization… of modernism" [15]. Since then, it seems, we can trace a specific hope in the act of dematerializing art, in the means that in that way it is thought to be able to escape commodification logic of the market, and avoid becoming a high-efficiency container for commercial value". In the age of digital art, the first example of this sort is certainly net art, and for a couple of decades digital art, as a continual line drawn from net art subverting practices, was regarded as an art that is able to contradict the logic of institutionalized and monetized art world. Furthermore, it was regarded as "temporary autonomous zone with no tolerance for institutional dogma and the ideologically suspect economic value systems propagated by institutional bureaucracies of the art world". Perhaps that is why there is still so much resistance towards blockchain technology within the scope of digital art practices, as well as among digital artists and digital art theoreticians. These resistance voices usually point to the dangers of hypercommodification of art; also, they are worried about the possibility of creating a technologically dependent society, or of strengthening the capitalist logic in the domain of art [16]-[20].

To be able to understand relations between digital art and blockchain technology, we have to go back to possible definitions of digital art. According to Christiane Paul, digital art is "art that is digital-born, computable art that is created, stored and distributed via digital technologies and uses the features of these technologies as a medium" [21]. Similarly enough, Hui defines digital objects as appearing to human users "as colourful and visible beings. At the level of programming they are text files; further down the operating system they are binary codes; finally, at the level of circuit boards they are nothing but signals generated by the values of voltage and the operation of logic gates" [22]. Both the authors stress the informational element of the digital object, or digital art object that needs digital technology to be able to exist, or to present itself to the audience. This certainly opens up the path for thinking about contemporary digital technologies, together with blockchain technologies, as relevant factors of enhancing the digital media. In the end, in order to serve and inspire digital art and artists, blockchain system should be both medium and subject of artistic practices, both conventional and critical/subversive. In order for this to happen, artists should have a voice in how this technology might be applied to the field they act in, but to be able to gain this voice, they would have to say *yes* to technology, and to research the ways it could work for their ideas, and not against them. All the different artworks of emerging blockchain artists already show this, and they successfully overcome duality between human and machine factor, as well as between materiality and immateriality of the artwork (see Figure 1 as an example, also [23]).

Figure 1: Primavera De Filippi, Plantoid

As innovations in art are always linked to innovations in technology and vice versa, and as making of art, in its quality of *technē* is always in the same time a knowledge creation, the authors of this text will propose a model of technology research in art, or, differently said, of art research in technology. One of the ways to do it is to work on the intersections of technology, art and learning, and to create gamification strategies of learning about art through blockchain system. One of such examples is *Discover DaVinci* project, an interactive system produced at the University of Florida Digital Worlds Institute.

4. Introduction to "Discover DaVinci" Project

In order to illustrate the emerging potential at the intersection of art and blockchain, we present a case study of a new interactive system produced at the University of Florida Digital Worlds Institute. This software was created by Marko Suvajdžić, associate professor and associate director of Digital Worlds Institute. Suvajdžić's experience includes various digital startups and educational projects involving artificial Intelligence-intensive video game titles for major corporate clients. Suvajdžić also co-founded 5 of his own startups.

The technologies of mobile computing, augmented reality (AR), and blockchain are starting to merge, creating new opportunities and scenarios to interact with our environment. In AR we can look at virtual objects superimposed within a real environment and resize them, rotate them, explore and interact with them on multiple levels. With the combination of AR and blockchain, we can create a system capable of keeping track of

digital assets located virtually in 3D space (i.e., spatial computing). The global scale of blockchain and related technologies heightens the potential for trade and digital distribution with a fully automated and trusted way to keep track of their creations without a "middle-man".

Discover DaVinci is a novel educational tool that teaches concepts of blockchain technology through an augmented reality experiential learning game.

5. Discover DaVinci Project Partnerships

This project was developed in collaboration with several units from the University of Florida and industry partners:

- Digital Arts & Sciences Faculty (Computer Science and Digital Worlds Institute)
- Digital Worlds Studios' Artists and Programmers
- Gator Blockchain Club (gatorblockchainclub.com) – Student-run blockchain club at the University of Florida
- Center for Innovation and Entrepreneurship (College of Business)
- Creative Campus Committee at the University of Florida

Industry Partners:

- DLUX, decentralized content network (dlux.io)
- Steem (steem.com), and Steemit (steemit.com)
- A-Frame, web VR platform (aframe.io).

6. Research Study: "Discover DaVinci"

Discover DaVinci (Fig.2) utilizes the format of a digital, collectible trading & drafting card game with AR elements on the STEEM blockchain. Although each player "owns" their cards, all transactions are public. This is enabled by the blockchain system itself, that defines private and public key for every user. The private key always stays private and cannot be recovered, and the public key is visible to the others. In this way blockchain system guarantees all the participants both their privacy, and in the same time public visibility of every change in the chain, being enacted by purchases and transfers in ownership.

Figure 2: Discover DaVinci game interface, University of Florida (2019)

Discover DaVinci collectible cards are all unique tokens, owned by the player - a digital asset registered to the player's account. The aim is to draw new question cards daily, answer the questions about Leonardo DaVinci, collect the special AR invention cards, and ultimately submit the accumulated card collection into a drawing for prizes. The app was developed to honor the 500th anniversary of Leonardo DaVinci by promoting new and innovative technologies. In the same time, this project is an homage to Leonardo DaVinci's fascination with art and technology, which is an *oeuvre* of the artist of the Renaissance period, but it may also be a quality of artists *maneuvering* their own artwork pathways in the age of blockchain technology development.

By using a novel combination blockchain, and spatial computing/AR technologies, the *Discover DaVinci* mobile app allows users to explore holographic images of DaVinci's devices. By combining AR with blockchain technology, we have created a system capable of assigning unique values and user comments to these digital assets, thus providing a personal sense of ownership, the potential for trade, and a fully automated and trusted way to keep track of these digital artifacts (e.g. collected 3D objects in AR space). Finding the artifacts themselves is an act of discovery, further enhanced by subsequent investigation, collection, and sharing of the artifacts across a diverse social media ecosystem.

The sense of ownership, and not only accumulating of goods is provided by the technology of non-fungible tokens (NFTs), which is especially valuable in the field of art. In the same time, using NFT in tokenization of an artwork or a collectible makes it possible to distinguish the art token from any other cryptovalue, or a cryptocurrency, which are defined as fungible tokens. In a way, this gives us an answer to a previously posed question – is blockchain based art only about the money? Following the logic of non-fungible tokens, it is not. NFTs are not interchangeable, nor convertible in any other value. They possess their own unique value – just as it is the case with material, traditional art. In this way we can see that blockchain system can be regarded as the next step of efficiency and inspiration material within the already existing art forms, art market and art institutions. Of course, as there is always a room for critique and resistance in traditional art world, it is the same in technological world. That is why it is important for the artists and art/art theory teachers to delve deeper into blockchain technology research, and to tailor new media to their own expressions and needs – just as it has always been, as it is shown in the case of Leonardo DaVinci.

Back to *Discover DaVinci* project, the blockchain system was implemented using the STEEM API, which handled the writing of new user accounts and the reward transactions to the Blockchain (Fig. 3 bottom). The front end was developed as a web app that rendered the various 2D and 3D/AR components of the game (Fig. 3 top right). Finally, the authentication and account creation was handled using a third-party service, called STEEM Ninja (Fig. 3 top left).
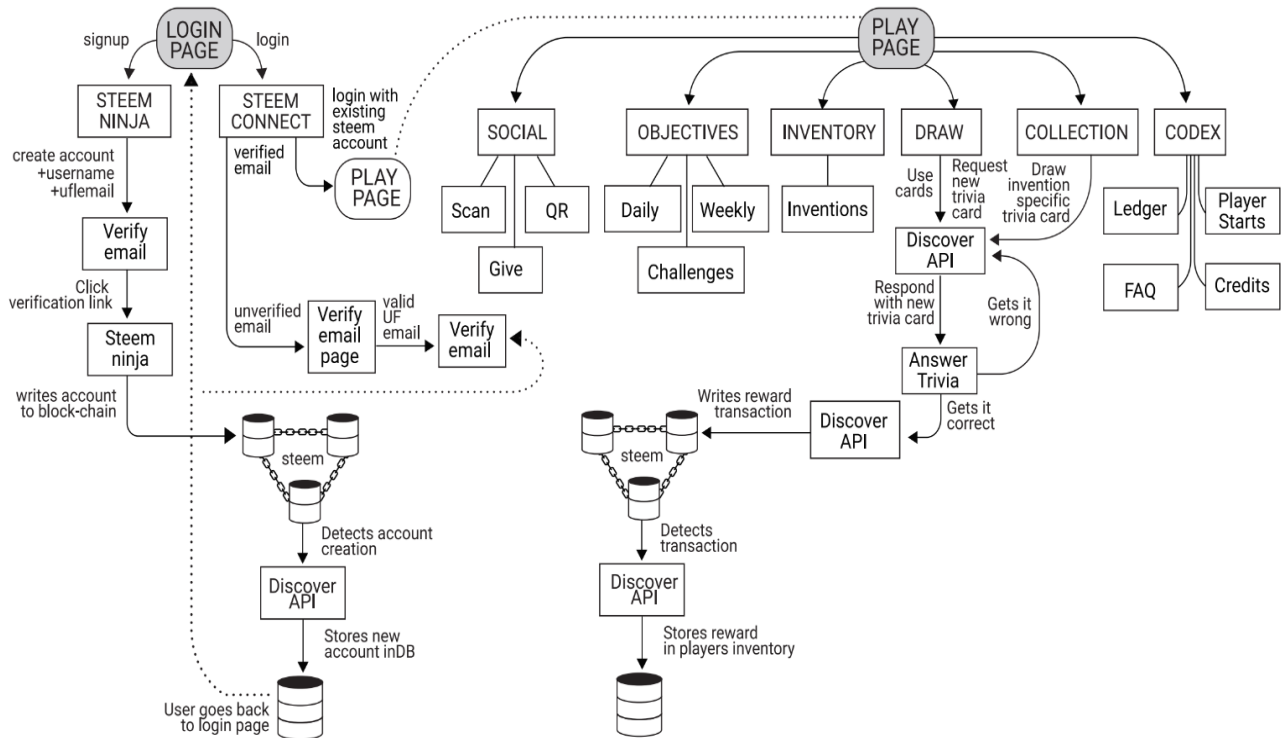
Figure 3: Discover DaVinci-System Architecture illustration, University of Florida (2019)

7. Discussion

We launched Discover DaVinci app to the initial audience of University of Florida students and faculty in November of 2019. Over the period of 60 days, we engaged ~150 participants who played the game, and left their feedback on the project's discussion board.

The player feedback has been analyzed, and the subsequent versions in the "Discover" series are currently being evaluated. The project is intended to continue as a platform for gamified education that promotes participant interaction, and supports focused learning on a topic of choice.

In this review paper with the main example of Discover DaVinci App, the project conducted by the team at the University of Florida, Digital Worlds Institute, the authors wanted to refer to Discover DaVinci application and its potential within the fields of gamification and learning concepts. Certainly, there is still a space for further tests to be done in the future, so we could effectively reveal the impact of such an application on learning methods and contemporary educational approaches involving blockchain technology. However, at this point, the authors would like to propose a formal conclusion, stating that these kinds of applications offer effective possibilities for understanding teaching and learning processes within generation that adopt knowledge at the present moment. What has been noticed is certainly an ease of a new technology adoption in the entertaining educational setting, and that is the point from which we can continue to think about the intersection of technology and education.

Discover DaVinci is an innovative app that uses blockchain technology as a tool that utilizes gamification of education, specifically focused on presenting life and works of Leonardo DaVinci. The app has been custom created for this purpose by the team at the Digital Worlds Institute at the University of Florida, and as such is not directly comparable with general educational tools on the market. However, it is used to illustrate how blockchain technology can be introduced to new users through gamified content.

We believe that this novel confluence of art, technology, and digital communications represents a tangible example of how artists and cultural researchers can leverage the potential of diverse digital resources to support both creative development and cultural inquiry. The Discover DaVinci app allows a diverse player community to utilize their existing mobile devices and participate in acknowledging DaVinci's life work while exploring a new paradigm of digital convergence. The outcomes included an enhanced cross-campus appreciation of personal creativity, social and cultural engagement, and a working framework that exemplifies how emerging technologies can serve esthetic ends while also ensuring fair use and exchange of intellectual property in the digital age.

Acknowledgements

References

[1] R. Cronk, "The Pseudo-Autonomy of Art," Art on the Rebound: A Collection of Essays on Art and Culture, https://www.westland.net/venice/art/cronk/autonomy.htm, 1996, Retrieved 23 March 2021.

[2] G. Dickie, Aesthetics, An Introduction, Pegasus, 1971.

[3] R. Rojcewicz, The Gods and Technology: A Reading of Heidegger, State University of New York Press, 2006.

[4] D. Carpi, Bioethics and Biolaw through Literature, Walter de Gruyter, 2011.

[5] L. Shiner, The Invention of Art, University of Chicago Press, 2001.

[6] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," www.bitcoin.org, Retrieved 22 August 2020.

[7] N. Chalaemwongwan, W. Kurutach, "State of the Art and Challenges Facing Consensus Protocols and Blockchain," ICOIN 2018, 957-962, 2018. doi: 10.1109/ICOIN.2018.8343266.

[8] S. Ren, "The Application of Blockchain and its Advantage to the Field of Art," 2nd International Symposium on Big Data and Applied Statistics, Journal of Physics: Conference Series, IOP Publishing, 2019.

[9] R. Myers, "Blockchain Poetics," Catlow R., Garret M., Jones N, Skinner S. (Eds.), Artists Re:Thinking the Blockchain, Torque Editions & Furtherfield, 239-250, 2017.

[10] A. Danto, "Symposium: The work of Art, the Artworld," The Journal of Philosophy 61(19), 571-584, 1964. doi: 10.2307/2022937.

[11] R. O'Dwyer, "Limited Edition: Producing artificial scarcity for digital at on the blockchain and its implications for the cultural industries," Convergence: The International Journal of Research into New Media Technologies, Sage, 1-21, 2018, doi: 10.1177/1354856518795097.

[12] H. Kaplinsky, "Collections Management on the Blockchain: A return to the Principles of the Museum," Catlow R., Garret M., Jones N, Skinner S. (Eds.), Artists Re:Thinking the Blockchain. Torque Editions & Furtherfield, 261-274, 2017.

[13] M. Zeilinger, "Everything you've always wanted to know about the Blockchain*(*but you were afraid to ask Mel Ramsden)," Catlow R., Garret M., Jones N, Skinner S. (Eds.), Artists Re:Thinking the Blockchain. Torque Editions & Furtherfield, 287-295, 2017.

[14] N. Jones, S. Skinner, "A Quasi Proto Preface," Catlow R., Garret M., Jones N, Skinner S. (Eds.), Artists Re:Thinking the Blockchain, Torque Editions & Furtherfield, 9-19, 2017.

[15] A. Alberro, B. Stimson, Conceptual Art: A Critical Anthology, MIT Press, 1999.

[16] M. Zeilinger, "Digital Art as 'Monetized Graphics': enforcing intellectual property on the blockchain," Philos.Technol, 2016, doi: 10.1007/sl3347-016-0243-1.

[17] L. Lessig, Free Culture, Penguin, 2004.

[18] A. Moore, K. Himma, "Intellectual Property," Stanford Encyclopedia of Philosophy, http://plato-stanford.edu/entries/intellectual-property/, Retrieved 15 November 2016.

[19] J. Boyle, The Public Domain: Enclosing the Commons of the mind, Yale University Press, 2008.

[20] R. J. Coombe, The Cultural Life of Intellectual Properties: Autorship, Appropriation, and the Law, Duke University Press, 1998.

[21] C. Paul, A Companion to Digital Art, Wiley Blackwell Publishers, 2016.

[22] Y. Hui, "What is a Digital Object?," Metaphilosophy 43(4), 380-395, 2012. doi: 10.1111/j.1467-9973.2012.01761.x.

[23] M. Suvajdzic, D. Stojanovic, J. Appelbaum, "Blockchain Art and Blockchain Facilitated Art Economy: Two Ways in Which Art and Blockchain Collide,"2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON), 2019, doi: 10.1109/TIMES-iCON47539.2019.9024403.

Comparative Analysis and Modern Applications of PoW, PoS, PPoS Blockchain Consensus Mechanisms and New Distributed Ledger Technologies

Caglar Arslan[*], Selen Sipahioğlu, Emre Şafak, Mesut Gözütok, Tacettin Köprülü

HAVELSAN Inc., Ankara, 06510, Turkey

ARTICLE INFO

ABSTRACT

Central authorities have registered economic transactions based on trust since the dawn of time. With the advent of paper, methods for documenting transactions and records became more difficult and detailed. With the widespread use of computers in recent years, digital recording has made record keeping easier and since then every technical advancement has made data recording more convenient and faster. Transactions have accelerated as technology has advanced, but the maintenance of records has remained under the jurisdiction of the authorities. The use of distributed ledger systems allows data to be released from the control of central authorities. A distributed ledger is a peer-to-peer network in which non-centralized data is shared by participants. The Bitcoin payment system and cryptocurrency, which arose after the 2008 financial crisis, was the first distributed ledger application. Central authorities' poor economic decisions aided in the rise of Bitcoin. The aim of Bitcoin is to make the current financial systems liberal from the influence of central authorities. With the rise in popularity of Bitcoin, the blockchain technology that underpins it has begun to garner interest. While blockchain technology was initially associated with the financial sector due to Bitcoin, research into its use in other sectors such as supply chain, records management, electoral systems, notary services, file systems, energy and artificial intelligence has begun. Several Blockchain infrastructures have been built to allow the use of blockchain technology in a variety of industries. However, it has been revealed in practical approaches that blockchain technology has limitations in terms of speed and scalability. As a result, new distributed ledger technologies with increased speed and scalability have been established. Hashgraph, Tangle, Tempo, Holochain are examples of newly developed distributed ledger technologies. Different influential features distinguish new generation distributed ledger technologies from the conventional Blockchain methods, which can yield into several and practical modern applications. In this study, Blockchain and new generation Distributed Ledger Technologies are compared and possible future applications are outlined.

1. Introduction

Blockchain is a distributed database technology in which a database is stored and maintained without the use of a central server across an entire network or among similar parties. No one on the network, inside or outside, has the right to modify or remove recorded data. Transparency is also established by disseminating the data to all chain participants by validating the transmission data via a consensus algorithm [1].

A consensus algorithm is a fault-tolerant mechanism used in computer and blockchain systems to achieve the required agreement among distributed processes or multi-agent systems on a single data value or a single network state.

A Blockchain infrastructure, as a type of distributed system composed of multiple entities, requires a variant of consensus algorithm in order to operate. The characteristics of any kind of Blockchain are determined by its internal consensus architecture.

A Blockchain, which is defined by a consensus mechanism, enables us to have an immutable database of transaction record among the participating peers. A sample transaction flow is summarized in 'Figure 1' below in order to gain a better understanding of the underlying working process of Blockchain.

[*]Corresponding Author: Caglar Arslan, HAVELSAN HQ Mustafa Kemal Mah. 2120. Cad. No: 39 06510 Cankaya Ankara Turkey, +903122195787, Email: caglara@havelsan.com.tr
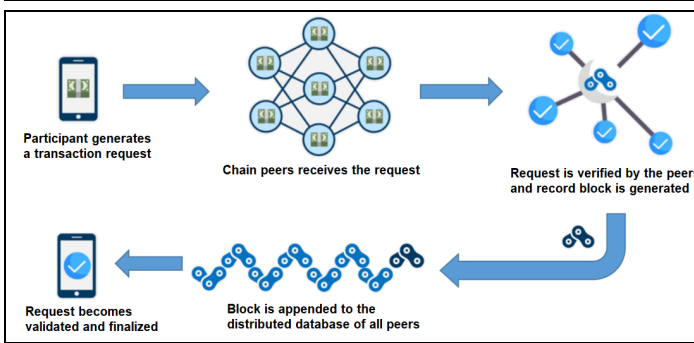
Figure 1: Sample Blockchain Transaction Flow

In summary, the information registry that exists within a group of nodes is made up of multiple blocks of data in which the data is stored and then consolidated within the network after a predefined validation process which is called the consensus.

The necessity of having an immutable distributed record database/table for a digital currency exchange system and applicability of Blockchain to this necessity gave rise to the first useful Blockchain application, which is widely known today as Bitcoin [2]. Since then the blockchain has been synonymous with financial transactions since the rise of Bitcoin, research is currently underway to see whether it can be applied to a variety of areas such as supply chain, records management, energy market, and artificial intelligence.

A variety of consensus algorithms as well as alternative distributed ledger technologies (DLTs), other than the classical Blockchain with the basic Proof-of-Work consensus, have been vastly proposed in recent years each with its own set of performance and security characteristics. It has to be noted that every application's requirements cannot be met by a single consensus algorithm [1]. Hence, it is critical to compare the consistent consensus algorithms as well as other distributed ledger technologies (DLTs) on a technical level in order to identify their strengths, weaknesses, and potential future applications.

Blockchain variants, which are defined by their accompanying consensus algorithms, can have problems regarding throughput, scalability, power consumption and storage usage which can handicap the potential usage areas besides digital currency exchanges/cryptocurrencies. As a result, consistent consensus variants and new distributed ledger technologies (DLTs) should be technically investigated in order to achieve more successful engineering designs of distributed, immutable, manageable, and scalable applications and thus allowing the widespread adoption of trustworthy distributed applications and products in everyday life.

In this study, Proof of Work (PoW), Proof of Stake (PoS), Pure Proof of Stake (PPoS) and new Distributed Ledger Technologies which are named as Hashgraph, Holochain, Tangle and Tempo are introduced and explained. After the introduction and explanations of the mentioned mechanisms, a comparison is made under the performance criteria of Throughput (transactions per second), Scalability (number of nodes), Energy Consumption (J per transaction) and Storage. After this comparison, possible modern applications that can be developed with Blockchain and DLTs are explained and then application-specific minimum performance requirements are given in order to outline the achievability of the

applications. Afterwards, suitability of the explained consensus mechanisms and DLTs to each application are explained.

2. Blockchain and Consistent Consensus Mechanisms

Blockchain is a decentralized distributed database technology that acts as a trust layer for the data exchange amongst participating peers.

The following key features define any type of Blockchain: Immutability, transparency, privacy, durability, decentralism, and security.

- Immutability: No participant can change the data stored in the blockchain.
- Transparency: Transparency is provided to all participants in the Blockchain by associating the transaction data with the concerning participants. All transactions are traceable.
- Privacy: The use of private and public keys is a crucial aspect of blockchain privacy. Asymmetric cryptography is used in blockchain systems to secure transactions between users.
- Durability: A Blockchain can operate indefinitely since a copy of the data is distributed across the participating nodes.
- Decentralism: The transfer of control and decision-making from a centralized entity (individual, organization, or group thereof) to a distributed network is referred to as decentralization in blockchain.
- Security: The structure of data created by blockchain technology has built-in security properties. Trust in transactions is ensured by the use of cryptographic principles, decentralization, and consensus. Each new block in a cryptographic chain connects to all the blocks before it in such a way that tampering is nearly impossible. A consensus mechanism validates and agrees on all transactions within the blocks, ensuring that each transaction is true and correct.

It has to be noted that throughout this study, Blockchain is referred as the two most popular consensus mechanisms namely the Proof of Work (PoW) and Proof of Stake (PoS) and a new consensus algorithm that is consistent with the key Blockchain features (i.e. decentralized, secure and scalable) and named as Pure Proof of Stake (PPoS).

2.1. Proof of Work (PoW) Consensus

The PoW consensus mechanism has been popularized by Bitcoin as one of the most widely used methods in blockchain (BTC). Miners and the electricity needed to perform the calculations that verify BTC transactions are the distinguishing features of PoW systems. Miners use computer hardware to run network nodes that use computational power to solve mathematical puzzles known as proofs of work. The miner who solves the puzzle first confirms the most recent block of transactions on the blockchain.

The successful miner then broadcasts the new block to all other nodes, who confirm its accuracy and add it to their copy of the blockchain, creating a verifiable record of data for the entire network. Consensus is represented by the verification process. A new block can only be added to the network after this data has been confirmed. For being the first to validate a new block of data and

add it to the PoW blockchain, miners receive newly minted cryptocurrency, the block reward (in the case of Bitcoin, they receive BTC).

Proof of Work blockchains aim to generate blocks at regular intervals — for example, Bitcoin generates one block every ten minutes. PoW networks are limited in speed and scale because of the energy-intensive proofing process. Furthermore, PoW networks are engineered to be more or less challenging depending on the network's computing capability.

Despite their speed and scalability restrictions, PoW blockchains have typically given better security while maintaining significant decentralization. Because PoW systems are distributed, it is extremely difficult for a malicious actor to take control of the blockchain by controlling the majority of computing power on the network. Typically, the hardware, power, and computing expenses are prohibitively expensive [3].

However, the same characteristics that make PoW blockchains secure also make it difficult to participate in the network as a node. Mining on many major networks has been monopolized by large-scale mining operators who have amassed influence in network governance. Since operating a mining rig and paying for the associated hardware and electricity costs is too expensive for the average user, large-scale mining operators have collected major influence in the network governance.

2.2. Proof of Stake (PoS) Consensus

PoS is the second most often used consensus method, and it addresses many of the issues that plague PoW blockchains, such as slowness, scalability, inefficient energy usage, and expensive requirements for entering the network.

PoS blockchains use validators instead of miners to validate transactions. Validators are network node operators who validate data in the same way that PoW systems do, but there is no energy-intensive computational process required to earn the right to validate. Instead of doing proofs of work, validators "stake" part of the blockchain's native currency in order to be considered as a validator node. The prospective validator mainly uses blockchain-born crypto tokens to provide collateral. The system randomly selects a validator to confirm data when validating the data held in a transaction block on a PoS blockchain. Certain factors, such as the number of tokens staked, can increase the likelihood of a validator being chosen. The validator is rewarded with network transaction fees when a block is confirmed, and then the process starts over [3].

Proof-of-Stake blockchains keep the network secure and validators honest by demanding validators to stake their tokens. If validators are acting deliberately or incompetently, their involvement and access to the network is lost by a 'slashing' procedure. This stimulus structure makes it possible for validators to earn through legal operation rather than by breaching regulations.

Accessing the PoS blockchains is easier as the validators on the PoS blockchains do not need to invest in high electricity costs and expensive hardware. PoS blockchains are arguably better for the environment than PoW networks in terms of sustainability because they consume significantly less electricity [4].

2.3. Pure Proof of Stake (PPoS) Consensus

PPoS is a new fully decentralized and promising consensus mechanism. The amount of influence each user has over the selection of a new block is proportional to their stake (number of tokens) in the system. Users are chosen at random and in secret to propose and vote on block suggestions. Every online user has the opportunity to be chosen to propose and vote. The probability of a user being picked, as well as the weight of its proposals and votes, is proportionate to its stake.

This means that only two blocks of PPoS cannot be sent to the chain simultaneously since the required threshold of committee votes can be reached by a single block. Up to 1 block in a given round is confirmed and written on the chain. When a block is created, users may immediately depend on the transactions it includes. Thus users can rely on that the block, since it will be part of the chain forever.

PPoS does not place a small group of users responsible for generating blocks, and users do not have to transfer their voting power to a limited number of users. All users are allowed to propose and vote on blocks with a probability that is directly in proportion to their stake.

PPoS doesn't force a user to put a part of their stake aside for a consensus process, and a user can not restrict his ability to spend their stake by participating in a consensus protocol.

In a completely decentralized fashion, PPoS guarantees full participation, protection and speed [4].

3. Distributed Ledger Technologies

The authors first mentioned distributed ledger technology in their 1991 article "How to time-stamp a digital text" [5]. Distributed ledger technology (DLT) is a decentralized framework in which data is shared among network participants. Despite the fact that the approaches to distributed ledger technology are technically distinct, they all have three fundamental components in common.

Every distributed ledger technology should support peer-to-peer transactions, have participants who store the network, and use a consensus mechanism to manage it.

The ability of two or more parties to operate without the use of a central authority is known as a peer-to-peer network. Participants in a distributed ledger are independent computation nodes that record, share, and synchronize transactions.

DLTs are managed without the central authority being required [6]. With the independent conduct of transactions, the efficiency can be improved by reducing errors.

To address the speed and scalability issues in Blockchain technology, new distributed ledger technologies have emerged, such as the Tangle, Hashgraph, Tempo, and Holochain. This newly developed distributed ledger infrastructures aim to improve blockchain technology's limited features.

It has to be also noted that every Blockchain is a type of DLT but not every DLT is a Blockchain [7].

3.1. Hedera Hashgraph

Swirlds developed Hashgraph, a proprietary distributed ledger technology. The consensus algorithm and the Directed Acyclic Graph (DAG) data structure distinguish it from Blockchain technology. Hashgraph does not use conventional Blockchain consensus. Instead of conventional consensus mechanisms, it employs the random gossip protocol and the virtual voting consensus algorithm.
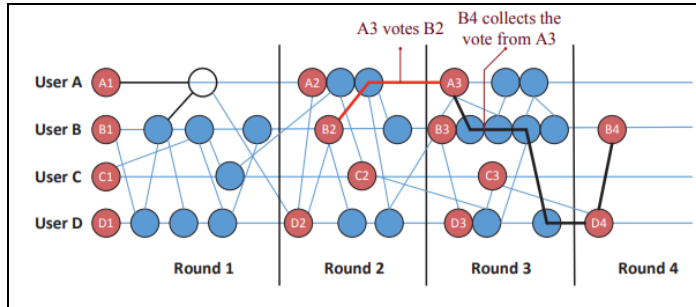
Figure 2: An example of Hashgraph's operation. Image courtesy of [8]

Any transaction that occurs in the Hashgraph network is referred to as an event. Each event includes a timestamp, a hash of the previous two events, transactions, and electronic signature information. The random gossip protocol distributes the time stamp of the verification process to neighboring nodes at random until the data is applied to the ledger [9]. Refer to the Figure 2 as an example of the Hashgraph algorithm.

In order to validate the transaction, this information must be transmitted to the majority of network participants. Transaction timestamps are documented in the case of a separate hashgraph ledger maintained by the participants. Transactions that are agreed by two-thirds of the network through virtual voting are considered legitimate and are registered by each node. Virtual voting can require several rounds in the absence of a majority of participants,

3.2. Tangle

A directed acyclic graph (DAG) is a specialized subject in mathematics, particularly graph theory and computer science. A DAG is made up of vertices and edges (also known as arcs), with each edge pointing from one vertex to another in such a way that following those directions will never result in a closed loop. This mathematical model is being considered as an alternative approach for developing alternative Blockchain structures for applications consisting of many devices [10].
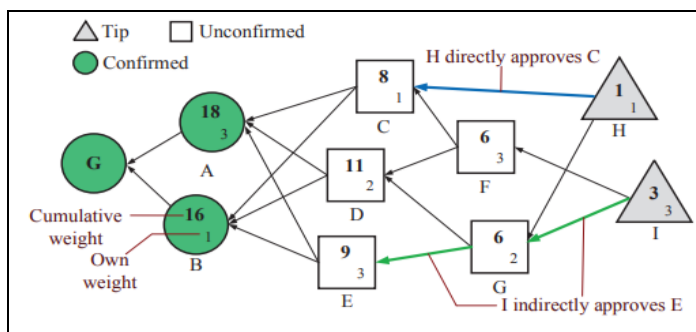
Figure 3: An example of Tangle's DAG based operation. Image courtesy of [8]

Tangle is a new type of DLT which is considered as one of the major DLT that vastly uses the DAG model. Refer to the Figure 3 for visualizing the Tangle's operation.

Since the current mainstream Blockchain infrastructures (i.e. PoW) are unable to meet the efficiency and scalability requirements of IoT systems that require a Trust Layer, Tangle DLT is becoming increasingly important in establishing trust and security for the Internet of Things (IoT) applications.

There is no transaction fee in Tangle since there are no blocks or miners to add data to the network. Transactions are not kept in a sequential order. Transactions are not recorded in a linear fashion. Each transaction in the consensus protocol verifies the previous two transactions and records the transaction data. In terms of security, it is less secure than Blockchain technology especially in the early stages of the network. When the transaction volume is low, network coordinators are used for protecting the network.

The use of network coordinators may result in the network not being fully distributed. The goal is to improve the scalability, speed, and efficiency of the Tangle protocol. The speed of the network increases in direct proportion to its size, especially as the network grows.

Tangle, which is used in IOTA, is one of the most important applications that use the DAG model [11].

It is appropriate to use this advanced DAG based DLT particularly in IoT applications or any other applications that require providing speed and scalability to a network composed of large number of devices and transactions.

3.3. Holochain

Holochain is an open-source distributed ledger technology. Unlike Blockchains, Holochain relies on individual chains rather than a single chain. While transactions in Blockchain technology are made over a single chain for all network participants, participants in Holochain store data in their local chains as part of the public network, Instead of recording all of the data on the network, participants only record their own transactions. This situation should be considered as the main reason why this DLT requires low amount of storage on participating nodes. The Distributed Hash Table (DHT) [12] ensures the correctness of local chains. DHT stores the chain data's DNA (hash) information and ensures that the chain's invariance is maintained. As indicated in Figure 4, DHT is stored in parts of the mesh topology [13]. While some users are offline, the network can continue to function. The disadvantage of this protocol is that information can be lost if a registered peer's device completely malfunctions. A backup procedure is additionally required for all peers in order to prevent data losses.

Distributed applications (dApps) built on the Holochain,which are named as Holochain Apps (hApps), can run quickly, and the network's scalability is limitless. Millions of transactions can be processed per second with Holochain. Holochain aspires to transform the current internet's center-based services.

3.4. Tempo Radix

Sharding method is used by Tempo distributed ledger technology in order to distribute data amongst many nodes in the

network. Hence, this situation eases the amount of storage required on the peers.
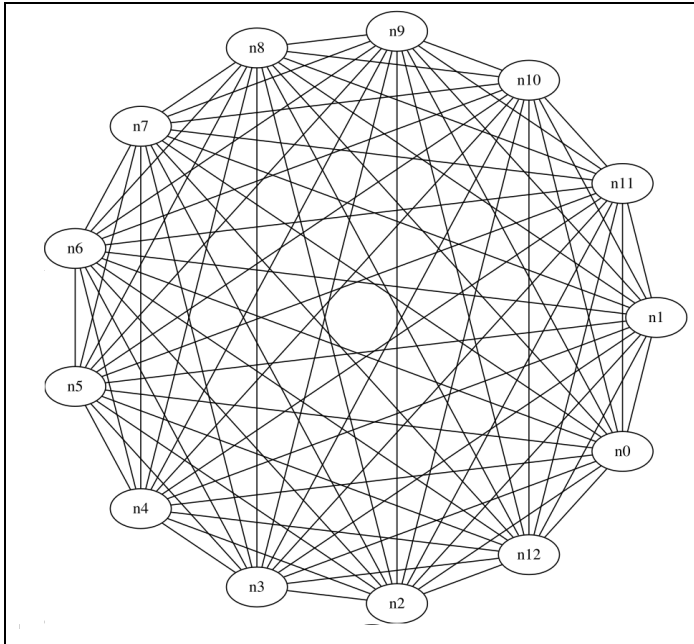
Figure 4: A representation of Holochain's DHT topology. Image courtesy of [13]

Tempo relies on Lamport's Logical Clocks, see Figure 5 for a sample representation of this concept, to come to an agreement on the distributed ledger. Any time a transaction is recorded, the logical hour number is also recorded. Each node increments its logical counter for every transaction it witnesses. Due to the unreliability of using the physical clock in distributed systems, the logical clock system is used. Tempo logical hours are smart counters that count upwards in increments of one when a participant sees a new and verified transaction in the ledger,
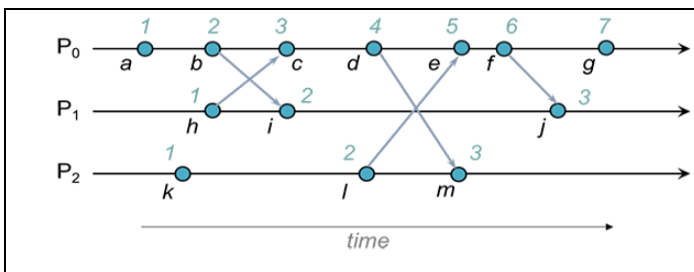
Figure 5: A representation of Lamport Logical Clocks. Image courtesy of [14]

Tempo participants use the Gossip protocol in order to synchronize the timeliness of the information in their tracks [15].

Participants in a blockchain network have access to the entire ledger, but the Tempo infrastructure uses shredding technology to limit participants' access to a subset of the ledger. Some sections of the participants' data are also being recorded by neighboring nodes as a data-loss prevention mechanism.

4. Comparison of PoW, PoS, PPoS Blockchain Consensus Mechanisms and Distributed Ledger Technologies

4.1. Blockchain versus Hedara Hashgraph

In terms of speed and scalability, Hedera Hashgraph can outperform the Blockchain technology. The transaction volume is high due to the Hedera Hashgraph's random gossip protocol, which allows transactions to be released quickly to the network. When a node conducts a transaction, it transfers transaction information to a randomly chosen neighbor node [9]. The node that discovers the transaction will also pass this information on to a random neighbor node. This procedure is repeated until the entire network is aware of the information. As a result, the network is immediately alerted when a new transaction occurs.

Hashgraph DLT uses virtual voting to validate each transaction, rather than the costly mining process in some Blockchain infrastructures. If two-thirds of the entire network accepts the transaction, the transaction is valid. Virtual voting can take place several times if the majority is not reached. In Hedera Hashgraph distributed ledger technology, a maximum of 250,000 transactions per second can be performed [16].

Hedera Hashgraph is more scalable than Blockchain. The Hedera Hashgraph network can potentially have an infinite number of participants.

4.2. Blockchain versus Tangle

Tangle distributed ledger technology has increased scalability and quantum durability compared to Blockchain technology, while reducing transaction fees in open blockchains.

To confirm a transaction in Tangle, each node must verify at least two transactions from previous transactions. The participant's transaction becomes more valid as it validates more transactions. It is the most resistant distributed ledger technology even it is also considered as quantum-safe.

Tangle arises amongst the other DLT alternatives with its lowest transaction fee. When compared with Blockchain, the security of Tangle is weaker when the network size is smaller.

The number of transactions per second in Tangle DLT his estimated between 500 and 800 [17]. Due to the nature of Tangle's operation, it is assumed to allow an unlimited number of participants.

Tangle infrastructure can be preferred in applications that require speed and scalability such as instant mobile voting applications. Also, it is alleged that Tangle can also have a good impact on the Energy Trade industry [8].

4.3. Blockchain versus Holochain

In comparison to Blockchain, resources are used efficiently and scalability is increased with Holochain. Data on the network is kept in separate chains rather than a common ledger in Holochain technology .

As part of the public network, each participant keeps its own local records. Each participant keeps a separate ledger. Distributed Hash Table (DHT) ensures the accuracy and immutability of local ledgers. Holochain DLT allows a truly distributed architecture in which everyone maintains their own ledgers.

As opposed to Blockchain technology, energy efficiency is ensured by avoiding costly mining operations via the Holochain consensus mechanism.

Since there is a central point through which all transactions must pass, Holochain does not set a TPS (transactions per second) value like other blockchain-based or blockchain-derived projects [18]. Holochain, on the other hand, is a generalized distributed computing protocol. Due to the underlying mechanisms, throughput and scalability are assumed as theoretically infinite.

4.4. Blockchain versus Tempo Radix

In terms of speed, scalability, and resource efficiency, Tempo distributed ledger technology can outperform Blockchain. As opposed to Blockchain's single chain, Tempo distributes the subportions of the entire ledger onto the network participants' storage. In other words, Tempo participants store the puzzle pieces by viewing the entire ledger as a puzzle [15].

Each node keeps a piece of its neighbor node in the fragmentation process in order to prevent data loss. Tempo distributed ledger technology's shredding process has advantages in terms of speed and resource efficiency. In Blockchain, a lot more resources are used since the entire block data / ledger is kept on all participants. Additionally, distributed ledger synchronization is slower and less scalable than Tempo distributed ledger technology.

Tempo distributed ledger technology has been tested to handle 1,400,000 transactions per second [19]. The Tempo network has a significant advantage over Blockchain technology because the maximum number of participants is theoretically unlimited.

The Gossip protocol is used in Tempo distributed ledger technology ensures that nodes have up-to-date information about the distributed ledger. The Gossip protocol allows the nodes in a network to propagate new configurations to the rest of the network. This protocol allows network nodes to communicate with one another and exchange information about their components.

4.5. Comparison of Blockchain (PoW, PoS, PPoS) and DLTs

In a typical setting, performance and evaluation metrics for developing a distributed application depends on four key parameters, Throughput (total number of transactions per second on the distributed network), Scalability (Supported number of nodes in the distributed network), Energy consumption per transaction and the level of storage requirement/usage.

'Table 1' summarizes the key performance and evaluation metrics of the consensus mechanisms and DLTs which are outlined in this study.

Throughput, scalability values and the storage usage levels are gathered from the internally conducted research and from the following references: [16], [17], [20]-[22].

For the energy consumption of Blockchain consensus protocols, in a network of 10000 nodes, which is approximately the number of active full nodes on PoS, note that this amounts to approximately 1000 J per transaction for PoS, it is many orders of magnitude less than for the current PoW blockchains which is about 1 billion J per transaction [23]. It has to be also noted that PPoS uses 30 J per transaction [24].

As for the DLT segment, Tangle uses 3.6 J per transaction [25]. It is claimed that Hashgraph uses 3600 J per transaction [26]. Holochain's energy consumption is estimated as 325 J per

transaction [27]. Energy consumption of Tempo is not available in the literature, however it is considered very low since no mining is involved [28].

Table 1: Key performance comparison and evaluation metrics of Blockchain (PoW, PoS, PPoS) and DLTs

	Throughput (transactions per second)	Scalability (number of nodes)	Energy Consumption (J per transaction)	Usage of Storage
Blockchain - PoW	~7 tps	~10.000 nodes	~1 billion J/transaction	High
Blockchain - PoS	~25 tps	~10.000 nodes	1000 J/transaction	High
Blockchain - PPoS	~10000 tps	500.000 nodes	~30 J/transaction	High
DLT - Hashgraph	~250.000 tps (Theoretically) ~10.000 tps (recent practical result)	Theoretically Unlimited	3600 J/transaction	Moderate
DLT - Tangle	~500 tps	Theoretically Unlimited	3.6 J/transaction	Moderate
DLT - Holochain	Theoretically Unlimited	Theoretically Unlimited	325 J/transaction	Low
DLT - Tempo	~1.400.000 tps	Theoretically Unlimited	N/A (Considered very low)	Low

5. Modern Applications of Blockchains and Distributed Ledgers

In empirical settings, blockchain and distributed ledger technologies can be used in a variety of ways. Like all technologies, a Blockchain application must be viewed as a sociotechnical system. A sociotechnical system consists of a technological artifact and the social environment in which it is used, as well as the interactions between the two. When viewed as a sociotechnical system, the core characteristics of Blockchain and DLTs can be applied to a variety of situations, resulting in a wide range of use cases.

5.1. Cryptocurrency

A cryptocurrency is a digital or virtual currency that is mostly based on consistent Blockchain consensus mechanisms and protected by cryptography, making counterfeiting and double-spending nearly impossible. Cryptocurrencies are distinguished by the fact that they are not issued by any central authority thus allowing them to be theoretically immune to government intervention or manipulation.

Cryptocurrencies have the potential to make it easier to transfer funds between two parties without the use of a trusted third party such as a bank or credit card company. In addition, users can avoid the high fees charged by banks and financial institutions for wire transfers by preferring cryptocurrency based fund transfers with minimal processing fees [29].

Major disadvantageous situation of Cryptocurrencies is that the semi-anonymous nature of Cryptocurrency transactions makes them ideal for a variety of illegal activities, including money laundering and tax evasion. However, Bitcoin is a poor

choice for conducting illegal operations because forensic analysis of the Bitcoin has aided authorities in detecting criminal activity. It has to be noted that more privacy-oriented coins like Dash, Monero, and ZCash, on the other hand, are far more difficult to track than the Bitcoin [30].

5.2. Decentralized Finance (DeFi)

Decentralized Finance, also known as DeFi, is a movement that allows users to access financial services such as borrowing, lending, and trading without relying on centralized entities. These financial services are delivered through Decentralized Applications (Dapps), the vast majority of which are built on the PoS powered Ethereum platform [31].

DeFi is a collection of products and services that act as a replacement for institutions such as banking, insurance, bonds, and money markets, rather than a single product or company. DeFi Dapps allow users to combine their services in order to expand their options. DeFi Dapps typically require collateral to be locked into smart contracts in order to function. The Total Value Locked is a term used to describe the total collateral locked in DeFi Dapps [32].

By eliminating the need for any middlemen, DeFi Dapps have the potential to revolutionize traditional financial services. It should be noted, however, that DeFi in its current state is still very new and experimental, with many projects being improved on a daily basis. DeFi may evolve further over time and become completely unrecognizable from what it is now.

5.3. Non-Fungible Tokens (NFT)

NFTs (non-fungible tokens) are a new type of blockchain-based token that is one-of-a-kind and indivisible. They first appeared in late 2017. NFTs are digital tokens that can be used to show who owns one-of-a-kind items. Unlike PoW powered Bitcoin, where each coin is identical, NFTs are unique, with their own set of features. Tokenized items include art, valuables, and even real estate that no one can alter the ownership record. NFTs can only have one official owner at a time and are mainly secured by the PoS based Blockchain infrastructures [33].

5.4. Digital Identity

Any document that can be used to prove an individual's identity is known as an identity document. We frequently share our identity information to authenticate in our daily lives. Personal data is violated when identity information is stored, processed, and shared against the person's will. There is no universally accepted identity in the Internet.

A digital identity system can be used for information confirmation as well as authentication. Users will have complete control over their digital ID and personal information. Verification can be handled with or without sharing entire personal information in the digital identity system.

Blockchain has progressed significantly since its inception as a distributed ledger system for tracking bitcoin ownership. This technology has the potential to replace traditional identity management systems with a highly trusted mechanism. Users will be able to have more control over their own identities owing to the infrastructure provided by the Blockchains [34] and DLTs.

Organizations can only use the data with the consent of customers, and no central entity will be able to compromise a customer's identity. Self-sovereign identity, which is inherently unalterable and more secure than traditional identity systems, has been made possible by Blockchain [35]. This has the potential to completely transform how we connect to various online services using our identities. Individuals would verify their identity using their self-sovereign ID, eliminating the need for passwords.

As with any life-changing innovation, there has been a long period of evolution, with experts exchanging ideas and little agreement on what self-sovereign ID means. It is based on the idea that an individual must have control over how his identity is managed. With user consent, the ID cannot be tied to a single system and must be interoperable across multiple platforms. Experts considered combining various identifying data, such as demographic and employment-related information, as well as individual information revealed by others.

In the Blockchain domain, digital identities and user information will be stored in a digital wallet. There will be no need to carry or store many cards or documents. In a Blockchain based digital identity system, documents such as our citizenship card, bank cards, diplomas, certificates, driver's license, and passport are being planned to be stored in self-sovereign id oriented digital wallets.

5.5. Supply Chain Management

One of the most important areas in which blockchain projects are concentrated is considered as supply chain management. A trust issue arises because of the large number of parties involved in supply chain processes. Supply chain transaction information is being planned to be securely stored by Blockchain [36]. The current method of running these processes on paper makes it difficult to manage and unreliable. By using Blockchain, it is possible to conduct transactions securely on digital media and to view all stages of the processes transparently with confidence. Circulation of counterfeit products can also be prevented from the market using this approach. Preventing the sale of smuggled goods will protect citizens' rights and countries from incurring tax losses.

Consumers may not fully confirm whether the procured products were genuine or not in a traditional supply chain process. Consumers can reliably view the production and distribution history of products by using Blockchain enhanced Supply Chain Management Systems. With this understanding in mind, countries may want to look at the manufacturing, distribution, and shipping history of all imported goods in the future.

5.6. Digital Notary Services

A notary, also known as a public notary, is a state-appointed official who has the authority to certify a specific action or deed. Contracts, documents, affidavits useful in other jurisdictions, and deeds are all included. In order to perform the role of notary, an official must be objective when making various declarations. A notary's duties include everything from signing a property sale record to administering oaths and keeping accurate records of other important deeds and documents. The officer can use an

electronic or digital notary to perform notarizing functions digitally.

The authorized seal will be applied to the certified document by the electronic notary. This activity requires cryptography and a secure public key in order to manage, develop, store, and distribute the digital certificate since a digital notary must keep an up-to-date electronic record of all completed notary tasks. Remote implementation of a significant digital solution for public notary activities is also possible. This notarization process is expected to be aided by Blockchain and DLT technology. As a prospecting result, users of this service can trust the integrity of data on the Blockchain or DLT.

Nonrepudiation, tamper-resistance features and the well-known characteristics of distributed ledgers make it an excellent tool for enhancing notary capabilities [37].

5.7. Land Registry and Property Ownership Management

Land registration entails gathering information such as the ownership and property size information of a land piece. The entire process of land registry maintenance is extremely time-consuming because it demands the safekeeping of massive volumes of written registers. The current method is insecure since the majority of this procedure is opaque and continuous sales of a property must be accurately recorded. There have been several proposals to automate land register data recordings by eliminating the need for handwritten records. This is initially accomplished by storing the data in massive databases. However, in terms of data security, such a strategy is ineffective because the data contents are easily accessible, and data tampering can occur in poorly maintained databases.

A distributed ledger based solution for property registration could significantly improve efficiency and possibly eliminate fraud [38]. Contracts and ownership data can be stored in a distributed manner too. Blockchain and DLT implementations can make it easier to track data transactions because it eliminates the need for physical involvement thus enhancing the overall system security for users. Developing countries are considering Blockchain and DLTs as a tool for updating their existing land registry systems.

5.8. Health Records Management

Preventing illnesses or detecting them early would improve society's living conditions and reduce the workload of hospitals. Citizens' health status should be constantly tracked using preventive health technologies. The volume of health data will grow as a result of constant surveillance of citizens' health status. This data can be immutably stored and exchanged for clinical trials using Blockchain and DLT technology.

Health data is vital information that countries must safeguard. Theft of these details poses a threat to the country's national security. Health information is also essential to the pharmaceutical industry. As a result, health data has become an obvious priority for financial benefit.

Blockchain and DLTs can serve as a link between healthcare researchers and health data records in order to facilitate data exchange. Blockchain and DLTs can provide secure storage and

exchange of health data at any time. Anonymized health data can have economic benefits to both users and organizations [39]. A health consortium with Blockchain infrastructure, comprised of organizations that generate and need data, can be created. As a result, the data can be safely exchanged with the desired entities and organizations for scientific purposes.

5.9. Insurance

The insurance industry can use Blockchain and DLT technology in order to improve core internal processes like claim-submission and processing, as well as fraud detection and prevention. Blockchain can also be used to modify and possibly create new insurance brands, as well as improve existing ones. It has also come to light that Blockchain faces a number of issues, including scalability, security, and privacy, as well as taxation and regulation. The insurance industry has yet to reap the benefits of new generation blockchain architectures and distributed ledger technologies [40].

5.10. Electronic Voting Systems

It has long been a challenge to create a secure voting structure that offers the same fairness and privacy of current voting schemes while also providing the transparency and flexibility of electronic systems. To compete with the conventional ballot system, an electronic voting system must meet the same security and privacy criteria.

A Blockchain or DLT based e-voting technology has the ability to reduce election fraud while also increasing voter access [41]. Such an e-voting system could decentralize controls, enabling electors to take on certain duties while maintaining a copy of the electoral register. Since the copies of the ballots are a voting history that cannot be reversed, there can be no null votes [42].

5.11. Energy Distribution and Trade

Given the finite nature of resources, countries must increase their production of renewable energy resources. As a result, the number of local producers who use renewable energy is expected to grow in the future. The structure of energy distribution and trade will be altered as the number of local energy producers increases. To encourage local producers, an energy trading platform would be developed, allowing producers to load their excess energy to the grid immediately.

It is necessary to ensure that energy exchange between the local producer and the consumer is secure. Blockchain and Distributed ledgers could be used to create a platform that allows local producers and consumers to transact usage certificates instantly. For such an energy distribution and trade infrastructure, all parties involved in the energy distribution process must form a distributed ledger based consortium. In the energy sector, Blockchain and DLT will bring together market regulators, distributors, consumers, home users, and producers on a single, reliable energy market and sharing platform [43].

This platform will enable real-time allocation of required and preferred types of energy, as well as the approval of energy consumption. This platform would provide incentives to renewable energy producers. Blockchain has the potential to provide an infrastructure for instant energy trade for energy consumers and local conventional or renewable energy producers [44]. Smart meters will also need to be developed in order to measure the produced or purchased energy.

Active customer engagement and energy trading would require immutable, open, and dependable smart contracts. Such digital energy market systems can ensure that energy is priced immediately and it is delivered on demand, and allows the customer to determine their own energy consumption preferences. It would also make the consumer more efficient in deciding the frequency of energy usage.

It has to be noted that especially for the energy industry, DAG based structures can have a critical role [8].

5.12. *Distributed File Systems*

Distributed file systems (DFS) have gained a lot of interest recently. Instability, auditing, and incentive mechanisms are all weaknesses in traditional Peer-to-Peer (P2P) distributed file systems. Although Blockchain-based DFSs successfully provide enough incentives and security assurances, a number of problems such as scalability and privacy concerns are preventing the next generation of Blockchain DFSs from being developed. As a result, it is anticipated that DLTs have much greater potential than conventional Blockchain in order to establish a globally distributed file system due to their storage-wise and scalability advantages [45].

5.13. *Artificial Intelligence and Federated Learning*

A new machine learning paradigm known as federated learning (FL) has arisen as a result of the increasing computing capabilities at end user devices, as well as the growing privacy concerns over sharing sensitive raw data. FL has the potential to avoid exchanging data directly by training models locally at each client and aggregating learning models at a central server, hence avoiding privacy leaks. The standard FL structure, on the other hand, is strongly reliant on a single central server, which could fail if that server acts maliciously. To address the single point of failure problem, new approaches are being developed that focus on establishing a DLT-assisted decentralized FL framework that can effectively prevent malicious clients from poisoning the learning process while also providing clients with a self-motivated and reliable learning environment.

Distributed Ledger Technologies have the potential to improve transparency and trust in the distributed machine learning process and generating consistent models by permanently logging it into a distributed ledger. Since the model is available to the public and the history is preserved, the blockchain tracks not just the current model but also all of the previous iterations. Since decentralization is inherent in both, federated learning can benefit from distributed ledgers [46].

6. Application-specific Minimum Performance Requirements for Achievability of the Applications

Minimum required levels of performance requirements for the above mentioned applications are given in Table 2. This section does not provide the desired ideal state of applications. It has to be noted that; there does not exist any universal solution for all types of applications. Hence, it is necessary to outline the needed basic and minimum performance requirements for having a better decision-making action in order to select an appropriate protocol or conduct a better protocol implementation strategy during the initial phases of a specific distributed project. All of rationals of Table-2 are given in the subsections.

A scoring system of Low, Moderate and High is used in order to outline the minimum requirements. It has to be that 'DC' denotes, as a scientific and engineering abbreviation, "Don't Care".

Table 2: Application-specific Minimum Performance Requirements for the Achievability of the Applications

	Throughput (transactions per second)	Scalability (number of nodes)	Energy Efficiency (J p/tx)	Level of Storage Requirement
Crypto Currency	Low	Low	DC	DC
DeFi	Low	Moderate	Moderate	DC
NFT	Low	Low	DC	DC
Digital ID	Moderate	DC	DC	Moderate/High
Supply Chain	Moderate	Moderate	DC	High
Digital Notary	Low	Low	DC	High
Land Registry	Low	Low	DC	Moderate
Health Records	Moderate	Low	DC	High
Insurance	Moderate	Low	DC	Low
Electronic Voting	High	High	High	Moderate
Energy Trade	Moderate	High	High	Moderate
DFS	High	Moderate	DC	High
AI/FL	Moderate	Moderate	DC	High

6.1. Cryptocurrency – Minimum Performance Requirements

Cryptocurrency applications can be achieved with low rate of transactions per second and low number of participating nodes. Any cryptocurrency system can be achieved with low levels of transaction rate with a small number of validating or participating nodes. The most basic and stable example of this situation can be given as the PoW based Bitcoin in which the performance metrics are given in Table 1.

6.2. DeFi – Minimum Performance Requirements

As in the case of Cryptocurrency, any DeFi project can be achieved with low transaction rate. Major difference is the necessity of having a more energy efficient and moderately scalable infrastructure for the purpose of having a more future-proof DeFi system. The reason is that; there is the possibility of having multiple mobile devices participating in the DeFi domain in the near future.

6.3. NFT – Minimum Performance Requirements

NFT portals or NFT systems recently does not have and need many transactions per second and participating nodes, since NFTs are used mainly for the occasional transfer of the ownership of any concept.

6.4. Digital ID – Minimum Performance Requirements

The required transactions of Digital ID records will be moderately high enough. In an interconnected e-government environment, the transactions between the official Digital ID issuers, validators and peers will be moderately high enough in a daily setting. This situation will also be true for keeping and updating the Digital ID records of the objects.

Considering the large number of files in a countrywide setting, the storage requirement will be essentially very high for the issuer and verifier sides of a Digital ID system.

As a special note; if there exits multiple IoT (Internet of Things) devices in an Digital ID system then the level of storage requirement needs to be moderately large enough in order to store the entire history of record in the IoT node's lifecycle.

6.5. Supply Chain – Minimum Performance Requirements

In a worldwide supply chain setting, the amount of transactions between suppliers, manufacturers, distributors, retailers and consumers is considered very large. As a result, the required level of storage is needed to be large enough in order to collect and store the entire decades of transaction records.

The required minimum transactions need to be moderately high enough due to the speed requirements in a global environment for a supply chain operation and considering the amount of participating nodes in this setting, the scalability level should also need to be moderate.

6.6. Digital Notary – Minimum Performance Requirements

On a daily urban life setting, the necessity and frequency of notary services is considered high. In a Blockchain/DLT based Digital Notary scenario, the notary branches in a city can be considered as validating nodes. When compared with any large scale application, the number of notary branches are not large enough and the approval time needed in a branch does not require high levels of scalability. As a result, the minimum throughput and scalability requirements are selected as low.

One major situation about the notary services is considered as the increasing intensity of the volumes of documents. Hence, the minimum required level of storage is selected as high.

6.7. Land Registry – Minimum Performance Requirements

Land registry services can be considered as a subsegment of digital notary services. The only difference between the digital notary requirements is the volume of data that needs to be stored. It is considered lesser than the digital notary. Hence, the minimum storage requirement is set as moderate.

6.8. Health Records – Minimum Performance Requirements

It is required to have fast approvals of transactions of data in a network of hospital environment due to the criticality of individual health; therefore a Blockchain based immutable health records database minimally should need to have a moderate level of transaction rate of health recordings. The healths data is defined as laboratory test results, reports, prescriptions, diagnostic info and imaging data. It has to be also noted that due to the intensity of large amount of imaging data, the minimum required level of storage level is considered as high.

6.9. Insurance – Minimum Performance Requirements

In a global or multi-cliented realtime digital insurance scenario, daily and instant insurance contracts are being forecasted to be part of society's daily life. Anticipated participants will be mostly rental items such as automobiles. Due

to the fact that; the number of registered validating nodes of an insurance company is considered as low enough. Hence, the scalability is selected as low. Whereas the speed of transactions should be moderately high enough in order to reduce the waiting durations during insurance processes. In addition to the transaction rate and scalability requirements, since the needed records data is text based, the minimum requirement for level of storage is determined as low. However, if there will be many items to be insured frequently then the requirement should be considered as moderate.

6.10. E-Voting – Minimum Performance Requirements

Electronic Voting, as one of the most suitable and politically sensitive application on the Blockchain and DLT domain, have high demanding transaction rate and scalability requirements. The setting environment in this application area is considered as fully mobile and distributed voting system. Hence, the energy efficiency requirement is selected as high. Given that; the necessity of having the voting/election as rapid as possible in a large scale voting event, the transaction rate requirement is also considered as high and due to the required mobile operation the scalability should also be considered as high.

6.11. Energy Trade – Minimum Performance Requirements

Blockchain-enabled Energy Trade sector is being regarded as a digital transformation technology in a Smart City context. In a practical sense, this concept is considered as a large scale IoT application in which the participants measure, share and certificate the energy production and consumption. With the inclusion of prosumers to the consumers and producers, the frequency of transactions of the energy certifications would be moderately high enough in a city/country wide environment. Hence, the minimum transaction rate is selected as moderate. Because of the prosumer and consumer-approval support for such an application, the minimum scalability requirement is considered as high. Because of the scenario of consumer-approvals and instant-billing and energy monitoring cases, the minimum energy efficiency should be selected as high. The data transactions are moderately high enough and the data size to be stored is essentially low. Hence, the minimum level of storage requirement should be kept as moderately high.

6.12. DFS – Minimum Performance Requirements

A globally effective Distributed File System should need to have immutable and zero-loss structure. It should be also accessible fast enough, the participating nodes should be fully decentralized and the number of nodes should be moderately enough. Energy efficiency is not considered as a requirement. Due to the variety of multiple data types such as text, audio, video and other possible large file types, the minimum required level of storage is considered as high.

6.13. AI/FL – Minimum Performance Requirements

As a promising future application, federated learning and distributed neural networks model training will require high levels of data storage and fast execution times and multiple nodes.

7. Applicability of Blockchain (PoW, PoS, PPoS) and DLTs on Modern Applications

The Table 3 above summarizes Blockchain's (Note that Blockchain is considered as PoW, PoS, PPoS in this study) and new DLTs' suitability for the explained modern applications which are outlined throughout this study by following the information of minimum performance requirements per application given in Section 6 and Table 2. For better assessment the values in Table 1 is converted into the Low, Moderate, High performance scoring.

Table 3: Performance scoring levels of Blockchain (PoW, PoS, PPoS) and DLTs

	Throughput (transactions per second)	Scalability (number of nodes)	Energy Efficiency	Storage Efficiency
Blockchain - PoW	Low	Low	Low	Low
Blockchain - PoS	Low	Low	Moderate	Low
Blockchain - PPoS	Moderate	High	High	Low
DLT - Hashgraph	Moderate	High	Moderate	Moderate
DLT - Tangle	Low	High	High	Moderate
DLT - Holochain	High	High	High	High
DLT - Tempo	High	High	High	High

Table 4 is constructed based on the scoring values of the performance criteria provided in Table 3 and minimum performance requirements provided in Table 2.

Table 4: Applicability of PoW, PoS, PPoS and DLTs on Modern Applications

	PoW	PoS	PPoS	Hashgraph	Tangle	Holochain	Tempo
Crypto Currency	✓	✓	✓	✓	✓	✓	✓
DeFi			✓	✓	✓	✓	✓
NFT	✓	✓	✓	✓	✓	✓	✓
Digital ID			✓	✓		✓	✓
Supply Chain			✓	✓		✓	✓
Digital Notary	✓	✓	✓	✓	✓	✓	✓
Land Registry	✓	✓	✓	✓	✓	✓	✓
Health Records			✓	✓		✓	✓
Insurance			✓	✓		✓	✓
Electronic Voting						✓	✓
Energy Trade			✓			✓	✓
DFS						✓	✓
AI/FL			✓	✓		✓	✓

One major factor about the mapping of the consensus/DLT protocols is the distinction of consensus/protocol specific storage efficiency and application-specific minimally required storage levels. During the selection process of the suitable protocol, designer should be aware of that the needed amount in the peer's storage will be directly dependent on the selected protocol's storage efficiency and the desired application's required minimum level of storage. As an example, if the selected application requires high levels of storage and the selected consensus/protocol has low storage efficiency, then the resultant distributed system's peers will possibly require an enormous storage space. Hence, the storage related columns are not taken into consideration for the consensus/protocol selection process. The remaining Throughput, Scalability and Energy Efficiency criteria are taken into consideration for the selection of suitable consensus/protocol for the listed applications.

8. Conclusions

By deploying DLTs and Blockchain to suitable applications, parties can perform stable transactions for the first time in human history without relying on a centralized authority. Blockchain and Distributed Ledger Technologies have the power to modify most structures which are currently under central authority's control because of the concept of trust on peer transactions.

Blockchain infrastructures are inefficient in applications with high participation and transaction volume. As a result, new distributed ledger technologies have been created that primarily address the speed and scalability issues of Blockchain. Hashgraph, Tangle, Holochain and Tempo distributed ledger technologies are the recent alternatives to Blockchain. While these new distributed ledger technologies claim to solve the problems of Blockchain technology, their effectiveness has yet to be proven.

Participants and peers should should have the rights to conduct transactions without the use of intermediaries, and all Blockchain and Distributed Ledger Technologies should ensure decentralized management and participant data security. Under the hood of this prospect, a decentralized, flat and fully democratized structure with applications on Finance, Market Exchanges, Digital ID, Voting, Energy Trade, Distributed Storages, Insurance, Records Managements, Land Registry and AI will certainly constitute the future of humanity.

The completion of necessary country-wide regulations, as well as technical studies, will be a critical step toward the global adoption of blockchain and distributed ledger technologies, as well as related critical applications, in the near future.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] N. Chaudhry and M. M. Yousaf, "Consensus Algorithms in Blockchain: Comparative Analysis, Challenges and Opportunities," 2018 12th International Conference on Open Source Systems and Technologies (ICOSST), 54-63, 2018, doi: 10.1109/ICOSST.2018.8632190.

[2] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System" [Online]. Available: https://bitcoin.org/bitcoin.pdf [Accessed: 05-08-2020].

[3] L. M. Bach, B. Mihaljevic and M. Zagar, "Comparative analysis of blockchain consensus algorithms," 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 1545-1550, 2018, doi: 10.23919/MIPRO.2018.8400278.

[4] Cryptopedia, "Types of Blockchain: PoW, PoS, Private, and DLT" [Online], Available: https://www.gemini.com/cryptopedia/blockchain-types-pow-pos-private [Accessed: 01-Jul-2021]

[5] S.Haber, W.S.Stornetta, "How to time-stamp a digital document.", J. Cryptology 3, 99–111, 1991, doi:10.1007/BF00196791.

[6] A.Sunyaev, Distributed Ledger Technology. In: Internet Computing. Springer, Cham, 2020, doi:10.1007/978-3-030-34957-8_9.

[7] A. Panwar, V. Bhatnagar, "Distributed Ledger Technology (DLT): The Beginning of a Technological Revolution for Blockchain," 2nd International Conference on Data, Engineering and Applications (IDEA), 1-5, 2020, doi: 10.1109/IDEA49133.2020.9170699

[8] H. Alen, C. Tomislav & Z. Ivana, Demystifying Distributed Ledger Technologies: Limits, Challenges and Potentials in the Energy Sector. IEEE Access, 7, 1-1. 2020. Doi: 10.1109/ACCESS.2020.3007935.

[9] Hedera Hashgraph, "Hedera: A Public Hashgraph Network & Governing Council"

[10] F. M. Benčić and I. Podnar Žarko, "Distributed Ledger Technology: Blockchain Compared to Directed Acyclic Graph," 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), 1569-1570,2018, doi: 10.1109/ICDCS.2018.00171.

[11] N. Zivi et al., Directed Acyclic Graph as Tangle: an IoT Alternative to Blockchains. 1-3, 2019. 10.1109/TELFOR48224.2019.8971190.

[12] Holochain, "Holochain" [Online]. Available: https://github.com/Holochain/holochain-proto/blob/whitepaper/holochain.pdf [Accessed: 06-08-2020].

[13] Ceptr, "Holochains for Distributed Data Integrity", [Online]. Available: http://ceptr.org/projects/holochain [Accessed:08-Jul-2020]

[14] Rutgers University, Logical clocks, [Online]. Available: https://www.cs.rutgers.edu/~pxk/417/notes/logical-clocks.html [Accessed:07-Jul-2021]

[15] Tempo, "Building Tempo" [Online]. Available: https://www.radixdlt.com/post/tempo-consensus-lessons-learned/ [Accessed: 05-08-2020].

[16] Cao, Bin & Li, Yixin & Zhang, Lei & Zhang, Long & Mumtaz, Shahid & Zhou, Zhenyu & Peng, Mugen. (2019). When Internet of Things Meets Blockchain: Challenges in Distributed Consensus. IEEE Network. PP. 1. 10.1109/MNET.2019.1900002.

[17] K. Enis Data Storage in the Decentralized World: Blockchain and Derivatives, 2020. Doi: 10.26650/B/ET06.2020.011.03.

[18] Holochain, "What is the TPS (Transactions Per Second) on Holochain? - Learning Library / FAQs - Holochain Forum" [Online]. Available: https://forum.holochain.org/t/what-is-the-tps-transactions-per-second-on-holochain/191 [Accessed: 08-Jul-2021]

[19] Radix DLT Tempo, "10 years of Bitcoin history, replayed in under 30 minutes | The Radix Blog | Radix DLT" [Online], Available: https://www.radixdlt.com/post/replaying-bitcoin [Accessed: 01-Jul-2021]

[20] Holochain, "The Holochain Guidebook FAQ",

[21] Hedera Hashgraph, "Hedera Hashgraph", [Online]. Available: https://hedera.com/ [Accessed: 05-Jul-2021]

[22] Lepore, Cristian & Ceria, Michela & Visconti, Andrea & Rao, Udai Pratap & Shah, Kaushal & Zanolini, Luca. (2020). A Survey on Blockchain Consensus with a Performance Comparison of PoW, PoS and Pure PoS. Mathematics. 8. 1782. 10.3390/math8101782.

[23] Sedlmeir, J., Buhl, H.U., Fridgen, G. et al. The Energy Consumption of Blockchain Technology: Beyond Myth. Bus Inf Syst Eng 62, 599–608 (2020). https://doi.org/10.1007/s12599-020-00656-x

[24] Sustainable Blockchain: Estimating the Carbon Footprint of Algorand's Pure Proof-of-Stake | Algorand, [Online]. Available: https://www.algorand.com/resources/blog/sustainable-blockchain-calculating-the-carbon-footprint . [Accessed:07-Jul-2021]

[25] IOTA, "Tangle Energy Benchmarks for IOTA", [Online]. Available: https://blog.iota.org/internal-energy-benchmarks-for-iota/ [Accessed: 06-Jul-2021]

[26] Leafscore, "Sustainable Crypto 2021", [Online]. Available: https://www.leafscore.com/blog/the-9-most-sustainable-cryptocurrencies-for-2021/ [Accessed:07-Jul-2021]

[27] Holochain, "Holoport Electricity Usage", [Online]. Available: https://help.holo.host/support/solutions/articles/36000035595-will-holo-use-lots-of-electricity-how-much-electricity-will-my-holoport-use- [Accessed: 07-Jul-2021]

[28] A. Hany & W. Gary. Intersections between IoT and distributed ledger, 2019. Doi: 10.1016/bs.adcom.2018.12.001.

[29] Hashemi Joo, M., Nishikawa, Y. and Dandapani, K., "Cryptocurrency, a successful application of blockchain technology", Managerial Finance, Vol. 46 No. 6, 715-733,2020, doi:10.1108/MF-09-2018-0451

[30] Decrypt, "What Are Privacy Coins? Monero, Zcash, and Dash Explained - Decrypt" [Online]. Available: https://decrypt.co/resources/what-are-privacy-coins-monero-zcash-and-dash-explained [Accessed: 30-Jun-2021]

[31] Ethereum, "Decentralized finance (DeFi)" [Online]. Available: https://ethereum.org/en/defi/ [Accessed: 30-Jun-2021]

[32] Investopedia, "DeFi (Decentralized Finance) Definition and Why it Matters" [Online]. Available: https://www.investopedia.com/decentralized-finance-defi-5113835 [Accessed: 29-Jun-2021]

[33] Cryptopedia, "Non-Fungible Tokens (NFTs): Crypto Collectibles | Gemini" [Online]. Available: https://www.gemini.com/cryptopedia/nft-non-fungible-token-crypto-collectibles [Accessed: 29-Jun-2021]

[34] M. Takemiya and B. Vanieiev, "Sora Identity: Secure, Digital Identity on the Blockchain," 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), 582-587,2018, doi: 10.1109/COMPSAC.2018.10299.

[35] Consensys, "Zug Digital ID: Blockchain Case Study for Government Issued Identity" [Online]. Available: https://consensys.net/blockchain-use-cases/government-and-the- public-sector/zug [Accessed: 17-08-2020].

[36] N.Hackius, M.Petersen "Blockchain in Logistics and Supply Chain : Trick or Treat?" , Epubli, 23, 2017, doi:10.15480/882.1444

[37] Gao, Ying & Pan, Qiaofeng & Liu, Yangliang & Lin, Hongliang & Chen, Yijian & Wen, Quansi. (2021). The Notarial Office in E-government:A Blockchain-Based Solution. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3066184.

[38] Barbieri, Maurice, and Dominik Gassen. "Blockchain–can this new technology really revolutionize the land registry system." Responsible Land Governance: Towards an Evidence Based Approach: Proceedings of the Annual World Bank Conference on Land and Poverty. 2017

[39] Agbo, C.C.; Mahmoud, Q.H.; Eklund, J.M. Blockchain Technology in Healthcare: A Systematic Review. Healthcare 2019, 7, 56, doi:10.3390/healthcare7020056

[40] Anokye Acheampong AMPONSAH, Adebayo Felix ADEKOYA and Benjamin Asubam WEYORI, "Blockchain in Insurance: Exploratory Analysis of Prospects and Threats" International Journal of Advanced Computer Science and Applications(IJACSA), 12(1), 2021. http://dx.doi.org/10.14569/IJACSA.2021.0120153

[41] N. Kshetri, J. Voas, "Blockchain-Enabled E-Voting," in IEEE Software, 35(4), 95-99, 2018, doi: 10.1109/MS.2018.2801546.

[42] R. Hanifatunnisa, B. Rahardjo, "Blockchain based e-voting recording system design," 2017 11th International Conference on Telecommunication Systems Services and Applications (TSSA), 2017, 1-6, doi: 10.1109/TSSA.2017.8272896.

[43] Andoni, V. Robu, D.Flynn, S. Abram, D. Geach, D.Jenkins, P. McCallum, A.Peacock,"Blockchain technology in the energy sector: A systematic review of challenges and opportunities",Renewable and Sustainable Energy Reviews, 100,143-174,2019,doi:10.1016/j.rser.2018.10.014.

[44] K. Mannaro, A. Pinna, M. Marchesi, "Crypto-trading: Blockchain-oriented energy market," 2017 AEIT International Annual Conference, 1-5,2017, doi: 10.23919/AEIT.2017.8240547.

[45] H. Huawei et al., When Blockchain Meets Distributed File Systems: An Overview, Challenges, and Open Issues. IEEE Access, 7, 2020 10.1109/ACCESS.2020.2979881.

[46] M. Chuan et al., When Federated Learning Meets Blockchain: A New Distributed Learning Paradigm, IEEE Access, 7, 1189, 2020. Doi: 10.1109/ACCESS.2020.2971189.

Ensemble Learning of Deep URL Features based on Convolutional Neural Network for Phishing Attack Detection

Seok-Jun Bu[*,1], Hae-Jung Kim[2]

[1]Deptartment of Computer Science, Yonsei University, Seoul 03722, Korea

[2]Department of Computer Science, Kyungil University, Daegu 38428, Korea

A R T I C L E   I N F O

A B S T R A C T

The deep learning-based URL classification approach using massive observations has been verified especially in the field of phishing attack detection. Various improvements have been achieved through the modeling of character and word sequence of URL based on convolutional and recurrent neural networks, and it has been proven that an ensemble approach of each model has the best performance. However, existing ensemble methods have limitations in effectively fusing the nonlinear correlation between heterogeneous features extracted from characters and the sequence of sub-domains. In this paper, we propose a convolutional network-based ensemble learning approach to systematically fuse syntactic and semantic features for phishing URL detection. By learning the weights that integrating the heterogeneous features extracted from the URL, an ensemble rule that guarantees the best performance was obtained. A total of 45,000 benign URLs and 15,000 phishing URLs were collected and 10-fold cross-validation was conducted for quantitative validation. The obtained classification accuracy of 0.9804 indicates that the proposed method outperforms the existing machine learning algorithms and provides plausible solution for phishing URL detection. We demonstrated the superiority of the proposed method by receiver-operating characteristic (ROC) curve analysis and the case analysis and confirmed that the accuracy improved by 1.93% compared to the latest deep model.

1. Introduction

Network security based on information technology for protecting personal information and system resources from various types of threats may be defined through policies and methods. Various methods for network administrator have been developed to protect networks and cyber assets including detection mechanism against active attacks [1]. However, few studies have been conducted to analyze the characteristics of phishing attacks, which steal entire input information from users. Phishing attack in its broadest sense can be defined as a scalable act of deception whereby impersonation is used by an attacker to obtain the information from an individual [2]. Considering that the most common form of online phishing attack is malicious hyperlinks embedded in messages, the recent technological trend in which personal connections are reinforced due to the explosive growth of social media services is particularly vulnerable [3].

Existing security systems primarily conduct rule-based detection mechanism using phishing databases to identify malicious URLs [4]. However, phishing URLs based on web applications have zero-day exploit characteristics that frequently involve novel attack instances, as URLs can be generated very conveniently in such applications. For this reason, phishing URLs hardly detected by predefined databases or simple detection rules [2, 5, 6].

Meanwhile, previous study based on ensemble of the convolutional neural network (CNN) and recurrent neural network (RNN) for the modeling the character and word-level features found that classification of malicious URLs was improved [7,8].

In Figure 1, we visualize the phishing URLs into feature space generated by the t-SNE dimension reduction method. Blue and red dots represent normal URLs and phishing URL instances, respectively. The Euclidean distance was determined based on the similarity of character combinations constituting the URL, and a cluster of short and regular URLs was mainly formed at the

*Corresponding Author: Seok-Jun Bu, Yonsei University, sjbuhan@yonsei.ac.kr

bottom. On the other hand, in the center, instances where it is difficult to distinguish between normal and phishing URLs due to subdomains are intricately confused.
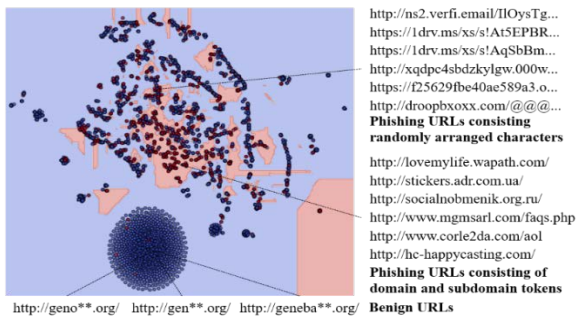

Figure 1: The feature space of phishing URLs and the necessity of ensemble learning

Phishing URL features can be distinguished into a syntactic feature consisting of a sequence of randomly arranged characters and a semantic feature consisting of a sequence of domain and subdomain words. In the existing deep learning-based ensemble approach, a simple rule-based ensemble that averages the output of syntactic-semantic convolution and recurrent networks at log-scale was applied, but it showed the limitation to effectively model the complex nonlinear correlation of features and resulted the degradation of accuracy and recall.

Taken together, we propose an ensemble learning network based on CNN that can systematically utilize the syntactics and semantics of URL features using CNN and RNN. The proposed ensemble network is a deep learning algorithm that can extract effective features for phishing URL classification using filter operations that can be trained using data. The joint learning of ensemble rule based on deep representations of URLs provides plausible solution for phishing detection. We collected a total of 45,000 benign URLs and 15,000 phishing URLs and the proposed method was validated through 10-fold cross-validation, and chi-squared test. The analytic results indicated the best performance among the machine learning-based phishing detector. To the best of our knowledge, this is the first attempt that convolutional neural network is incorporated to learn the ensemble rule for phishing detection. The main findings of this research can be summarized as follows:

- The convolutional neural network works well for learning the ensemble rule of fusing heterogeneous features of URL representations, resulting the best accuracy and recall for phishing detection.
- We categorized the features of URLs into character and word levels, and demonstrated the convolutional and recurrent neural networks to effectively model each feature.

The remainder of this paper is organized as follows. In Section 2, we review the previous URL modeling methods based on machine learning and clarify the contributions of this paper by discussing the differences between them. In Section 3, we illustrate how the heterogeneous URL features are extracted by the deep learning and fused with convolutional neural network. The performance of the model is evaluated in Section 4 through various experiments, including the 10-fold cross-validation and ROC curve analysis. Finally, section 5 concludes the paper with some discussion of future directions.

2. Related Works

Previous studies on phishing URL classification can be classified into the following categories as summarized in Table 1: those on phishing URL detection based on the blacklist, which were mainly performed before 2010; those on modeling of words extracted from the text based on traditional machine learning; and those on text feature extraction through the latest deep learning algorithms.

The author proposed a system that extracts lexical features from the text according to ex-pert-defined rules, constructs a blacklist on known phishing URLs, and detects new phishing URLs through a simple comparison algorithm [9]. However, this method has the limitation of detecting new phishing URLs in terms of generalization performance. To confirm the validity of the machine learning method in the field of phishing URL classification, the authors applied fundamental machine learning methods including naive Bayes classifiers to the word combination found in URLs and classified phishing URLs that were not included in training datasets [10]. The authors enhanced the performance of phishing URL classification systems based on machine learning by applying a support vector machine (SVM), which is widely known to perform more complex nonlinear mapping [11]. Verma significantly increased phishing URL classification accuracy through the implementation of a random forest algorithm that was designed to perform effective modeling of hierarchical elements of lexical features in the URLs [12].

The researchers extracted semantic features from phishing URLs using a word-to-vector model capable of embedding word vectors based on their statistical meaning using deep learning algorithms. Furthermore, they applied long short-term memory (LSTM) and gated re-current unit (GRU) deep learning algorithms specialized for time series modeling, including gate operations, to enhance the phishing URL classification performance of existing modeling methods [8,13]. It was proposed a convolution-recurrent network to effectively model semantics extracted from the word-to-vector model [14].

The majority of the current research in deep learning-based phishing detection focuses mainly on optimizing the operation of the neural network [16]. In particular, the comparative study in [17] proves the superiority of the ensemble approach based on CNN variations. This motivates our decision to consider the ensemble learning approach proposed in this paper. The proposed method deviates from existing work in that it implements and learns the ensemble rule with convolutional operation based on CNN to consider the heterogeneous URL features.

Table 1: Related works on phishing URL detection with respect of URL features and modeling method.

URL Features	Method	Author
Bag-of-words	Naive Bayes	Prakash [9]
Lexical Features	Matching Rules	Ma [10]
Bag-of-words	SVM	Le [11]
Lexical Features	Random Forest	Verma [12]
Word embeddings	LSTM	Bahnsen [8]
Word embeddings	GRU	Zhao [13]
Lexical Features	Generative adversarial network (GAN)	Anand [15]
Word embeddings	CNN-LSTM	Yang [14]
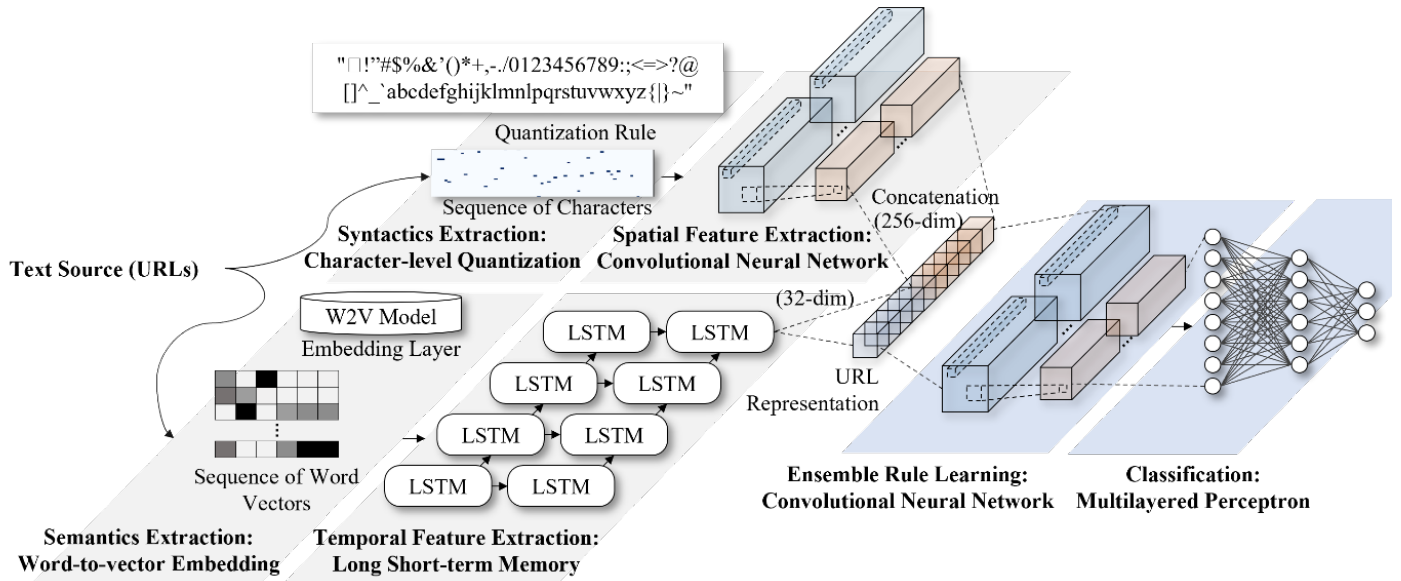
Figure 2: Proposed CNN-based ensemble learning method for phishing URL detection

3. Proposed Method

In this section, we describe the combination of the convolutional neural network and recurrent neural network to extract URL features and its ensemble learning method. Figure 2 visualize the diagrams of deep learning-based phishing classifier that extract the syntactic and semantics from URLs, as well as the proposed CNN-based ensemble learning network for the late-fusion of URL features.

3.1. Deep Learning-based Phishing URL Feature Extraction

Two types of deep learning algorithms and individual preprocessing steps were applied to conduct the modeling of syntactic and semantic features of phishing URLs. First, an integer was assigned to each character, and modeling of a low-level signal obtained through this process was performed by the CNN to model the syntactic features of random characters, including enumerated special characters, which are frequently observed in phishing URLs. Second, each word was embedded based on the word-to-vector model, and the modeling of a sequence of words obtained through this process was performed by the LSTM to model the semantic features of domains and sub-domains composing the internal URLs.

In detail, a preprocessing step for each character was performed to replace the characters with their unique Unicode values based on UTF-8 encoding, and an integer sequence of up to 100 characters was extracted in consideration of the average length of URL characters in the datasets collected. In total, 139 types of characters were used, and a vector in the dimension of $n \times 100 \times 139$ based on n of observations were inputted into the character-level CNN.

The convolution operation $\phi_c^l(\cdot)$ in Equation 1 applies a parameterized filter to the input vector and extracts syntactics from sequence of characters in URL. A filter size $m \times m$ is applied to the ith row and jth column nodes of the lth layer.

$$\phi_c^l(x_{ij}) = \sum_{a=0}^{m-1}\sum_{b=0}^{m-1} w_{ab}x_{(i+a)(j+b)} \tag{1}$$

The pooling operation $\phi_p^l(\cdot)$ in the lth pooling layer performs the extraction of representative value and defined as Equation 2, with the pooling distance τ in the region $k \times k$ among the input vectors, and outputs the maximum activation value from the region.

$$\phi_p^l(x_{ij}) = \max_{\tau \in R} x_{ij \times \tau} \tag{2}$$

The learning of the convolution operation is the process of optimizing the weight of the filter w that extracts the syntactics while preserving the spatial correlation between characters, and the pooling operation is based on extraction of emphasized features.

Meanwhile, representative features of URLs include semantics that can be derived from a sequence of words such as domains and sub-domains. Phishing URL classification accuracy can be enhanced through the parallel utilization of deep learning algorithms for additional modeling of a sequence of subdomains [18].

The modeling of semantics of phishing URLs was carried out through word embedding based on the word-to-vector model and LSTM deep learning algorithm application for time series modeling. Moreover, 20 words that appeared in sub-domains were additionally extracted since phishing URLs generally included various sub-domains. Each word was replaced as vectors in 32 dimensions using the word-to-vector model, and URLs formed as $n \times 20 \times 32$ sized vector according to n observations were input in the phishing word-level LSTM.

The LSTM network is a type of RNN in which three types of nonlinear gates are implemented. The LSTM $\phi_L^l(\cdot)$ performs the time-series modeling of sequence of domain and subdomains.

$$\phi_L^l(x_{ij}) = o_t \odot \tanh(c_t) \tag{3}$$

The input gate(i), forget gate (f), output gate (o), and LSTM cell state (c) were defined based on the input domain sequence of

x = (x(t),…, x(t-ω)) with word sequence length ω, as shown in Equation 4. b, σ and ⊙ refer to the bias added to each neural network, the sigmoid activation function of neural networks, and Hadamard multiplication, respectively.

$$\begin{aligned}
i_t &= \sigma(W_{ix}x(t) + W_{im}h_{t-1} + b_i) \\
f_t &= \sigma(W_{xf}x(t) + W_{hf}h_{t-1} + b_f) \\
o_t &= \sigma(W_{xo}x(t) + W_{ho}h_{t-1} + b_o) \\
g_t &= tanh(W_{xh}x(t) + W_{hc}h_{t-1} + b_o) \\
c_t &= f_t \odot c_{t-1} + i_t \odot g_t
\end{aligned} \tag{4}$$

3.2. Ensemble Learning based on the Convolutional Network

The proposed ensemble network utilizes the deep URL representations with a size of (n × 256) and (n × 32) with n observations from the intermediate layer of CNN and LSTM in Section 3.1. Contrary to the existing CNN-LSTM ensemble-based phishing URL detector, the model is optimized to weight the outputs from multi-level URL representations.

The character-level and word-level representations of phishing URL derived from the character-level CNN and word-level RNN were concatenated to form a vector of size (n × 288). The proposed fusion neural network was trained to minimize errors that might occur in the process of mapping the input vector to the benign or phishing URLs.
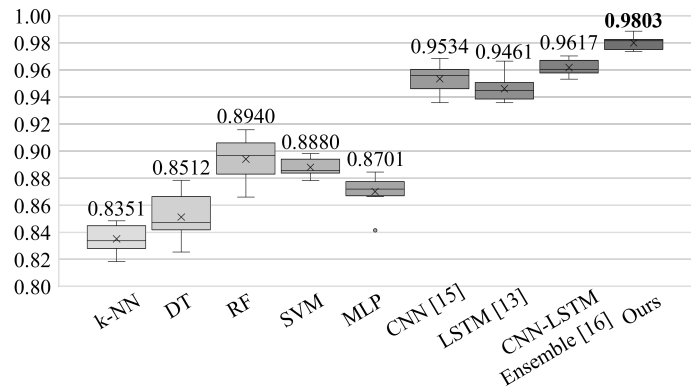
Figure 3: Results of 10-fold cross validation comparing the proposed ensemble learning with other existing methods (k-NN: k-nearest neighbor; DT: decision tree; RF: random forest; SVM: support vector machine; MLP: multilayered perceptron; CNN: convolutional neural network; LSTM: long short-term memory).

As the proposed method contains a convolution layer in order to systematically fuse level-based features, effective features are selected from input vectors from 288 dimensions and output the predictive label $\hat{y}$, as shown in Equation 5.

$$p(\hat{y}_i|x_i) = argmax \frac{exp(\phi^{l-1}(x_i)w^l + b^l)}{\sum exp(\phi^{l-1}(x_i)w^l + b^l)} \tag{5}$$

At this stage, the Softmax function, which is an activation function of the neural network, was applied to facilitate the encoding of the output vector at the probability of [0,1] range and to promote the differentiation process that operates during the optimization of the loss function. The entire mapping results obtained from inputs to outputs in the character-level and semantic-level neural networks, including the proposed ensemble network, is differentiable and can be learnt by the massive URL observations.

The entire weights of neural networks are tuned by applying a backpropagation algorithm based on gradient descent to the cross-entropy loss function L_{CE} shown in Equation 6.

$$L_{CE} = -\sum_i y_i log(\hat{y}_i) \tag{6}$$

The proposed ensemble network, which fuses features according to deep URL representations, performs the optimization of ensemble rules in consideration of joint learning of CNN and LSTM.
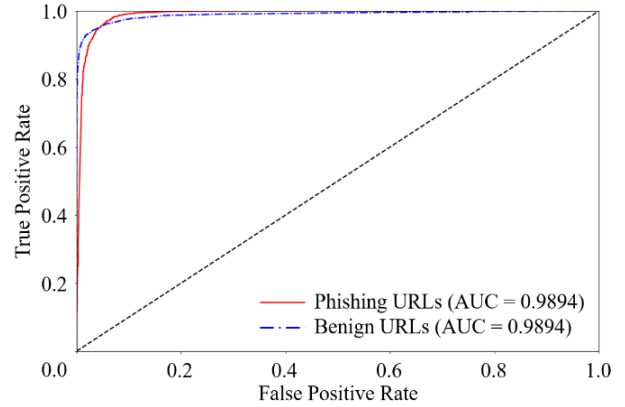
Figure 4: ROC curve and area under the curve (AUC) of classification result based on ensemble learning approach.

4. Experimental Results

4.1. Phishing URL Dataset

A total of 45,000 of URLs were collected by crawling method and 15,000 of phishing URLs were collected from Phishtank [19,20] where provides a blacklist of phishing URLs. Benign URLs were mainly collected from open directory project(ODP), which is the URL database to categorize URLs. The number of observations was intentionally adjusted because of the data imbalance issue, to reflect the conditions in which the number of phishing URLs was much lower than that of benign URLs. We noted the length of the URL is one of the critical phishing features, given that the average lengths of phishing and benign URLs are 75.74 and 35.83, respectively.

4.2. Phishing URL Classification Performance

Figure 3 shows the result of 10-fold cross validation on the proposed ensemble network and other machine learning-based models to verify the phishing URL classification performance of the proposed method. The average accuracy of the random forest algorithm, CNN, and RNN were achieved 0.8940, 0.9534, and 0.9461, respectively, with the CNN and RNN exhibiting significantly higher accuracy performance than the random forest algorithm.

The comparative result based on the ensemble of CNN [7] and RNN [8] achieved 0.9641. Based on the performance improvement, we confirmed the significance of proposed ensemble learning approach in consideration of character and word-level URL modeling. Regarding the proposed ensemble network designed to fuse Multi-level URL features, the best classification accuracy was achieved as 0.9803.

Table 2: Qualitative evaluation of complementarity based on the case analysis of CNN and LSTM (0: benign, 1: phishing).

Category		URL (accessed date: 19-10-2020)	CNN	LSTM
CNN Advantages	Phishing	https://1drv.ms/xs/s!AhtvzT3KrwqMZzLMKnTc8clHnRA?wdFormId=%7BA0F7982D%2D71A4%2D4DE0%2DB4C4%2DC16A0F044	0.9874	0.7385
	Benign	http://market.security***.net	0.0031	0.8441
LSTM Advantages	Phishing	http://bitcoin24-wallet.site	0.0722	0.9837
	Benign	http://www.knightfeatu***.com/kfweb/content/features/kffeatures/puzzlesandcrosswords/kf/sudoku/sudoku_classic/sudoku_classic.html	0.8384	0.0073
Misclassified	Benign	http://archives.seattletimes.nwsou***.com/cgi/bin/texis.cgi/web/vortex/display?slug=will&date=199903	0.8815	0.8764
	Phishing	http://tesla-present.site/ethereum/	0.0584	0.0354

Table 3: A confusion matrix for phishing URL classification based on the ensemble network

		Predicted (w/o ensemble learning)		
		Benign	Phishing	Recall
Actual	Benign	9035 (8902)	74 (207)	**0.9919 (0.9773)**
	Phishing	115 (266)	2776 (2625)	**0.9602 (0.9080)**
	Precision	**0.9874 (0.9710)**	**0.9740 (0.9269)**	Accuracy: 0.9843 (0.9606)

Since it is essential to minimize false negatives and improve recall in the field of phishing URL detection, the ROC curve and AUC are described in Figure 4. Table 3 summarizes the classification results based on the model that exhibited the best accuracy. Considering the false negatives and the recall of phishing instance of 0.9602, it is inferred that additional modeling should be carried out mainly focusing on the generalization strength of the model.

4.3. Discussion

Table 2 indicates the advantages of each model based on practical classification cases, to aid in the classification of the performance of deep learning models according to multi-level URL representations. The two upper rows show the robustness against random character enumeration of CNN. The character-level CNN classified URL as phishing instance with a probability of 0.9874, considering that a phishing URL feature is hidden in the sequence of random characters. On the other hand, in the second case, the word-based LSTM misclassified benign URL as phishing with a probability of 0.8441 because the benign words are included in the URL.

The word-level LSTM supplements the entire system by reflecting a sub-domain that was not used by the character-level CNN in the form of words. The LSTM was able to classify certain words such as 'security' and 'bitcoin' based on the massive observations that such words are frequently used in phishing URLs. The CNN, however, misclassified benign URLs as phishing based on the number of sub-domains.

5. Concluding Remarks

5.1. Conclusion

This study introduced a character-level CNN and word-level RNN for phishing URL representation and proposed an ensemble network that can effectively fuse the syntactics and semantics of phishing URLs extracted from each model. The proposed ensemble network, implemented as convolutional neural network, provide the plausible solution of parameterization and optimization of the ensemble rule. Specifically, it exhibited a classification accuracy of 0.9804, which is the highest compared to other machine learning methods including deep learning models.

5.2. Future Work

Further studies should be performed to guarantee the generalization strength of model, in consideration of zero-day attack characteristics of phishing attacks. The latest deep learning algorithms, such as one-shot learning, should be thoroughly examined. Meanwhile, the scope of this study is limited to the modeling of syntactics and semantics of URL and optimizing the ensemble rule. In this regard, a symbolic AI approach to fully exploit and utilize the domain knowledge is promising. In addition, the neural-symbolic integration approach that can calibrate the deep learning model should be additionally considered in the future studies.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] H.-J. Kim, Image-based malware classification using convolutional neural network, Springer: 1352–1357, 2017, doi:10.1007/978-981-10-7605-3_215.
[2] S.-J. Bu, S.-B. Cho, "Deep Character-Level Anomaly Detection Based on a Convolutional Autoencoder for Zero-Day Phishing URL Detection," Electronics, **10**(12), 1492, 2021, doi:10.3390/electronics10121492.
[3] V. Suganya, "A review on phishing attacks and various anti phishing techniques," International Journal of Computer Applications, **139**(1), 20–23, 2016, doi:10.5120/ijca2016909084.
[4] K.L. Chiew, K.S.C. Yong, C.L. Tan, "A survey of phishing attacks: Their types, vectors and technical approaches," Expert Systems with Applications, **106**, 1–20, 2018, doi:10.1016/j.eswa.2018.03.050.

[5] I. Qabajeh, F. Thabtah, F. Chiclana, "A recent review of conventional vs. automated cybersecurity anti-phishing techniques," Computer Science Review, **29**, 44–55, 2018, doi:10.1016/j.cosrev.2018.05.003.

[6] S.-J. Bu, S.-B. Cho, "Integrating Deep Learning with First-Order Logic Programmed Constraints for Zero-Day Phishing Attack Detection," in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE: 2685–2689, 2021, doi: 10.1109/ICASSP39728.2021.9414850.

[7] H. Le, Q. Pham, D. Sahoo, S.C.H. Hoi, "URLNet: Learning a URL representation with deep learning for malicious URL detection," ArXiv Preprint ArXiv:1802.03162, 2018, doi:10.475/123_4.

[8] A.C. Bahnsen, E.C. Bohorquez, S. Villegas, J. Vargas, F.A. González, "Classifying phishing URLs using recurrent neural networks," in 2017 APWG symposium on electronic crime research (eCrime), IEEE: 1–8, 2017, doi:10.1109/ECRIME.2017.7945048.

[9] P. Prakash, M. Kumar, R.R. Kompella, M. Gupta, "Phishnet: predictive blacklisting to detect phishing attacks," in 2010 Proceedings IEEE INFOCOM, IEEE: 1–5, 2010, doi:10.1109/INFCOM.2010.5462216.

[10] J. Ma, L.K. Saul, S. Savage, G.M. Voelker, "Beyond blacklists: learning to detect malicious web sites from suspicious URLs," in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 1245–1254, 2009, doi:10.1145/1557019.1557153.

[11] A. Le, A. Markopoulou, M. Faloutsos, "Phishdef: Url names say it all," in 2011 Proceedings IEEE INFOCOM, IEEE: 191–195, 2011, doi:10.1109/INFCOM.2011.5934995.

[12] R. Verma, K. Dyer, "On the character of phishing URLs: Accurate and robust statistical learning classifiers," in Proceedings of the 5th ACM Conference on Data and Application Security and Privacy, 111–122, 2015, doi:10.1145/2699026.2699115.

[13] J. Zhao, N. Wang, Q. Ma, Z. Cheng, "Classifying malicious URLs using gated recurrent neural networks," in International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, Springer: 385–394, 2018, doi:10.1007/978-3-319-93554-6_36.

[14] W. Yang, W. Zuo, B. Cui, "Detecting malicious URLs via a keyword-based convolutional gated-recurrent-unit neural network," IEEE Access, **7**, 29891–29900, 2019, doi:10.1109/ACCESS.2019.2895751.

[15] A. Anand, K. Gorde, J.R.A. Moniz, N. Park, T. Chakraborty, B.-T. Chu, "Phishing URL detection with oversampling based on text generative adversarial networks," in 2018 IEEE International Conference on Big Data (Big Data), IEEE: 1168–1177, 2018, doi:10.1109/BigData.2018.8622547.

[16] F. Tajaddodianfar, J.W. Stokes, A. Gururajan, "Texception: A character/word-level deep learning model for phishing URL detection," in ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE: 2857–2861, 2020, doi:10.1109/ICASSP40776.2020.9053670.

[17] D. Vasan, M. Alazab, S. Wassan, B. Safaei, Q. Zheng, "Image-Based malware classification using ensemble of CNN architectures (IMCEC)," Computers & Security, **92**, 101748, 2020, doi:10.1016/j.cose.2020.101748.

[18] Q. Li, M. Cheng, J. Wang, B. Sun, "LSTM based phishing detection for big email data," IEEE Transactions on Big Data, 2020, doi:10.1109/TBDATA.2020.2978915.

[19] L.L.C. OpenDNS, "PhishTank: An anti-phishing site," Online: https://Www.Phishtank. Com, 2016 (accessed: 1 Oct. 2021).

[20] Q. Cui, G.-V. Jourdan, G. V Bochmann, R. Couturier, I.-V. Onut, "Tracking phishing attacks over time," in Proceedings of the 26th International Conference on World Wide Web, 667–676, 2017, doi:10.1145/3038912.3052654.

ASTES

Reading Acquisition Software for Portuguese: Preliminary Results

Ana Sucena*, Ana Filipa Silva, Cristina Garrido, Cátia Marques

Research and Intervention Reading Centre, Polytechnic Institute of Porto, Porto, 1800-412, Portugal

A R T I C L E I N F O	A B S T R A C T
	The persistent difficulties in reading and spelling acquisition are a risk factor for learning motivation. Play-like intervention tools have been developed to face these difficulties. "I read" is a software that seeks to develop and introduce systematic reading and spelling skills training in a playful and complementary way. This software is intended for children at the beginning of their school journey, as well as for those who reveal reading and/or spelling difficulties. This article intends to present the goals and the structure of this software, as well as the preliminary results of its implementation and game enjoyment with 244 children between 5 and 7 years old. Results indicate that 58% of the participants completed the activities dedicated to alphabetical decoding, and 42% were able to reach the last stage of the game, dedicated to orthographic decoding. Regarding the enjoyment with the software, 96% of the participants classify the games as fun games. In conclusion, training with this software revealed to be beneficial for reading and spelling skills promotion, as well as to increase the overall enjoyment and motivation for learning.

1. Introduction

Research highlights persistent difficulties in reading acquisition as a key risk factor for school motivation, learning and children's self-esteem [1]. These difficulties lead to long-term negative consequences, such as behavior problems, depression and other mental health issues [2]. If reading disabilities are not early addressed, difficulties tend to generalize to other domains, exposing students to consecutive experiences of failure, thereby diminishing their motivation to learn [3], jeopardizing future knowledge acquisition [4]. In turn, when these difficulties are identified early and lead to a prompt intervention, the likelihood of reversing failure trajectories is very high [5], [6].

Thus, early intervention to face learning difficulties, namely on reading and spelling [7]-[10] is of utmost importance. Early intervention can have an even more ambitious purpose: preventing the child from experiencing learning difficulties, hence the term "at-risk" (of experiencing reading and writing learning difficulties). The risk for retention, school dropout and low motivation decreases with the adoption of early intervention focused on the key competencies for reading acquisition [11]. Phonemic awareness, letter-sound knowledge and decoding skills are the key competencies for successful reading and writing acquisition [11], [12].

As a result, the past decade has been witnessing the fast development of information technologies, together with the rapid development of serious games [13]. Serious games consist of entertaining tools with educational purposes, presented as digital games throughout which children are required to overcome numerous obstacles. The serious games primary goal is to contribute to increasing the players' motivation to acquire and expand knowledge and practising skills [13]. Educational aspects are included in the gameplay, which will be acquired and automatized by the players during the gaming process [13]. To ensure effective learning in serious games many aspects are important. Designers are not only faced with the challenge of creating appealing graphics and a compelling narrative, but also with the additional challenge of incorporating suitable pedagogical strategies into their narratives [14].

In sum, serious games are congruently described in the literature as digital games that have an explicit and carefully thought-out educational purpose and are not solely intended for amusement [15], [16]. The development of these types of games, in a motivating, challenging and playful environment has proven to be a more effective tool at conveying knowledge and achieving behavior change [14] than the traditional teaching and learning methods [17], [18]. Serious games also present an opportunity for proactive learning by involving players in an immersive learning experience where they can apply their knowledge, learn from

*Corresponding Author: Ana Sucena, asucena@ess.ipp.pt

experience, and test complex or risky scenarios in a safe and play-full environment.

Player enjoyment is the term used to express the satisfaction perceived when performing tasks and activities that require high levels of concentration and skills in digital games [19]-[21]. Enjoyment is a form of fun that serious games can promote, resulting from the challenges the players face and must be able to overcome to advance in the game [22]. Player enjoyment is an important aspect for educational serious games, as they are commonly introduced to students by professionals responsible for the teaching process who are leading students to play, rather than the students starting to play willingly or through a recommendation from friends. Therefore, promoting player enjoyment can keep them interested in continuing to play along with all the levels, continuing to develop their skills in the contents covered by the game, potentially enhancing the results from the use of these educational tools in specific domains.

The game levels are used to show the progress of users. It is a common practice for serious games to adopt levels to show increasing difficulty and complexity levels. This option is based on shaping - a behaviour changing strategy. Shaping consists of breaking tasks down into small discrete steps and training in a systematic and precise way [23]. This can be linear (e.g., you level up after earning 50 points), or exponential. Creating the levels at the right difficulty and complexity (level balancing) is important to avoid drop-outs. Difficulty can also serve as a way to motivate the users as they passed a specifically difficult level the child gets points and move forward in the game [24]. The different levels give players a direction in order to reach the goal that needs to be achieved. Players have something to accomplish within their overall experience [25].

Due to these features, serious games have been successfully applied in various domains such as medicine, military, or complex processes training, among others; example success-cases include commerce, environment and ecological behaviour, cartography, machine learning, software development, innovation, health and medical issues, tourism, energy, mobility and transportation, accessibility fashion, usability, risk management, marketing [26] and with children with learning disabilities [11].

Serious games have become increasingly popular in education, supporting children with learning disabilities [27]-[32]. It can be used as a tool to reinforce the teaching process and provide enriching learning contexts for its users [26]. Most children with learning disabilities can be successful in reading acquisition with the help of a tutor or a specialized teaching program [33]. Emotional support also plays an important role for these children while dealing with learning disabilities [33]. Several studies have shown that children prefer playing a "fun" computer game than performing a "boring" exercise [34]. The idea is that the user does something useful, for his/her knowledge. If the user has fun, she/he will probably continue to play, achieving the serious goals of the game [34]. Serious games can be a great help in these situations, not substituting, but complementing the work done by therapists [35]. Moreover, improving the life quality of children with learning disabilities has a crucial impact on their learning and future path.

The results of serious games as a tool to intervene with children with learning disabilities are supported by the empirical findings of several studies [36]. There is an increasing demand for innovative and cost-effective evidence-based solutions to the educational challenges faced by teachers and students. Reading related difficulties represent 80% of the overall learning difficulties [37], which makes it all the more important to invest in developing reading acquisition serious games.

The Graphogame reading promotion software is one of the pioneer's projects regarding serious games in learning acquisition. The Graphogame provides children with training in phoneme awareness, letter-sound and early word decoding training. This play-like software was developed in a research project, originally developed in Finland, then adapted for over 20 language versions (e.g., Finnish [9], [38]-[40]; German [41], [42]; English [12], [43]; Portuguese [11], [48]; Spanish [45]). In this computer game set, the child uses headphones and hears a sound that corresponds to a letter, word or pseudoword, and he/she must select the written option corresponding to the sound.

Recently, a new Portuguese serious game was developed – I Read [46], [47]. This serious game aims to promote phonemic awareness, letter-sound knowledge and decoding. Compared to the Portuguese Graphogame the I Read has a greater focus on letter-sound knowledge as well as on spelling tasks. For each set of letter-sound, activities are focusing on monosyllabic words (CV and CVC syllabic structure) reading and spelling. The I Read serious game intends to facilitate reading and spelling acquisition in a playful and complementary environment. In this study, we describe the I Read software, along with preliminary results of its implementation on reading and spelling skills, as well as on the player enjoyment levels.

In this study we present the I Read software that aims to practise basic literacy skills, we present the main results of its implementation, as well as, assess the children enjoyment with the game.

2. Method

2.1. Participants

244 first graders participated in this study, aged between 5 and 7 years old, studying in Portuguese public schools. All children were identified as at risk of experiencing difficulties in reading. None of the participants revealed any sensory, motor or cognitive disturbances according to school registers.

2.2. Instrument

The I Read is a recently developed computer software, for reading and spelling training with Portuguese speaking children. I Read [36], [37] focus on the development of the alphabetical principle - training letter-sound and phonemic awareness - and decoding through systematic training, along with different games. This software is intended for children at the beginning of their school path and withchildren at risk of experiencing difficulties in reading acquisition. The I Read can be played on any computer or tablet, with internet access. Each child logs in with a username and password. Information about the game is recorded online, which allows the child to resume the exact game in the next session.

The general and specific goals of the I Read software are described in table 1. There are 13 game levels. Between level one and nine, the goal is to promote the alphabetic principle and alphabetic decoding, thus all stimuli correspond to simple graphemes and have simple syllabic structure. Between level 10 and 13, the focus shifts to orthographic decoding, with stimuli composed of complex graphemes and complex syllabic structure.

Table 1: I Read goals

General goals	Specific goals
Simple vocalic grapheme (level 1)	Phonemic awareness Letter-sound knowledge Alphabetical decoding
Oral diphthongs (level 2 and 3)	Phonemic awareness Alphabetical decoding Spelling decoding
Nasal diphthongs (level 4)	Phonemic awareness Alphabetical decoding Spelling decoding
Simple consonant grapheme (level 5 to 9)	Phonemic awareness Letter-sound knowledge Alphabetical decoding Spelling decoding
Complex consonant grapheme (level 10 to 13)	Phonemic awareness Letter-sound knowledge Spelling decoding

Within each level, the difficulty increases along with the games. For example, at level five (p, t, s) there are 13 sub-levels.

The 1st and 2nd sub-levels focus on phonemic awareness and letter-sound knowledge training; from the 3rd to the 13th sub-level, the goal shifts to alphabetical decoding. The same difficulty continuum applies to the syllabic length, as words/ pseudowords presented on the initial levels are monosyllabic CV, CVG– whereas two-syllable words/ pseudowords are introduced in more advanced levels - CV.CV, CVG.CV/CV.CVG. Finally, reading tasks precede spelling tasks - from the 3rd until the 8th sub-level the focus is on reading games; spelling games appear from the 9th level.

The narrative behind the software is that of a challenge to discover planets and satellites in a galaxy, along with spatial-like music, intending to contribute to a playful appearance. Each planet represents a level, whereas satellites (around each planet) represent the game sublevels. Each game is presented with a different scenario (e.g., under the ocean; in space; on a submarine) and contains multiple trials. Each trial contains multiple stimuli, always presented along with distractors.

There are two types of game: a) identification/ reading – the child hears a sound corresponding to a letter, word or pseudoword and he/she identifies the orthographic match on the screen; b) spelling – the child hears a word or pseudoword and is required to select and order the constituent letters (presented in the screen, in random order).

In all games, there is a button represented by a loudspeaker, which the child can press whenever he/she needs to hear the target stimulus again. The accuracy feedback is provided immediately after the correct or incorrect answer (through an auditory and visual stimulus). If the answer is inaccurate, the stimulus visually trembles and disappears from the screen, thus reducing the number of stimuli on the screen (until only the right option appears on the screen). The software adapts to the performance of the player, by eliminating the possibility for repeated mistakes. As the child starts improving performance, the number of distractors increases again. It was adopted the errorless learning strategy. This strategy aims to minimize opportunities for errors, increasing the frequency at which the child encounters reinforcement. Minimizing errors also reduces the likelihood that the child engages in challenging behavior [39]. When the child selects the correct answer, a green checkmark appears on the screen (along with a "correct-like" sound). At the end of each game, there is feedback on the game's performance, through the attribution of one to three stars.

I Read can be played in two modes, free or pre-set mode. Thus, it is possible to select specific levels/planets (free mode, designed for the teacher/ adult accompanying the training) or to follow the preset outline (designed for the child playing on her own) by playing the levels in increasing difficulty, with the advantage of the software adaptation to each child learning rhythm [39].

The player enjoyment with the I Read was assessed by asking the players to classify their satisfaction with the game using a three Likert scale using smiles. If the players are very satisfied they answer three smiles, more or less satisfied two smiles and if they are not satisfied one smile.

2.3. Procedures

This software was implemented with at-risk first graders. Training occurred 3 to 4 times a week in sessions between 15 to 20 minutes, in a school context, between November and May of 2020, under the supervision of an expert in reading and spelling promotion, with specific training on the software. Each child played individually, on a computer, in a noiseless environment and/or with the use of headphones. Children played in the preset mode, according to their learning pace. The information of each player gaming sessions was recorded, so the child could resume playing at the point where he/she stayed the last time he/she played, on any computer, as long as he/she has access to his/her user name and password.

3. Results

The ultimate challenge – playing all the 13 planets and respective satellites up until the last one - was reached by 48 children (15% of the total children). On average, participants took 18 sessions to reach the final game, with a frequency of play of about once per week.

On a level-by-level analysis, all participants concluded the vowels and diphthongs levels. The percentage decreases when we analyse the alphabetic decoding and the orthographic decoding (respectively, 58 % and 42%) (Table 2).

Table 2: Distribution of participants reaching each level of difficulty

Levels	Number Participants (%)	Number Participants (%) by levels of groups
1	0	0
2	0	
3	0	
4	0	
5	50 (20.5%)	142 (58.2%)
6	27 (8.6%)	
7	21 (6.7%)	
8	25 (7.9%)	
9	19 (6.03%)	
10	32 (10.2%)	102 (41.8%)
11	8 (2.5%)	
12	14 (4.4%)	
13	48 (15.2%)	

Regarding player enjoyment, 233 (95.5%) participants found fun to play the games (Figure 1).
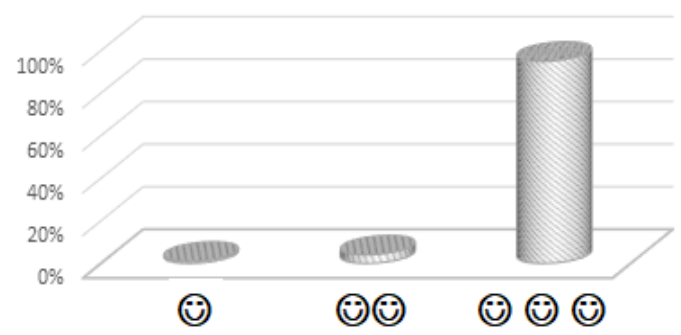

Figure 1: Player's enjoyment

4. Discussion

Ensuring literacy acquisition in early grades is a fundamental investment. In this study, we intended to (i) present the I Read software that aim to practice basic literacy skills, (ii) the main results of its implementation as well as (iii) children enjoyment with the game. The development of serious games has a large impact on education, prompting the need for upgraded policy making, curriculum development, and teacher training. Providing teachers and educational technicians with research technology empowers them to take their places as qualified experts and advisers that policymakers need for making informed choices for better education quality [4].

The results of this study show that all participants managed to finish the first stage, focused on the Alphabetic Principle - vowels, diphthongs, and oral and nasal diphthongs. Around 60% of children reached the second stage – focused on alphabetic decoding. These are promising results, as these students finish the first grade dominating the alphabetic decoding. Having started the 1st grade with weak pre-reading skills, these children conclude the 1st grade with reasonable reading skills, thus moving on to the 2nd grade without the at-risk label. The I Read software positively

impacts the attentional and motivational level [11], which are, for at-risk children, often negatively affected by the growing awareness of their difficulties [8].

There is, yet, potential to increase the average results. Regarding the last stage - focusing on orthographic decoding – the results are modest. This stage was attained by 42% of the children and only 15% managed to finish all games. It is important to develop strategies to increase the percentage of children who finish all games. One possible strategy could be to count on the support of an adult during the time when the child is playing I Read. Complementary activities with the supervision of an adult in order to promote orthographic decoding are also an important strategy to help children facing difficulties. The results indicate that children at risk for reading acquisition difficulties benefit from a remedial reading intervention that is enriched with the use of the I Read software at the very beginning of the first grade.

Regarding player enjoyment, the results confirm the playful nature of the software as an effective tool to intervene with children with learning disabilities. These results also encourage us to continue the intervention, in particular with children who are facing difficulties. The focus of the software on phonemic awareness, letter-sound knowledge and decoding skills, which are the key competencies for successful reading and writing acquisition [8], [9], [11], [12], [16] along with the playful environment, proved to be highly beneficial to promote reading acquisition. Thus, a simple learning game can provide learning and playful opportunities for struggling readers.

In future studies, it is important to test the impact of the software in terms of effective learning, creating two groups of at-risk children – an intervention group that plays regularly and a comparative group that does not use the software. Also, in order to understand the role of the teacher's training in the application for the software, in future studies this variable should be controlled. Future studies could also include a thorough analysis of the intervention process to better understand the evolution of participants from session to session, along with the teachers' perspective regarding the software efficacy and a follow-up study of the participants, thus contributing for this educational tool to become even more effective for teachers and students. Since literacy is closely tied to social and economic development, quality research on this subject, along with its dissemination play a key role in the maintenance and increase of economic and social increase. Adaptations of the I Read to different Portuguese speaking countries and for other languages could be an important step for the worldwide promotion of reading acquisition with at-risk children, by focusing the training on the very specific skills, highly associated with success in reading acquisition.

5. Conclusion

Successful reading and writing acquisition is associated with an overall and long-lasting good performance in the educational, social and professional path. In the same vein, any impairment in the early reading acquisition can negatively affect the life trajectory in general. To intervene with at-risk children regarding reading acquisition is known to have high economic revenue, as a successful reading acquisition increases the odds for broad school success. Losing one year in the labour force (after school retention

caused by learning difficulties) corresponds to a loss of around $80,000, on an estimate for a college-educated male working full time and year-round, retiring at age 67 in the USA [42].

Serious games are highly effective on reading acquisition as well as, on broader terms, on the learning process. Serious games are a strong complement to more conventional instruction methods, with a higher capacity for motivating children in result of the game-like and technology-like appearance. Research has shown that serious games have the potential to prevent and remediate reading difficulties. The use of the playful-like I Read software, focusing on reading and spelling skills, has proven to be an important complementary tool for early intervention with at-risk children. It is the author's belief (and expectation) that by captivating the children's attention for reading-related training, the I Read will play an important role in the foundations of the overall enjoyment and motivation for learning.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] R. Lovio, A.Halttunen, H.Lyytinen, R Näätänen, T. Kujala, "Reading skill and neural processing accuracy improvement after a 3-hour intervention in preschoolers with difficulties in reading-related skills," Brain Research, **11**, 42-55, 2012, doi:10.1016/j.brainres.2012.01.071

[2] A. M.Undheim, L.Wichstrøm, A. M.Sund, "Emotional and behavioral problems among school adolescents with and without reading difficulties as measured by the youth self-report: a one-year followup study," Scand. J. Educ. Res. **55**, 291–305, 2011, doi: 10.1080/00313831.2011. 576879

[3] H. Lyytinen, J.Erskine, "Early identification and prevention of reading problems," Encyclopedia on Early Childhood Development, 1–5, 2016. Retrieved from http://www.childencyclopedia.com/documents/Lyytinen-ErskineANGxp.pdf

[4] S.Raspin, R.Smallwood, S.Hatfield, L.Boesley, "Exploring the use of the ARROW literacy intervention for looked after children in a UK local authority," Educational Psychology in Practice, **35**(4), 1-13, 2019, doi:10.1080/02667363.2019.1632172

[5] M. Hall, M. Burns, "Meta-analysis of targeted small-group reading interventions," Journal of School Psychology, **66**(1), 54-66, 2018. doi:10.1016/j.jsp.2017.11.002

[6] H. Lyytinen, "New Technologies and interventions for learning difficulties: Dyslexia in Finnish as a case study". In Foresight Mental Capital and Wellbeing Project: The Government Office for Science. Government Office for Science, London: UK, 2008.

[7] P.Hatcher, C.Hulme, M.Snowling, "Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure" Journal of Child Psychology and Psychiatry, **45**(2), 338–358, 2004.

[8] U. Richardson, H. Lyytinen, "The GraphoGame method: The theoretical and methodological background of the technology-enhanced learning environment for learning to read", Human Technology, **10**, 39-60, 2014. doi:10.17011/ht/urn.201405281859.

[9] N.L. Saine, M.K.Lerkkanen, T.Ahonen, A.Tolvanen, H.Lyytinen, "Computer-assisted remedial reading intervention for school beginners at risk for reading disability", Child Development, **82**, 1013-28, 2011doi:10.1111/j.1467-8624.2011.01580.x

[10] H. Wimmer, H.Mayringer, "Dysfluent reading in the absence of spelling difficulties: A specific disability in regular orthographies", Journal of Educational Psychology, **94**(2), 272-277, 2002.

[11] A.Sucena, , J.Cruz, F.L.Viana, A.F.Silva, "Graphogame português alicerce:

[12] F.E. Kyle, J. Kujala, U. Richardson, H. Lyytinen, U.Goswami, "Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG rime and GG phoneme", Reading Research Quarterly, **48**, 61-76, 2013, doi:10.1002/rrq.038.

[13] Y.Zhonggen, "A Meta-Analysis of Use of Serious Games in Education over a Decade. Hindawi" International Journal of Computer Games Technology 2019, 4797032, doi:10.1155/2019/4797032

[14] O.Troyer, F.Van Broeckhoven, , J.Vlieghe, "Linking serious game narratives with pedagogical theories and pedagogical design strategies," J Comput High Educ **29**, 549–573, 2017, doi:10.1007/s12528-017-9142-4.

[15] C.C. Abt Serious Games University Press of America, 1987.

[16] M.A. Khenissi, F. Essalmi, M. Jemni, "Comparison between serious games and learning version of existing games," Procedia - Social and Behavioral Sciences, **191**, 487-494, 2015, doi:10.1016/j.sbspro.2015.04.380.

[17] N. Bäulke, M.Eckerlein, M. Dresel, "Interrelations between motivational regulation, procrastination and college dropout intentions" Unterrichtswissenschaft, **46**(4) 461-479, 2018, doi: 10.1007/s42010-018-0029-5.

[18] Y. Du, S. Chaaban, A.M. Sabah, L. Al-Thani, Wang, Active learning engagement in teacher preparation programmes—A comparative study from Qatar, Lebanon and China Asia Pacific Journal of Education, **40**(3), 283-298, 2020, doi:10.1080/02188791.2020.1717436

[19] M. Csikszentmihalyi, "Finding flow: the psychology of engagement with everyday life," Psychol. Today, 1-7, 1997.

[20] P. Sweetser, P. Wyeth Gameflow, "A model for evaluating player enjoyment in games," Comput. Entertain., **3** (2005), 1-3, doi:10.1145/1077246.1077253.

[21] J. Almeida, L. Machado, "Design requirements for educational serious games with focus on player enjoyment," Entertainment Computing, **38**(100413), 2021.

[22] M. Prensky, "Digital game-based learning," Comput. Entertain., **1**, 1-21, 2003.

[23] M. Spreckley, R. Boyd, "Efficacy of Applied Behavioral Intervention in Preschool Children with Autism for Improving Cognitive, Language, and Adaptive Behavior: A Systematic Review and Meta-analysis", The Journal of Pediatrics, 2019, doi:10.1016/j.jpeds.2008.09.012

[24] V.Insley, D. Nunan, "Gamification and the online retail experience," International Journal of Retail & Distribution Management, **42**(5), 340–351, 2014, doi:10.1108/IJRDM-01-2013-0030

[25] L.M. Lubbe, C.Gerritsen, M. Klein, K. Hindriks, "Empowering vulnerable target groups with serious games and gamification," Entertainment Computing, **38**(100402), 1-27, 2021, doi 10.1016/j.entcom.2020.100402

[26] Baptista, G., Oliveira, T., "Gamification and serious games: A literature meta-analysis and integrative model", Comput. Hum. Behav., **92**, 306–315, 2019. doi:10.1016/j.chb.2018.11.030

[27] M. Ronimus, K. Eklund, J. Westerholm, R. Ketonen, H. Lyytinen, "A mobile game as a support tool for children with severe difficulties in reading and spelling", Journal of computer assisted learning, **36**, 1011–1025, 2020, doi: 10.1111/jcal.12456.

[28] M. Niemelä, T. Kärkkäinen, S. Äyrämö, M. Ronimus, U. Richardson, H. Lyytinen, "Game learning analytics for understanding reading skills in transparent writing systemritish", Journal of Educational Technology, **51**(6), 2376–2390, 2020. doi:10.1111/bjet.12916

[29] K. Khowaja, S. Salim, "A framework to design vocabulary-based serious games for children with autism spectrum disorder (ASD)", Universal Access in the Information Society **19**, 739–781, 2020 doi:10.1007/s10209-019-00689-4.

[30] K.D. Godinez, L.S. Gaytán-Lugo, P.A. Alcaraz-Valencia, Rocío Maciel Arellano, "Evaluation of a Low Fidelity Prototype of a Serious Game to Encourage Reading in Elementary School Children", CLIHC '17, November 8–10, 2017, doi:10.1145/3151470.3156640.

[31] S. Serret, S. Hun, S. Thümmler, P. Pierron, A. Santos, J. Bourgeois, F Askenazy, "Teaching Literacy Skills to French Minimally Verbal School-Aged Children with Autism Spectrum Disorders with the Serious Game SEMA-TIC: An Exploratory Study", Frontiers Psychology, **8**(1523), 2017, doi: 10.3389/fpsyg.2017.01523.

[32] J.M. Thomson, N. Foldnesb, P.H. Uppstadc , M. Njåc , O. J. Solheimc , K. Lundetræ, "Can children's instructional gameplay activity be used as a predictive indicator of reading skills?", Learning and Instruction, **668**, 101348, 2020 doi:10.1016/j.learninstruc.2020.101348.

[33] A. J. Alcázar, E. Venegas, S. Criollo, S.L. Mora An Approach to Accessible

software de apoio a crianças disléxicas. In M. J. Gomes, A. J. Osório & L. Valente (Eds.), Atas da IX Conferência Internacional de TIC na Educação (pp. 396-405). Braga: Universidade do Minho/Centro de Competência em TIC na Educação, 2015.

Serious Games for People with Dyslexia Sustainability 2021, **13**, 2507. doi:10.3390/su13052507.

[34]   O. Gaggi, C. E. Palazzi, M. Ciman, G. Galiazzo, S. Franceschini, M. Rufino, S. Gorgi, A. Facoetti, Serious Games for Early Identification of Developmental Dyslexia, Comput. Entertain **15**, 2, Article 4 (February 2017), 1-24. doi:10.1145/2629558.

[35]   R.M. Tomé, J.M. Pereira, M. Oliveira, "Using serious games for cognitive disabilities". In Proceedings of the International Conference on Serious Games Development and Applications, Berlin, Germany, 9–10, October 2014.

[36]   M. J. Dondlinger, "Educational video game design: A review of the literature", Journal of Applied Educational Technology, **4**(1), 21–31, 2007.

[37]   S. Shaywitz, "Entendendo a dislexia: um novo e completo programa para todos os níveis de problemas de leitura", Porto Alegre: Artmed. 2006.

[38]   N.L. Saine, M.K. Lerkkanen, T.Ahonen, A. Tolvanen, H. Lyytinen, "Predicting word-level reading fluency outcomes in three contrastive groups: Remedial and computer-assisted remedial reading intervention, and mainstream instruction", Learning and Individual Differences **20**, 402-414, 2010 doi:10.1016/j.lindif.2010.06.004

[39]   V. Markham, A. Giles, G. Davies, V. Adshead, G. Tamiaki, R. May, Applications of within-stimulus errorless learning methods for teaching discrimination skills to individuals with intellectual and developmental disabilities: A systematic review", Research in Developmental Disabilities, **97**, 103521, 2020, doi: 10.1016/j.ridd.2019.103521

[40]   M.Ronimus, H.Lyytinen, "Is school a better environment than home for digital game-based learning? The case of GraphoGame". Human Technology, **11**(2), 123–147, 2015 doi:10.17011/ht/urn.201511113637

[41]   S.Brem, S. Bach, K.Kucian, J. V. Kujala, T. K.Guttorm, E. Martin, et al."Brain sensitivity to print emerges when children learn letter-speechbsound correspondences", Proc. Natl. Acad. Sci. U.S.A. **107**, 7939–7944, 2010, doi:10.1073/pnas.0904402107

[42]   S. Bach, U. Richardson, D. Brandeis, E. Martin, S. Brem, "Print-specific multimodal brain activation in kindergarten improves prediction of reading skills in second grade". NeuroImage, **82**, 605–615, 2013. doi: 10.1016/j.neuroimage.2013.05.062

[43]   J.Worth, J. Nelson, J. Harland, D. Bernardinelli, B. Styles, "GraphoGame Rime: Evaluation report and executive summary", Slough, UK: National Foundation for Educational Research, 2018.

[44]   A. Sucena, A. F.Silva, F.L.Viana, "Intervenção precoce nas dificuldades de aprendizagem da leitura com recurso ao software Graphogame", Revista Digital do Programa de Pós-Graduação em Letras da PUCRS Porto Alegre, **9**(2), 200-212, 2016.

[45]   R. Rosas, J.P.Escobar, M.P. Ramírez, A.Meneses, A. Guajardo, "Impact of a computer-based intervention in Chilean children at risk of manifesting reading difficulties", Infancia y Aprendizaje, **40**(1), 158–188, 2017. doi:10.1080/02103702.2016.1263451

[46]   A. Sucena, Eu Leio, Escola Virtual, 2015. Retrieved from https://www.escolavirtual.pt/Pagina-Especial/euleio.htm

[47]   A. Sucena, C. Marques, A.F. Silva, A. Ramalho, R. Machado, R. Santos, C. Santos, C. Garrido, H. Freitas, I. Santos, I. Rangel, M. J. Mata, "I Read: Reading and writing skills promotion software". In P. Arantes, & J. V. Sá (Eds.), ARTECH 2019 Proceedings of the: 9th International Conference on Digital and Interactive Arts. Braga: Portugal, 2019. ISBN 978-1-4503-7250-3

[48]   A. Sucena, J.F. Carneiro, M.T. Restivo, "Serious games for reading acquisition: A tentative prototype", Teaching and Learning in a Digital World, **2**, 686-692, 2018. doi:10.1007/978-3-319-73204-6_75

ASTES

TETRA™ Techniques to Assess and Manage the Software Technical Debt

Boris Kontsevoi[1,*], Sergei Terekhov[2]

[1]*Boris Kontsevoi, Intetics Inc, President & CEO, Naples, 34108, USA,*

[2]*Sergei Terekhov, Intetics Inc, Director of Quality Assurance, Minsk, 220004, Belarus*

A R T I C L E   I N F O

A B S T R A C T

The paper examines the company's proprietary means for determining the quality of a software product and measuring its technical debt. The paper's authors explain how a software product's quality is directly correlated with the amount of varying technical debts that the end-users receive. All debts can be paid, and technical debt is no exception: one can use various parameters, techniques, and dimensions to effectively measure and optimize the quality of a software product. The authors share information about the company's proprietary method to technical debt management, which is done via the Technical dEbT Reduction plAtform, otherwise known as TETRA™. They give details about the assessment's major dimensions, tools, and measurement parameters.

1. Introduction

Whether you've been a part of the software development industry for decades or just a few years, you've probably had to explain software quality to your clients. Everybody talks about software quality, but it's understood differently from person to person. Some define quality as the level of client satisfaction, while others say it's just about meeting the customer's requirements. And in the tech world, it's more about the software being free of any defects. We find the latter to be more realistic; however, it can raise even more concerns. At Intetics, we decided to clear up any understandings by carrying out intensive research: we aimed to determine what the quality of a software product is and how it can be managed efficiently. Throughout our investigations, we discovered that a software product's quality is directly correlated to the product's amount of varying technical debts that are passed on to end-users.

Best practices within the software development industry state that technical debt is any code added now that will take more work to fix later on. This kind of code is usually added to achieve rapid gains. The source code is one of the most important aspects of software. After all, the higher the number of problems within the source code, the more redevelopment is needed [1]. That's how the software development industry understands technical debt – and we agree. However, after dealing with technical debt for many years, we started to think about it differently. After thorough examination, we came to the conclusion that technical debt is "*a combined product non-compliance with technical guidelines and business objectives that negatively impact business results.*"

Technical debt can be further divided into three categories: intentional, unavoidable, and unintentional.

*Corresponding Author: Boris Kontsevoi, boris@intetics.com

- **Intentional tech debt:** As a business strategy, this kind of debt is knowingly taken on. If intentional debt is kept at bay, it's acceptable.
- **Unavoidable tech debt:** This kind compiles as a result of external factors, such as when you need to adapt your software after a third-party system upgrade.
- **Unintentional tech debt:** It stems from negligence. It might happen if several developers are simultaneously working on one solution and one error is layered over another.

There are also various kinds of tech debt types based on the debt's cause. Some causes include people debt, automation test debt, documentation debt, and requirement debt, among others [2].

A growing volume of technical debt in the project is usually a consequence of cutting corners during the project's early stages. It can stem from minor defects in the codebase, environment setup, documentation, or chosen third-party software. And as the impact of these early errors grows, refactoring is necessary – likely rewritten or adjusted to ease further development and maintenance. The development team has to work with tech debt while the software is already developed, released, and being actively used. And if the tech debt is not addressed in time, it will gradually build up, threatening to take down the whole software product.

The costs of technical debt can be quite significant; the cost of technical debt for an average application with 300,000 lines is just over $1 million, according to a report from CAST software. What's more, the repercussions can go beyond high costs. Take, for instance, an NHS glitch in 2018 – in this case, over 10,000 patients were at risk of being given the wrong medication.

For any software product, it is natural to accumulate some amount of technical debt in the development process. Therefore, it makes academic and practical sense to manage tech debt. Various reviewed sources suggest that technical debt should be managed within the Agile Triangle, where there are three different dimensions - budget, time, and scope [3] - conflicting since it is not feasible to develop high-quality software with a low budget in a short time. So, as shown in Figure 1, if we are focused on one corner, we are going to weaken the other two.
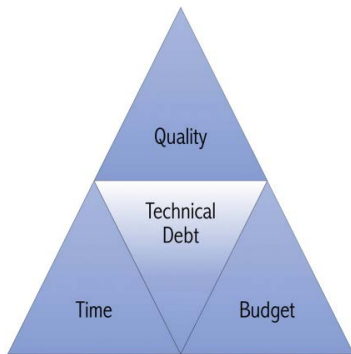
Figure 1: Agile Triangle

ᵃ Agile Triangle figure is taken from source [3].

But there are also more specific approaches that highlight the following steps:

- Identifying;
- Measuring;
- Prioritizing;
- Preventing;
- Monitoring;
- Repaying;
- Representing/Documenting;
- Communicating [4].

However, there is no commonly accepted theory of technical debt identification. Yet, technical debt is present in any real project – whether or not the developers recognize it as its own entity. That had encouraged us to propose our own different model.

While developing the model, we spent thousands of hours investigating and defining crucial aspects of software product quality that allow us to get a better perspective and measurement of technical debt. We used these dimensions to form the foundation of our platform for software quality assessment: TETRA™ (Technical dEbT Reduction plAtform). And why did we name it TETRA™? Technical debt is akin to tiny TETRA™ fish with thousands of species. That might be like small and insignificant at first glance trade-offs. Even though these "locally beautiful" quick solutions are easy to implement, they aren't the ideal solution overall. As technical debt compiles, you'll have to pay "interest" – in other words, more challenging maintainability, a less enjoyable user experience, worse productivity across your development team, and increased costs overall. We cram all those fish into one fishbowl: the TETRA™ platform. So, how do we decide what matters in software products? Is it possible to assess both the business and technical sides?

We measure and assess both sides using a specific set of tools and techniques, chosen based on original research from Intetics, best practice analysis, and expertise gained throughout multiple projects. The results enable us to calculate a software product's overall technical debt and its specific parameters. To get the big picture, we use eight dimensions to assess the product; each dimension is judged on a five-point scale. This enables uniform results and allows summary assessments to be obtained – for the product itself and for each dimension. Ranges of values were set for each metric – experimental and directly calculated, which acted as a scoring system. The final product quality assessment calculation was done via a weighted parameters system. The parameters determined the roles that each dimension would play in the overall score.

The final assessment is given within a table containing numbers and indices for each metric. The most advantageous part of using the TETRA™ methodology is that indices are introduced, rather than exact numbers. Thus, further investigation can be carried out based on the analysis of the TETRA™ Index, Metric Index, and Dimension Index.

Every value was given a letter score and color. The highest letter grade is A, which means that the parameter's component had no technical debt. The lowest, E, flagged the presence of critical problems that must be rectified before the product is worked on further.

An example of the assessment is shown in Figure 2. To see the big picture, we assess the product based on the following eight dimensions.
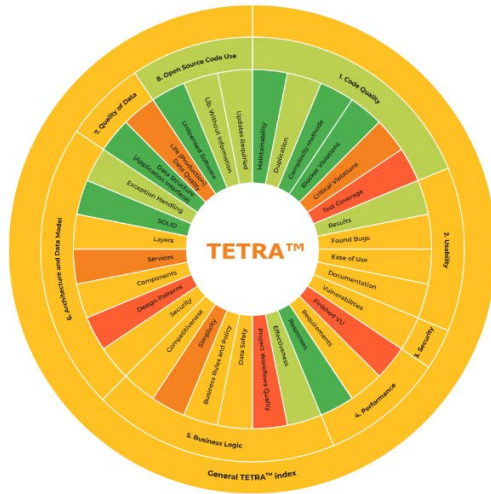
Figure 2: Example of TETRA™ assessment report.
ᵃ Note. Own development.

2. TETRA™ Dimensions

2.1 Dimension 1

Source code quality includes not only the completeness and correctness of the required functionality's implementation but also the ease of support and its modification [5]. It is evaluated by these parameters:

- **Cyclomatic complexity** of the code is a metric on which the complexity of the code maintenance directly depends and is calculated as the number of independent paths in the code.
- **Duplications** is one of the main characteristics that reflects how easy it will be in the future (or present) to make changes to the code. The fewer duplicates there are, the easier it will be to live with this code.
- The **code coverage** level is read as the ratio of the number of code points covered by the unit tests to the number of all existing ones.
- **Rules compliance** is the set of rules for compliance and conventions that verify the basic design principles for source code that's shared by the development team.
- **Maintainability index** is a software metric that measures how it`s easy to maintain the source code.
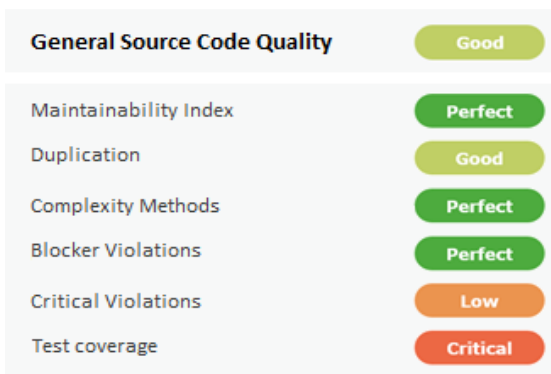
Figure 3: Source code quality assessment summary

Figure 3 shows the Summary Source Code Assessment results where the most pressing issues are related to the unit test coverage and critical rules compliance violations.

2.2 Dimension 2

Usability, UI & Documentation covers the assessment of a software product's usability, UI, and documentation, as the name suggests. This process includes a checklist and expert assessment, which address learnability, memorability, likeability, error tolerance, and efficiency metrics [6]. In order to give an accurate assessment, we use Usabilla, Browser Shots, W3C Markup Validation Service, and other tools. A usability assessment example is shown in Figure 4, where you can see that the usability quality level is good.

Figure 4: Usability assessment summary.

2.3 Dimension 3

Security delves into product vulnerabilities related to BID, CERT, OSVDB, CVE, CWE, and OWASP [7]. We use automated and manual approaches within the security assessment, using tools like OWASP ZAP, Burp Suite, and Nessus [8, 9]. An example of the security assessment is shown in Figure 5; 8 vulnerabilities were found there, giving the software a medium security quality level.
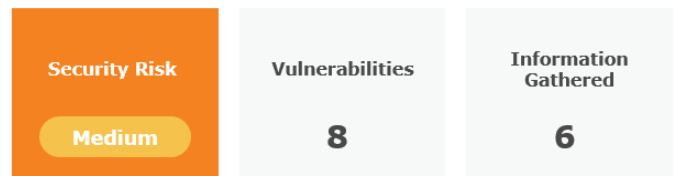
Figure 5: Security assessment summary

Figure 6: Performance assessment summary

2.4 Dimension 4

Performance assessment is the evaluation of whether the product meets the customer's requirements. We perform the assessment through load testing and by simulating the activity of virtual users. We use tools like LoadUI Pro, WebLoad, JMeter, and LoadRunner to assess certain performance parameters, including peak and average response times, concurrent users, and a separate set of server-state assessment parameters. An example

of the performance assessment is shown in Figure 6. The most impact on the final evaluation was made by the fallen virtual users.

2.5 Dimension 5

Business logic assessment determines whether software aligns with necessary business processes. This kind of assessment considers feedback from various focus groups, including customers, end-users, community, management, etc. The business logic assessment relies on six main metrics: data safety, project workflow quality, effectiveness, simplicity, business rules and policy, and competitiveness. An example of the business logic assessment is shown in Figure 7.
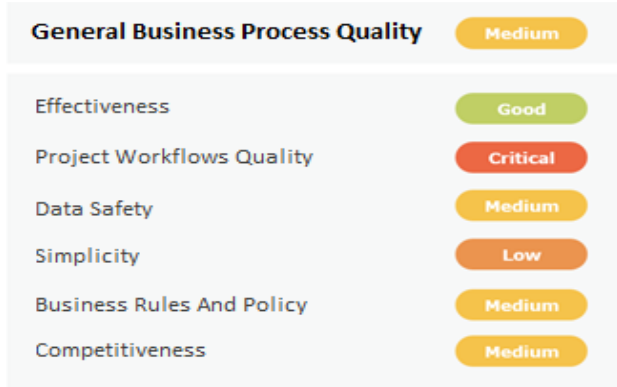
Figure 7: Business logic assessment summary.

2.6 Dimension 6

Architecture quality assessment contains an expert's conclusion regarding the data model and core structure of the software. During the assessment's administration, we analyze services, layers, exception handling, design patterns, infrastructure and recycling components, and more [10, 11].
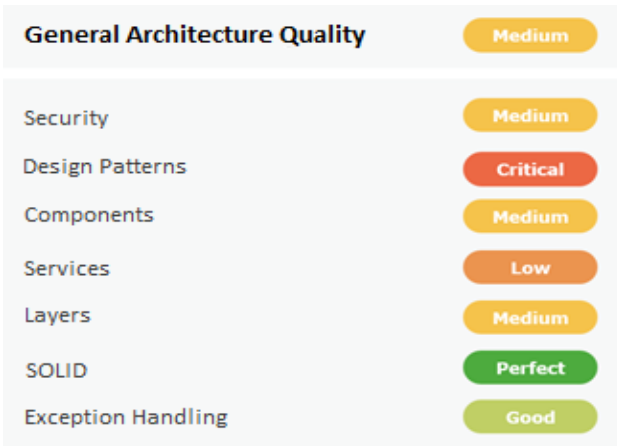
Figure 8: Architecture assessment summary.

2.7 Dimension 7

Data quality encompasses a product's resistance to bad data, how it handles exceptions, and its bad data preventative measures. We base the assessment on metrics like completeness, timeliness, validity, currency, accuracy, consistency, and accessibility. To ensure the assessment's accuracy, we use manual testing and the

Talend Open Studio. An example of the data quality assessment is shown in Figure 9. In this dimension, both metrics were assessed at the medium level.
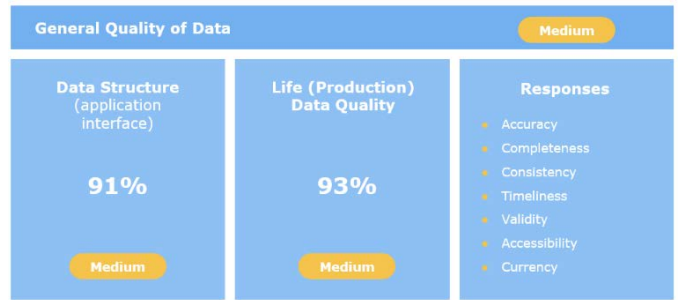
Figure 9: Data quality assessment summary.

2.8 Dimension 8

Open-source code use assessments find any open-source components contained in the software product. We included this dimension to pinpoint any operational, legal, security, or copyright issues that could occur from using open-source components. Inside the report, we give a list of all copyrights, open-source code, and licenses used in the software product. Furthermore, we analyze the software's open-source version. The assessment is done with Black Duck, Palamida, and other tools. See Figure 10 for an example of an assessment of open-source code. The biggest problems were caused by the inclusion of unlicensed software.

Figure 10: Open source code use assessment summary.

By using this method of analyzing and measuring technical debt, we create the final report that describes each weakness and determines the product's quality.

Because these dimensions significantly impact business needs, they must be communicated to stakeholders. Each component can be checked independently, and some might not be applicable to certain products. Separating technical debt dimensions necessitates manual action and review from a personal expert to be factored into the final scoring. We consider this to be less of a problem than not estimating a software product's technical debt.

3. Advantages of TETRA™

Using TETRA™'s framework to expand the concept of technical debt brings benefits to developers and all stakeholders. By understanding the total technical debt, the following can be achieved:

- Managers are able to observe whether project management is effective. They can identify bottlenecks that are currently

hindering the project's development or could become a problem in the future.

- Designers and developers get an unbiased third-party report on project coherence and code quality.
- Investors who are presented with the TETRA™ evaluation can use it as an additional quantifiable justification for product investment.
- Product owners can use TETRA™'s evaluation to assess the effectiveness of fund allocation to various development teams. Thus, they can place priorities within product development areas or decide to conserve until the technical debt has been removed.

4. Limitations of TETRA™

However, TETRA™ is not a one-size-fits-all solution, nor is it a miracle. While it is a highly effective evaluation tool, it does have certain limitations, namely:

- While TETRA™ allows technical debt to be identified, it does not provide answers on how to minimize said debt. There will need to be further investigation into the causes of technical debt, and the management and development team must change their established practices.
- Because technical debt goes past quantifiable technical aspects of code quality (e.g., number of vulnerabilities/ defects per code volume), the process cannot be fully automated and needs subjective, expert assessment.
- In order to minimize technical debt, TETRA™ assessments will be required on a regular basis.

5. TETRA™ Use Case: E-Learning Platform

One of our EdTech clients wanted to innovate and revamp its e-learning platform with the goal of adding to its user base and improving its market value.

The companies target audience consists of students and graduates who are studying English and trying to improve their language skills. Most of the platform's users are tech-savvy kids, teenagers, and young adults who are comfortable with modern technologies.

To ensure the platform was a good fit for their target audience, our client needed to measure its current technical capabilities and business efficiency before introducing new features.

The project contained about 120,000 lines of code and 7 logic sections/modules. 5 developers and 2 QA developed the platform for 2+ years. A wide array of tools and technologies were used, including Spring Boot, Angular, Java, MySQL, Git, Jenkins, and jQuery.

A profound platform assessment was required, so the client chose us for our extensive expertise; they requested that we carry out an unbiased analysis.

To meet the client's needs, we applied the proprietary TETRA™ approach to their platform, analyzing all eight components outlined in Section 2. Each stage of the assessment concluded with a report about the product's bottlenecks and pain points. The process took 2 months from start to finish.

Throughout the TETRA™ analysis, we found several areas that the client was satisfactory in, including their open-source code use and source code quality. However, there were also several areas that needed improvement: the platform had a medium security risk, as well as poor architectural, performance, and business logic elements.

See results in table the Final TETRA™ report below.

Table 1: The final TETRA™ report

Metric	Assessment Result	Evaluation Rank	Score	Metric Index	Dimension Index	General TETRA™ Quality Index
TETRA™ Eight (8) Dimensions of Quality						
1. Source Code Quality					B	C
Unit Test Coverage	0	X < 40	5	E		
Critical Violations	44	40 < X < 60	4	D		
Duplication	4.30%	4% < X < 5%	2	B		
Maintainability index	1.9%	0 < X < 2.1	1	A		
Complexity Methods	4.40%	0 < X < 5%	1	A		
Blocker Violations	0	X = 0	1	A		
2. Usability, UI & Documentation					C	
Documentation	Failed: 9%, Failed weight: 15%	Failed: 5% < X < 10% Weight: X < 20	3	C		
Found Bugs	Trivial: 4, Minor: 20, Major: 2	Major - Worst priority	3	C		

Ease of Use	Score: 3	X = 3	3	C		
UI Results	Failed: 8%, Failed weight: 10%	Failed: 5% < X < 10% Weight: X < 20	2	B		
3. Security					C	
Vulnerabilities	Trivial: 3, Medium: 4	Medium - Worst priority	3	C		
4. Performance					C	
Finished Virtual Users	10%	< 85%	5	E		
Customer Requirements	85%	80% < X < 90%	3	C		
Responses	100%	X = 100%	1	A		
5. Business Logic					C	
Project Workflows Quality	23%	X < 25%	5	E		
Simplicity	30%	25% < X < 50 %	4	D		
Data Safety	67%	50% < X < 75%	3	C		
Bussiness Rules and Policy	57%	50% < X < 75%	3	C		
Competitiveness	63%	50% < X < 75%	3	C		
Effectiveness	80%	75 % < X < 90%	2	B		
6. Architecture Quality					C	
Design Patterns	25%	X < 30%	5	E		
Services	41%	30% < X < 50%	4	D		
Security	65%	50% < X < 70%	3	C		
Components	59%	50% < X < 70%	3	C		
Layers	63%	50% < X < 70%	3	C		
Exception handling	85%	70% < X < 90%	2	B		
SOLID	98%	95% < X	1	A		
7. Data Quality					C	
Real-slice (Production) Data Quality	58 %	50% < X < 60%	4	D		
Requirements Data Quality	96%	95% < X	1	A		
8. Open Source Code Use					B	
Lib. Without information	8%	5% < X < 10%	2	B		
Updates Required	8%	5% < X < 10%	2	B		
Unlicensed software	0%	X = 0%	1	A		

[a] Note. Own development

6. Results

The final TETRA™ report is illustrated in the Table above, which demonstrates that the overall platform state was "Medium," showing the value "C" of the TETRA™ Index. Simultaneously, a detailed results interpretation showed that functionality updates sometimes lead to software crashes.

By carrying out the analysis, we were able to give the client the bigger picture about their platform; they understood which areas needed enhancements and were able to avoid crashes resulting from future updates.

We provided the client with a detailed list of recommendations, which the client followed. As a result of the improvements, the platform's user-base grew by 17.3%, and product redevelopment costs were cut by 22%.

7. Conclusion

TETRA™ isn't just applicable for tech experts – it benefits everybody who works on a product. Testers and developers alike receive an unbiased assessment of their work, users get a well-performing, reliable product, and managers get a better idea of the product's capacity. Furthermore, the processes used within the assessment encourages best practices from the team, thus increasing team proficiency – and thereby the product's quality

TETRA™ gives proper business information to product owners and investors. Companies can get a fair analysis of their product and determine whether it's ready to be launched. If the product already works as intended, the owners can get an estimate of its market value and efficiency. And regarding investors, TETRA™ enables them to better define the product's market value, evaluate the state of the purchasable product, and determine potential investment risks for a transaction.

Throughout the development of TETRA™, we had the overall goal of quantitatively evaluating software product conditions and helping development teams get better results. If you don't use comprehensive, standardized measurements, it isn't possible to understand and manage the product's quality. TETRA™ determines whether the needs of the investors, market, development teams, product owners, and users have been met. It allows you to efficiently take control of your project.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] G. A. Campbell, P. P. Papapetrou, and O. Gaudin, "SonarQube in Action," 1st Edition, Manning Publications, **392**, 2013, ISBN-13: 9781617290954.

[2] N. S. R. Alves, L. F. Ribeiro, V. Caires, T. S. Mendes and R. O. Spínola, "Towards an Ontology of Terms on Technical Debt," 2014 Sixth International Workshop on Managing Technical Debt, 1-7, 2014, doi: 10.1109/MTD.2014.9.

[3] P. Ciancarini, D. Russo, "The strategic technical debt management model: an empirical proposal," In: V. Ivanov, A. Kruglov, S. Masyagin, A. Sillitti, G. Succi (eds) Open Source Systems, OSS 2020, IFIP Advances in Information and Communication Technology, **582**, Springer, Cham, 131-140, 2020, doi: 10.1007/978-3-030-47240-5_13.

[4] J. Yli-Huumo, A. Maglyas, K. Smolander, "How do software development teams manage technical debt? – An empirical study," The Journal of Systems and Software **120**, 195-218, 2016, doi: 10.1016/j.jss.2016.05.018.

[5] A. Peixoto de Queirós, R. and A. Simões, and M. T. Pinto, "Code Generation, Analysis Tools, and Testing for Quality. Hershey," PA: IGI Global, **288**, 2019. doi: 10.4018/978-1-5225-7455-2.

[6] ISO/IEC 25063:2014, "Systems and software engineering — systems and software product quality requirements and evaluation (SQuaRE) — common industry format (CIF) for usability," 1st Edition, ISO/TC 159/SC 4 Ergonomics of human-system interaction, **33**, 2014.

[7] OWASP, "OWASP top 10 – 2017. The Ten Most Critical Web Application Security Risks," OWASP Foundation, **25**, 2017.

[8] CISQ, "List of Weaknesses Included in the CISQ Automated Source Code Quality Measures," CISQ, **31**, 2019.

[9] ISO/IEC 5055:2021, "Information technology — Software measurement — Software quality measurement — Automated source code quality measures," 1st Edition, ISO/IEC JTC 1 Information technology, **235**, 2021.

[10] Microsoft Patterns & Practices Team , "Microsoft Application Architecture Guide," 2nd edition, Microsoft Press, **560**, 2009, ISBN-13: 978-0735627109.

[11] P. C. Clements, R. Kazman, M. Klein, "Evaluating Software Architectures: Methods and Case Studies," 1st edition, Addison-Wesley Professional, **323**, 2002. ISBN: 0-201-70482-X.

A Theoretical Artificial Intelligence Framework for Electricity Generation Life Cycle

Ndala Yves Mulongo[*,1], Ernest Mnkandla[1], Grace Kanakana-Katumba[2]

[1]*School of Computing, Department of Information Systems, University of South Africa, Johannesburg, 2000, South Africa*

[2]*Faculty of Engineering and the Built Environment, Tshwane University of Technology, Pretoria West, 0183, South Africa*

A R T I C L E I N F O

A B S T R A C T

In human life, energy plays an indispensable role. The development of human society in some sense is inseparable from the development of high-quality energy and the use of advanced energy. Alongside economic and population growth, global energy production keeps growing, causing numerous dilemmas. Further, worldwide energy sector is facing enormous pressure to ensure stable and reliable power generation. In order to address these issues, various strategies have been developed over the past decades, however, they failed to meet the expected outcomes. Hence, this study adopted the Prisma methodological approach to develop an integrated theoretical artificial intelligence framework for electricity generation system. The proposed framework has the potential to assist the electricity utility to quickly track and respond on changes across the power grid and within the electrical energy distribution networks in real time. Further, it will help decision makers in energy sector to achieve efficiency in several aspects of their activities, from preventive plant maintenance to fuel-use decisions beyond.

1. Introduction

As a consequence of growing greenhouse gas emissions (GHG) rates in ambient ecosystems, global warming necessitates a decrease in the global usage of fossil fuels, which accounts for roughly 80% of global GHG emissions [1]. Substituting fossil fuel for green energies in power grids is also a concern because of the unreliable existence of green energy sources, such as wind and solar energy. Present electricity grids have a strongly regulated architecture that enables only a one-way transfer of electricity and data between producers and consumers, restricting the integration of decentralized and sporadic production of clean energy sources into networks [2]. Intelligent electricity grids, operated on the concept of bi-directional flows of power and information, offer increased energy demand and supply stability and can therefore support higher rates of renewables in electricity grids [3]. An active involvement of customers, whether by energy supply and/or the integration of energyresources, plays a key role in efficient smart grid's activity, as illustrated in the description of smart grids by [4]: "A smart grid is an electrical network that effectively incorporates the behavior and activities of all its linked users. This is a problem for the management of energy output and availability at all times. Weather awareness is then required to

effectively adjust the energy market to weather instability and global warming.

Climate forecasts have, however, been constrained for many factors in the energy sector: broad range of available climatic databases, which are heterogeneous in terms of population classes and projections of greenhouse gases; inability to adjust contextual and geographical climate model resolutions for effect modeling; model bias; absence of consumer guidance; no customer-friendly database repositories [5]. In order to determine the feasibility of a potential wind farm, the wind power and photovoltaic capability factor, which corresponds to the proposed site for a future wind and solar plant, may be a guide for long-term pattern assessments and performance estimates. Oversupply variance measure calls for successful planning for potential adjustments in cumulative and seasonal data variations. Influx shifts impact energy costs and hydropower plant service optimisation [6]. For instance, environment projections often suggest normal winter temperatures and therefore a lower energy demand for heating. Together, these two improvements would authorize the usage of greater hydropower production and restrict spillage in the event that the reservoir ability is not big enough to hold excess water [7]. For areas classified as vulnerable to the possibility of freezing rain incidents, improved weather plans will also examine the effect of freezing rain on energy resources. An in-depth evaluation

of incident period data, predominant wind patterns and weather research will help decision taking processes for future contingency steps. In the climatic time scale, the predictor for the conditions of bioenergy development lies in the season ideal for forest harvests. The energy demand predictor (estimated with population-weighted heating rates) seeks to help the energy sector forecast supply needs, thus helping to align high demand with weak renewable energy capacity [8]. In certain nations, in sure. France's energy usage is very much related to this measure, and a linear model will say a lot about the energy use variability in this situation [9]. Nonetheless, there are other drawbacks of use the dataset: issues related to the adequacy of the model outputs and energy requirements (e.g. wind speed of 10 m versus 100 m), availability of high-frequencies outputs, adequacy of spatial resolution model and energy requirements, drawbacks of bias correction (restors in many areas, unavailability of high-resolution measurement, project dependency [10]. To this end, this study develops an integrated artificial intelligence framework for electricity generation lifecycle.

To date, artificial intelligence is a critical catalyst for the automation of the energy sector, allowing the processing of vast volumes of data and refining increasingly complex structures. In the electricity sector, technological innovation effectively transforms data into value [11]. The skyrocketing value of AI in the energy sector will also be the product of development in two other technological patterns: decentralization and electricity generation. Decentralization is powered by the expanded installation of micro power stations, primarily solar photovoltaic (PV) rooftops, connecting to the power grid. Electrification of transport and houses (heating and cooling) requires a significant variety of modern loads, including electric vehicles, heat pumps and electric boilers. These emerging tools on the supply and demand side bring uncertainty to the energy system, rendering surveillance, governance and control critical to the sustainability of the energy transition. AI will benefit the renewable energy industry in a variety of areas, involving improved management, production and maintenance of renewable energy assets; more streamlined network operations and actual-time control; the introduction of emerging customer designs; and the creation of innovative business models. AI systems are independent and can work without human intervention and can learn and decide the way decisions can be made and draw specific conclusions based on the examination of various circumstances [12]. Investing in new AI-based technology was one of the key public sector approaches at different level of government in a number of countries worldwide. Nevertheless, while there have been many empirical research on this issue, the findings of this approach also need to be systematized. AI started to be quoted in research in the 1940s as to whether computers were able to determine [13]. Throughout the 1970s, practical AI approaches were developed and experiments were conducted throughout different areas [14]. More recently, the 2010s included many public sector studies and implementations [15]. Given the technical advances in AI, the government still provides services in an old way that can reflect the allocation of public spending, as the majority of the resources are for maintaining the existing structures [16]. This could rising public service trust and citizens' satisfaction, particularly in relation to private sector services [17]. The introduction of modern electronic systems will improve the efficacy and

productivity of government and satisfaction of the public [18]. In addition, modern online technologies will make networking and interactivity simpler, while raising red tape [19]. The adoption of AI-based technologies by energy sector administrators has greatly increased in order to achieve savings in the delivery of energy to end-users [20]. Nonetheless, [21] point to the need to consider technical developments in the public sector and note that work into AI is still lacking. This thesis is inspired by a theoretical void in the literature in relation to public sector work and applications of AI. The work begins with the problem of government functions and thus the current research is intended to address the current situation in the field of AI technology and its implementations in the public sector in a standardized way.

2. Evolution of Artificial Intelligence

The authors described AI as "a computer or program's ability to perform a task that would necessitate some kind of intelligence, if performed by a human being. In [22], the author figured that the design, behavior, functionality, tasks and principles could be used to describe AI. In [23], the author described AI as 'an operation that makes machines smart, and intelligence is the quality which allows a company to operate properly in the environment and with foresight. In addition, a lack of an appropriate and widely agreed definition may help to promote the development of AI. AI technologies could be related to human intelligence, in part through the question of addressing the interpretation, reasoning, understanding, and to a lesser degree emotional intelligence, and novelty of information. The AI definition is based on the premise that the human cycle of thinking can be highly automated. Just before the technological period, AI assertions was seen in various civilizations. In 1956, the word "artificial intelligence" was first introduced at Dartmouth College by American computer scientist and officially recognized as a research discipline [24]. The programs developed by AI have been very remarkable throughout the early years of AI development. Its contributions could have included solving algebraic word problems, geometry theorems and learning to speak English. AI science has attracted a lot of attention and support from government agencies over the same period. In [25], the author argued that it was about then that in the WABOT venture in Japan the world's first complete-scale autonomous machine was produced. Although, AI is a challenging area and scholars did not know its importance in the past. Their expectations boosted the previous AI scores too high and the desired improvement was still not made. AI experienced numerous criticisms and economic failures in the 1970s, which culminated in significant reductions in work and funding [26]. This was also the first winter of the AI.

In [27], the author stated that the first use of the Expert Programs in the early 1980s has brought AI science back from the brink. They was a big success at the onset. For one project, a specialist study called XCON yearly generated $40 million for 1986. This process, however, gradually would have been too costly to sustain. The key drawbacks of Expert systems have included the complexities of system upgrades and their fragility. It was the beginning of the second AI season. Nevertheless, after this time, the AI field did not emerge entirely. Rather, it continued to evolve under labels including artificial intelligence, cognitive systems. The modern era of AI eventually arrived in the last decade, which brought with it ever more rapid computing

capabilities, especially parallel computing power, new sub-fields including Deep Learning (DL), Ma-chine Learning (ML), Artificial General Intelligence (AGI) and Big Data (BD).

2.1. Artificial Intelligence description

AI may be classified either on the basis of features or on the basis of their development over time. On the basis of its functionality, AI could be divided into constricted AI and general AI. Narrow AI can be used in several everyday activities owing to the vast number of applications that continue to evolve daily. Many of these activities involve arranging schedules, both private and corporate, reacting to consumer requests, conducting visual reviews of various services and infrastructure, tagging specific types of objectionable material on-line, etc. At the other side, Specific AI (or Artificial Artificial Intelligence) is somewhat unique and quite similar to the kind of emotional intelligence present in humankind. Such a system could execute operations that need more accuracy upon understanding and ac-accumulating practice. Up to now, Generalized AI can be used only in videos, and AI specialists are working to make this vision come true quickly.

Another method of categorizing AI system is based on the AI transition. As stated by [28] the early advancement of AI after its inception is identified as the old-fashioned of AI or, in other terms, Good-Old-Fashioned Artificial Intelligence (GOFAI). GOFAI was a theory that was established at the Dartmouth conference in 1956. Mainly focused on hand-coded abstract machinations, GOFAI was quite close to what conventional programming looks like. It is often referred to as symbol for its effort to explain knowledge in conceptual terms [29]. An-other of its famous euphemisms is Conceptual AI. Symbolic AI was designed to generate human-like knowledge in a computer, whilst most current studies in the area of AI emphasizes upon particular sub-problems.

The most effective type of Conceptual AI is an expert model. Of artificial intelligence, an expert program is a software program which equals or exceeds the decision-making capability of a professional expert [30]. Expert Systems were developed in 1965 and were some of the first fully popular examples of AI software. The purpose of such types of systems was to render the knowledge available for the program to function directly rather than indirectly. In other terms, the guidelines controlling the functioning of the expert systems needed to be specified in a way that was understandable and simple to manage. Therefore, even industry experts and not even IT (Information Technology) experts may be edited. The key benefits of this clear interpretation were quick growth and simple maintenance. The development of expertise is one of the main drawbacks of the Expert Programs. Certain issues involve implementation challenges, accessibility to vast repositories and quality issues [31]. Many scholars have offered various explanations why GOFAI could not bear the time test. The key point of these critiques was that the depictions used in the early GOFAI were not valid depictions of the modern world [32]. Another claim suggested that the reliance on logical analysis and inductive reasoning neglected other forms of reasoning.

Traditional subsets in the AI area include ML (Ma-Chinese Learning) and DL (Deep Learning). While the expressions AI, ML and DL are used synonymously in the literature, there is a strong difference between these three concepts. DL is a subsystem of ML which, in effect, is a component (or unit) of AI. One element that separates ML from typical information networks and expert structures is its capacity to develop and evolve on its own. ML can also render certain algorithms of its own. Arthur Samuel, one of the ML's founders, described it as "a research area that allows computers the opportunity to know without being directly programmed" [33]. DL is another branch of ML, which usually applies to large, artificial neural networks, and often large, reinforced learning. In DL, machines master the strategies that naturally come to humans. This is the main technology behind speech recognition, picture recognition, driverless vehicles, and so on.

2.2. How Artificial Intelligence function

The first theoretical solution to AI is the rule-based expert systems. This method fits the basic laws of the AI programs: if X, and then Y. A set of if-then rules expresses experience and information that fits well with clear, well-defined problems. However, this method falls apart when the number of possibilities increases dramatically in policy-making and where the laws disagree with each other [34]. Machine learning has been developed to address the shortcomings of the rule-based expert systems. As a subgroup of AI, machine learning algorithms also use statistical methods to adapt algorithms to conditions and "read" from data without being expressly configured to do it anyway [35]. Based on the probabilistic paradigm, machine learning algorithms reflect ambiguity — a central aspect of decision-making — via a probabilistic viewpoint. The state-of-the-art developments in probabilistic machine computing involve predictive programming, Bayesian optimisation, data encoding, and automated system development [36]. Considering, as a practical illustration, a manager choosing a plan from several options to be introduced in an enterprise. Any of the various interventions suggested for consideration by the committee leaders has an uncertain likelihood of implementation in the policy-maker's (in this case the leader's) company. When make a judgment, the leadership assesses the scenario by evaluating a variety of variables (e.g. the assets required to execute each plan, the performance rate of each system in other organisations, and the similarities between the leadership company and other organisations). The results of all these variables was then used to measure the likelihood of performance of each option plan in order to notify the decision-making process of the manager. Such a predictive method may be really useful for educational officials who do not realize how all the complexities of data processing and review affect the understanding of results. After all, even computer scientists often make incorrect judgments employing DIDM [37]. In this probabilistic method, AI algorithms produce valuable knowledge and make suggestions for DIDM members.

2.3. Artificial Intelligence strategies

AI is a computer science subset. Different academics describe AI in various ways. The variations in the AI concept include two dimensions: firstly, human primacy and secondly, rationality. The key element of intelligence is the logical behaviour. Through this view knowledge approaches the solution to problems by the laws of thought, that is to say by clearly defined reasoning processes (Aristotelian reasoning). The logical method leads to structures

that incorporate mathematics and technology. AI also includes rationally operating systems. One concept of AI is: artificial intelligence is to investigate and perceive and act agents that exist in the world [5]. To this end, different AI methods have been designed and implemented in several various studies. These methods are presented and briefly listed in chronological order below.

2.4. The Analytical Hierarchy Process (AHP)

AHP developed by [32] is a common decision-making system. The benefits of the scheme include its positive mathematical properties, the simplicity of collecting the necessary input data and the fact that it is a decision-making method that could also handle complex questions. The organizational composition of the AHP comprises of requirements, sub-criteria, goals and options. Within AHP, dynamic issues should be first divided into elements that are organized into several organizational layers. For the next phase, policy-makers evaluate each clast in pairs depending on their expertise and information. Any degree of ambiguity can emerge when similarities are made by personal decisions. The greatest advantage of the AHP system is its ultimate process, named consistency verification. When the similarities have been rendered at all stages of the hierarchy and confirmed by the verification of accuracy, an aggregate priority classification is established that is focused on the value of each element and the related criteria rating.

2.5. Fuzzy Logic (FL)

Because FL is one of the simplest to grasp among the different AI strategies, the advantages of fuzzy logic (FL) are multiple. It is built on human language, and the benefits include its ability to draw on the expertise of specialists, its capacity to model multi-linear processes, its understanding to imprecise information, and, eventually, its ability to integrate with traditional control methods. The three key phases of the Fuzzy method are: fuzzification, intrusion and defuzzification [6].

2.6. Genetic algorithms (GA)

GA are strategies focused on the principle of fittest growth. Necessary approaches to the dilemma are referred to as "humans," and their progression in age is examined. The genetic algorithm uses triple primary providers: range, crossover and mutation [7].The fitness feature tests the capacity of each person to resolve the question, and this is repeated per phase in the reproductive cycle. The degree to which each person is duplicated is attributed to his health. In other terms, the further desirable an entity or a remedy, the better are the odds of his descendants being passed toward the next era. A GA conducts a particular replication of individuals to locate the possible solution to a question. Also the population with better odds of success continues to the next generation under the evolutionary theory of the fittest. Genetic operators assume the function of modern and enhanced descendants [8]. The method proceeds up to a given amount of years until an appropriate answer to the question has been discovered.

2.7. Artificial Neural Network (ANN)

ANN is a computational network or algorithm that is made up of artificial cells and could be used to design the biological endocrine system. The author introduced the first train-friendly

NN, in 1957. At a NN, weight factors are generally used to model the relation between an input and an output vector. The model error is estimated during preparation and weights are modified with stochastic gradient descent or other steps to minimize the model error [9]. A multi-layered NN perceptron is an extension of an RNN [10]. It models time dependencies in target data through input links. By contrast with other methods, NN approaches are ideal for the simulation of multivariable problems [11]. Variables exhibit complex relationships among themselves in a multivariable problem. NN approaches are able to derive nonlinear associations by observing from the training data from variables. The strongest use of the NN was in the estimation of electrical load in houses.

2.8. Simulated Annealing (SA)

SA is another AI search or optimization strategy. This is a heuristic quest tool that chemists and metallurgists originally developed to determine the most stable condition in a chemical system [12]. By comparison to other iterative forms of development, SA requires fewer desirable options in the quest cycle, thereby stopping the task from getting wrapped up in local minima. In [13], the author used GA and SA to optimize multi-module thermoelectric coolers and found that SA needed less time and energy in computation. Many of the above AI strategies are powerful and low. Similar to AHP or FL, GA is better at addressing complex issues. Fuzzy methodology is therefore effective in non-linear environments because it is easier to work with ambiguity and subjectivity [14]. The AHP could be really effective for decision support networks, being simplistic and comprehensive. SA should prevent local minima trapping of the quest. Recently, NNs are widely seen in ML and DL.

3. Research Methodology

The layout of the systematic literature review used in this study was built in compliance with the PRISMA protocols. The PRISMA approach involves excellently-defined steps of a comprehensive review, such as the establishment of eligibility requirements and the description of material sources, search methods, selection mechanisms, findings and data synthesis [15]. The current systemic methodology under the PRISMA paradigm includes the critical assessment of recent research in the area of artificial intelligence in the energy sector. This method has been used as it promotes detailed and factual information [16]. In [17], the author defines the process of systematic analysis as suitable for the identification, evaluation and interpretation of the current body of scientific evidence. The systematic review is important to add value and examine the accuracy and discrepancies found in existing literature [18], which lead to the growth of theory, experience and knowledge [19]. The development and analysis of the evaluation method has been updated from [20] and [21] utilizing the [23] methods. Further, the choice of this approach is explained by the consistency of the protocols that make things simple to repeat and validate the findings [24]. The PRISMA procedures are divided into four phases: (a) identification, (b) screening, (c) eligibility, and (d) inclusion as shown in the Figure 1 below [25]. At the first stage that was based on the identification of relevant literature. To date, Literature analysis of the study's focus concepts, such as "Artificial Intelligence in the Energy Sector." This has been undertaken to insure that an effective

search approach has been established as the search technique is compatible with the technical retrieval of specific documentation [26]. To order to create an initial understanding of the research field and to establish the nature of the analysis [27], various variations of search words were used in websites and databases. This method permitted the selection of relevant keywords for the analysis report, which therefore focuses on specific main area of interest, Artificial intelligence in energy sector. The search was conducted via SCOPUS database with an exact phrase in Title, Abstract, and Keywords as "Artificial Intelligence in Energy Sector", "Artificial AND Intelligence AND in AND energy AND sector". It came out with 336 articles. Further, came to the second stage. This stage consists of the search for literature and the compilation of the related articles for the research, based on the results of the search procedure for keywords. This method was carried out in four stages as shown in the Figure 1 below.

As shown in the Figure 1 above, subsequently, the findings of each phase were optimized on the basis of inclusion and exclusion criteria applicable to this analysis. As a consequence, in Phase 1, Scopus, was chosen for a literature review to ensure a thorough quest. The quest for the keyword was initially performed in the text, abstract in Scopus. The screening parameters used to remove possible papers related to this research from the corpus were; the compilation of papers from peer-reviewed publications since this approach contributes to higher content and scientific performance [28], the papers published in English, and no constraint on the year span because this analysis aims to investigate the

phenomenon. The findings of the tentative quest in Step 1 produced a sample of 336 documents. Additionally, to order to ensure the quality of this study, the authors only used journal papers that were published in English only and excluded conference papers, book chapters, editorial reports, etc. The authors considered the timeframe from 2010 to 2020. The refinements of these two requirements resulted in 131 papers. Throughout abstract review, the collection of papers was focused on the role of Artificial Intelligence in the energy sector, which resulted in the completion of 15 articles as shown in Table 1 below:
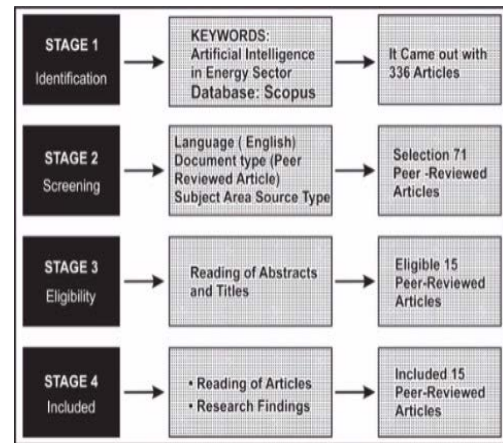
Figure 1: Prisma method's flow diagram (author's own construction)

Table 1: summary of the literature review (author's literature review)

Author & Year	Focus area	AI technics used	Outcomes
[8]	In this article, the Artificial Neural Network and the Genetic Method offer an intelligent scheduling algorithm. The pro-scheduling approach is targeted at reducing grid power consumption by 10%, 25% and 40%, according to weekly appliance schedules with a 15 minute increase.	ANN & GA	The technique offered decreases demand for energy in "first" times; maximizes the utilization of renewable sources (PV and wind turbines) and lowers grid dependency.
[14]	This article seeks, with a special attention on the industrial buildings sector, to foster the gap between high-level methods and the actual use of modelling algorithms. This is done by developing a revolutionary M&V 2.0 approach, which supports machine learning and allows for precise and reliable savings measurement.	ANN	The systematic approach lowers the requirement for extra metering facilities. This can minimize the necessary resources for a precise M&V in a specific project considerably.
[29]	The major aim of this effort was to decrease the number of such periodic evaluations in order to lower the maintenance costs of electrical lines. In particular, the aim is to decrease the number of regular assessments of transmission towers, to prevent step-by-step potentials that are particularly harmful to humans.	ANN & CBR	A sample of transmission towers from a specified set can be provided with the designed system for evaluation. The approach assures that the entire set has identical values without all transmission towers being examined.
[37]	The major aim of this study is to enhance sustainable energy planning procedures and provide a platform for additional GIS analysis rather than simulation of energy. The approach provided is to be integrated in a Space Decision Support System for energy decision-makers; by combining the currently existing data, it will assist to identify priority locations for regeneration initiatives	GIS	The main outcomes are the estimate and development of a technique which may be used at various scales of the geo referenced heating demand of residential structures for the research region (nearly 42,000 structures). Using the technique for the case study, the overall thermal demand of the building stock is overestimated somewhat,
[27]	This work offers an extensive and systemic analysis of the short-term forecasting approaches of artificial intelligence. This research aims mainly at reviewing, identifying,	ANN	The findings in this research reveal that the training capacity of the neural network is clearly shifted toward

	evaluating and analyzing the performance of the load forecast models and research gaps based on Artificial Intelligence (AI).		promising load prediction model outputs than conventional methods.
[22]	The main objective of this study is to examine, acknowledge and analyze the performance of both electrical energy consumption forecasting methodologies. The hybrid of two forecasting techniques can be used for more accurate outcomes compared to the use of a single forecasting approach.	ANN & SI	This work has showed substantial promise for enhancing the accuracy of the load predictions with hybrid interactions between AI and SI approaches. The SI method's capacity to balance exploration and use helps prevent the AI method's problems. The SVM and SI hybrids have also demonstrated that they are more able than the ANN and SI hybrids.
[13]	This report deals with the difficult control of wind energy-based systems by presenting a tool that will allow wind energy controls to be optimized in the near future, particularly the next 10 minutes.	ANN	For predicting wind density values, the precision achieved by the final ANN is enough to allow them to be entered in microgrids to regulate instantaneously.
[26]	Three different ANN models have been created to predict the LD of Tilos Island in Greece in the present time. Weather- and LD data from the period April 2015 to April 2015 have been used for this purpose.	ANN	The results indicated a remarkable predictive capacity for the created models, capturing three distinct perspectives. In this regard, for the medium-term prediction model, a moderate error of about 7,9 percent was identified (hourly predictions of the next 24 hours); for the 10-minute intervals for the next 24 hours a mistake was computed of around 7,2 percent (shorter predictions); and for the next 10 minutes a 1-minute error (short-term forecast) of around 4,0 percent was finally reported (very short- term prognosis).
[33]	This work presents a novel hybrid STLF model that combines enhanced artificial bee colony (ABC) optimization with ANN.	ANN	The results of this study have demonstrated that the LF model based on ABC is superior than the PSO and the GA method. More precision is attainable with MAPE of 1.89%. The accuracy test shows the importance of improving the forecast accuracy from the model.
[19]	In this research, an evaluation is presented on the power charge forecast for energy providers, the examination of wavelet approaches, time series and neural network approaches, and on the long-term predictions.	ANN	. Compared to those expected by power suppliers based on specialized information, the findings achieved were far more effective and so contributed to decision-making on the procurement/sale of energy on future demand.
[1]	The purpose of this study was to use two approaches for the prediction of environmental life cycle impacts and output energy of sugarcane production in plant or cattle farms. These techniques are the Artificial Neural Network (NNNs) and the Adaptive Neuro-Fuzzy Inference System (ANFIS), the models.	ANN & ANFIS	The results suggest that the most important influence on indicators in plant farms is on sugarcane yield, power, machinery, biocidal products and sugarcane stem cuttings. Results from ANN models showed that the determination coefficient (R2) in farms in plantation varies from 0,923 to 0,986, and in ratoon farms in environmental impact and outdoor energy training stages 0,942 to 0,982. Results of the model ANFIS, constructed using a hybrid learning algorithm, indicated that environmental impact prediction is necessary
[18]	The study suggests integrated models of supply of energy based on artificial intelligence, which can minimize machine energy saving effects, in response to anti-logy or common sense. A decision-making model analyzes to what degree the	ANN	As a consequence, three separate models based on artificial intelligence were appropriately regulated through interactions: energy supply model,

	source of the abnormal occurrences is related with the occupant's behavior through the use of design scenarios which assume certain unique occurrences.		thermal comfort model and decision-making model.

4. Research framework

From the table 1 above, it can be depicted that the energy industry is experiencing a significant transition with the increased installation of clean energy technology (solar PV and wind) offering intermittent energy supply, distributed energy systems, bidirectional electricity streaming, massive data flows captured by Internet of things as well as other tools, expanded usage of power storage, and the emerging function of electricity companies and customers. However, these measures are yet to address the current power crisis owing to a lack of an integrated holistic digital framework for the power system. Since, most system management actions are often made and executed manually or with a limited degree of robotics owing to a restricted amount of remotely responsive assets. Though, the aforementioned technological advancements will require a huge range of automatically manoeuvrable infrastructure to react to the demands of a variety of stakeholders (e.g. customers, utilities, transmission and distribution companies, suppliers). This innovative degree of control allows the network to be integrated for more distributed capital, thus optimizing device stability and lowering the cost of running a network for large shares. As a consequence, the function of AI is changing from a supporting and maximizing resource to a need for quick and rapid decision-making. As mentioned above, AI as an emerging technology will benefit the sustainable energy sector in a number of ways. Much of the innovations presently assisted by AI have been in the modeling and predictive maintenance of advanced environment and clean energy production. Throughout the future, AI would also improve decision-making and preparation, quality tracking, compliance, validation and supply chain management which would ultimately boost the performance of energy systems. Nonetheless, this overview reflects on promoting the deeper incorporation of AI into power systems, where Five key types of AI applications can be defined, as seen in Figure 2.
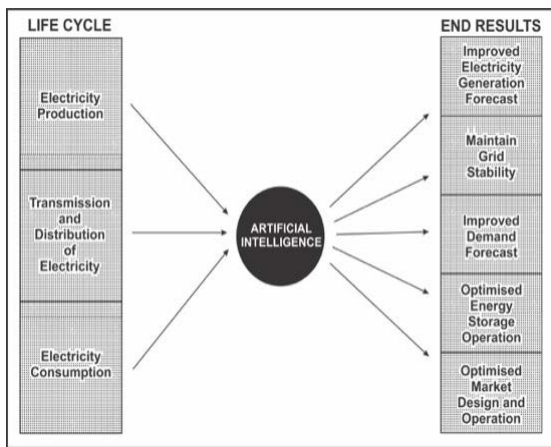

Figure 2: An integrated theoretical artificial intelligence framework for electricity generation lifecycle (own construction)

4.1. Improved electricity generation forecast

Enhanced meteorology is one of the key AI technologies that would boost the introduction of renewables into the electricity grid. Renewable energy sources generate an immense amount of information, and sustainable technologies have gained from long-established sensor technology. to this end, AI could generate reliable energy production predictions which would make it possible to incorporate even more solar and wind power into the grid. Since 2015, for example, IBM has demonstrated a 30 per cent increase since solar forecasts when collaborating with SunShot Program of the U.S. Department of Energy. The self-learning meteorological model and green energy forecasting technologies have combined broad databases of historical information and actual-time analysis from meteorological stations. Effective RES forecasts will help power stations and market factors to accurately predict their performance and bid on markets for wholesale and balancing while minimizing fines. For system technicians, reliable short term projections will improve the unit's contribution, improve transportation capacity and reduce maintenance problems, thus growing the machine's operational resources. From AI, both projects estimated power consumption utilizing data from solar sensors, wind turbine sensors and weather predictions, which helped reduce excess power consumption.

4.2. Maintain grid stability

By delivering reliable demand and supply forecasts, AI may further improve the function of the grid, especially in the case of decentralized systems with bi-directional energy flow, which enhances the intricacy of the energy systems. Energy supply electricity companies face big challenges as the amount of renewable electricity production technologies, such as solar PV, has grown exponentially. The introduction of clean energy technologies contributes to volatility and intermittent peak loads in the electricity system. AI will guarantee that the electrical grid runs at an acceptable demand and will maximize the electricity usage of its consumers. Theoretically, the energy produced by the solar PV system will be used in the household or in the community grid. For instance, in Riedholz, Switzerland, four companies along with the Canton of Solothurn are exploring whether AI systems will guarantee potential grid reliability and reduce expenditure in expensive grid extension in the SoloGrid pilot project. The project explores whether GridSense, an algorithm which studies consumer behavior via AI, will 1) monitor primary energy users, like heat pumps, heating systems, home batteries for electric vehicle charging points, and 2) incorporate calibration data from solar PV systems for optimum grid performance. The algorithm constantly calculates variables such as grid demand, usage and output, namely weather forecasts and energy prices, and maximizes power production and consumption. The system lowers peak demands in the electricity grid, regulates demands and stabilizes the control infrastructure.

Grid instability at the electricity distribution stage is a significant consideration which decreases the introduction of wind and solar PV energy into energy systems. AI will increase the power grid efficiency and raising the need for new lines by allowing greater use of existing infrastructure as a feature of environmental conditions. This is the case, for instance, in the

case of complex line ranking programs initiated by Ampacimon or under review at the Karlsruhe Institute of Technology in the "PrognoNetz" project. AI-based applications, utilizing vast volumes of weather information, will make optimum usage of current electricity grids by adjusting service to environmental conditions at any time and thereby reducing overload. AI may also boost the protection, stability and performance of the power network by automatically identifying disruptions. Operating system may allow autonomous data processing in real-time and identify scenarios of crisis or malfunction of the device. For instance, academics presented AI simulations with examples of typical system failures to enable the algorithm to slowly learn to discern – and precisely categorize – normal operating data from specified system malfunctions. The algorithm was able to make split-second assumptions as to whether there was anomaly fault, as well as the type and position of that disturbance. When one power plant collapsed, an immediate increase in the load on the other power plants might be anticipated. The added load speeds down the turbines and reduces the power. This calls for quick counter-measures (less than 500 milliseconds [ms]) since, if the frequency slips below the threshold value, the operator may be required to break off parts of the grid for the sake of network reliability. Since the algorithm would make a conclusion within 20 – 50 ms, there will be ample time to enforce the correct completely automated counter-measures. According to the study, the algorithm is able to be applied and work continues on the monitoring and enforcement of the related counter-measures.

4.3. Improved demand forecast

Accurate demand projections and RES predictions may be used to automate economic load flows and to boost demand management and performance. Consumers are generating an growing influx of data from the power grid itself. Considerable attempts have been made to deploy digital meters, which are capable of transmitting details to service providers as much as daily. Of this information, AI could forecast not only network demands, but also usage trends, and therefore can reliably create a trend of usage for each user. It becomes much more true with the recent installation of DERs, such as autonomous cars, heating systems and solar PV cells, something that fully alter the typical load form. Bee-Bryte, for instance, is a French start-up which employs AI to forecast a property's electrical energy consumption in order to deliver heat and cooling at the appropriate time, while preserving comfort and ventilation inside the consumer's operational range. It will lead in reductions of up to 40% on energy bills attributable to a mix of productivity improvements and transfers in loads to times where power is inexpensive where green power is active in the grid. Knowing user preferences, beliefs, desires and even personalities further improves the harmony and efficiency of the smart grid. This also encourages strategies to be implemented more efficiently and facilitates an awareness of the individual motives involved with the introduction of green energies and how to adjust customer behaviour to improve the electricity market as a whole.

4.4. Optimised energy storage operation

Power storage technologies are emerging as crucial enablers to clean power deployment in the context of large-scale batteries, aggregates ("behind the meter") or plugged-in electric cars.. AI

will help run those systems in a more productive way, optimizing the introduction of renewable energy (along with the elimination of production projection inaccuracies), reducing locally generated power costs and increasing returns for storage facility operators. In the case of large-scale energy generation projects, which involves decisions on retaining surplus solar power in the battery network and on discharging batteries to satisfy demand at a later date, taking into consideration future output, renewable energy production, costs and infrastructure latency, among other factors. Since storage batteries could be triggered easily and could be used to control excess loads and reduce the back-up resources required from diesel generators, fossil fuel-fired power plants or other maximum power stations, AI could be employed to forecast and make decisions on energy storage management. The nature and sophistication of power storage facilities operation, involving several variables in a fluid environment, includes integrated AI. AI study investigates decision-making on a size and with a sophistication that outperforms those of a human operator, in particular for systems of hundreds of integrated power storage systems (electric power, kinetic, etc.) built on the customer side, in homes or in manufacturing installations.

In fact, AI could assist to measure and expand the lifespan of a storage device by employing computational logic algorithms to data recharging and unloading. Owners install their warehousing packs in conjunction with the cost of battery services and the effect such services have on battery safety. California-based firm Stem produces Athena, which employs AI to schedule energy consumption and allows consumers to monitor electricity variability in storage capacity. For example, Tesla's Hornsdale battery was a wake-up call in Australia in accordance with the US software company AMS. Through leveraging AI, flexible energy storage devices will maximize incentives to buy energy from the grid while rates are small and then sold back to the consumer while prices are inflated. The Hornsdal battery function through a Tesla-developed auto bidder, which enabled the project to catch the best income streams to an extent which human potential buyers alone could not achieve. "Comparable to an individual dealer, algorithmic bidding software will boost the battery turnover by around five times," as per AMS. During its first year of activity, the Hornsdale battery created an additional USD 24 million in sales, thus the ancillary service frequency control expense between USD 40 and USD 50 million, and thereby saving customers in the long run. Price reductions such as this will contribute to an explosion in algorithm production which is most profitable in order to run batteries.

4.5. Optimised market design and operation

Advanced AI-based technologies are now being applied to automate near-real-time marketplace activities. This modeling builds on the study of wide flows of different data to allow fast reaction to changes in the market. Intraday trading is extremely useful to adapt to sudden shifts in the energy supply and usage by utilizing market processes before managing supplies become important. This helps an operator of the electricity generating station who unexpectedly loses capacity in one block to purchase extra electricity from other market players to retain the balance block. Intraday trade is also a critical element for the direct selling of green energy as quickly evolving environment results in an unexpected shortage or deficit of electricity from RES. Within a

competitive world that contains several factors, running intraday markets may be quick and complicated beyond the control of a human operator; this is the right scenario for sophisticated AI. As separate economies are combined to establish national economies, the difficulty of business operations grows even further. An AI-based algorithm named EUPHEMIA was established to measure day-to-day price of electricity throughout Europe and to assign day-to-day cross-border transmission power. EUPHEMIA is often used for coupled-day power rates for 25 European countries (Austria, Belgium, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden and the United Kingdom) with an average regular amount of the same. AI will improve time specificity on energy markets in terms of market architecture and enable for real time markets. The usage of AI in the fields of exchange and dispatch decisions for generating properties is being studied in near proximity to real-time commercial markets, with a view to ensuring that generators participate in business to optimize the benefit of versatile power choices. For instance, Origami Energy, a start-up firm based in Cambridge, UK, uses AI to forecast the supply of products and match market rates in almost real time in order to sell effectively in the transient response markets. By utilizing predictive analytics and deep learning, multiple organizational management challenges can be solved and new ideas can be extracted into medium and long-term planning – such as predicting when an asset would become available, the importance of versatility and the best means of deriving the greatest benefit from an asset.

5. Discussion and Conclusion

With the advent of intelligent appliances and associated software, and also the declining price of green energy, electricity users need an electricity supply system to be more smarter, safer and more efficient than ever. This need is entirely addressed by AI technology, which offers a decentralized sharing system for achieving renewable energy use and the establishment of a sustainable economy. This study attempted to respond to the use of AI technology in the energy sector. Today, the electricity industry is owned by South Africa government, and does not appear to favour the implementation of AI technologies. However, the cost of developing alternative energy sources is rapidly declining, creating major opportunities for the implementation of AI technology. since the electricity sector in South Africa still faces many technical and political issues. The AI technology itself is not yet established — not only in South Africa, but worldwide. The AI energy system is currently under investigation. Nonetheless, a regulation that is not adaptable to the business paradigm of the AI energy sector could be a greater challenge than technology. The restructuring of the electricity sector needs to be more deepened in order to establish the requirements for the development of the digitized energy network. In fact, the following protocol recommendations will be accepted by decision makers. First, the government of South Africa will enable businesses, universities and academic institutions to set up specific research platforms to study AI technology. Main fields of research cover cryptographic technologies, adaptive algorithms and risk management. Work will concentrate on seeking answers to concrete issues in South Africa. While there are some

implementations of AI technologies in South Africa, the number of self-developed AI systems in South Africa is still fairly limited. Third, the government of South Africa will encourage the development of centralized household power production. South Africa's power distribution household lags behind developing countries. Household distributed power production is conducive to the widespread usage of AI technologies in the energy sector. Smart meters are used to record the flow of power between household appliances, photovoltaic cells, etc. Redundant produced electricity can be marketed to neighboring power users and can offer shared resource assistance to household users. In turn, distributed power generation will allow consumers to engage effectively in the development of renewable energy and the elimination of carbon emissions. Thirdly, the related legislation regulating the energy industry ought to be modified. Currently, most green energy ventures in South Africa need to be accepted by government, or businesses need to buy licenses, permits, etc. to carry out a project. As an significant driver for the energy market, so many connectivity and licensing laws do not hinder the smart energy network initiative. South Africa will change the expired Energy Act in order to render people and private firms liable for electricity supply. Where required, additional laws will be implemented to guarantee access to the power grid and the integrity of transactions. In fact, regulators could promote smart monitoring by adding AI technology itself to the management of the energy sector. Eventually, in order to safeguard competition in the energy sector, the government should simplify management measures as well as enhance quality and performance.

References

[1] K Ali., S. Mohammad., M. Hossein., N. Ashkan., C. Kwok-wing ''Combined life cycle assessment and artificial intelligence for prediction of output energy and environmental impacts of sugarcane production''. Science of the Total Environment **664**; 1005–1019; 2019. Doi: 10.1016/j.scitotenv.2019.02.004

[2] N. Apergis., J. Payne ''Renewable and non-renewable energy consumption-growth nexus: Evidence from a panel error correction model''. Energy Economics **34**:733–8, 2021. Doi: 10.1016/j.2021.34.734

[3] N. Apergis, D. Danuletiu ''Renewable energy and economic growth: evidence from the sign of panel long-run causality''. International Journal of Energy Economics and Policy, **4**(4), 578, 2014

[4] N. Apergis, J. Payne ''On the causal dynamics between renewable and non-renewable energy consumption and economic growth in developed and developing countries'', Energy System, **2**, 299-312, 2011

[5] N. Apergis, J. Payne ''The electricity consumption-growth nexus: Renewable versus non-renewable electricity in Central America'', Energy Sources, Part B: Economics, Planning, and Policy, **7** (4), 423-431, 2012

[6] M. ASIF, T. Muneer ''Energy supply, its demand and security issues for developed and emerging economies''. Renewable and Sustainable Energy Reviews, **11** (7), 1388- 1413, 2017. Doi: 10.1016/j.rser.2005.12.004

[7] K. Ayoub, K. Payne ''Strategy in the age of artificial intelligence''. Journal of Strategic Studies, **39**(5–6), 793–819, 2016. Doi: 10.1080/01402390.2015.1088838

[8] Y. Baris., R. Yacine , M. Monjur ''ANN–GA smart appliance scheduling for optimised energy management in the domestic sector''. Energy and Buildings **111** (2016) 311–325, 2016.

[9] M. Beccali, G. Ciulla,, V. Lo Brano, A. Galatioto, M. Bonomolo ''Artificial neural network decision support tool for assessment of the energy performance and the refurbishment actions for the non-residential building stock in southern Italy''. Energy, **137**, 1201–1218, 2017.

[10] G. Bosque, I. del Campo, J. Echanobe, ''Fuzzy systems, neural networks and neuro-fuzzy systems: A vision on their hardware implementation and platforms over two decades''. Engineering Applications of Artificial Intelligence, **32**(1), 283–331, 2014. Doi: 10.1016/j.rser.2014.32.331

[11] S. Chen, A. Jakeman, J. Norton, ''Artificial intelligence techniques: An introduction to their use for modelling environmental systems''. Mathematics and Computers in Simulation, **78**(2–3), 379–400, 2008.

[12] T. Egger, M. Zellweger-Zähner, C. Schneider, C. Junker, G Lengeler, ''Antes, Language bias in randomised controlled trials'' published in English and German, Lancet. **350**, 326–329, 1997

[13] R. Fermín, M. Ane., T. Florez , F. Luis., G. Ainhoa ''Very short-term wind power density forecasting through artificial neural networks for microgrid control''. Renewable Energy **145**, 1517-1527, 2020. Doi: 10.1007/s10163-016-0575

[14] V. Gallagher, L. Kevin, O. Peter, B. Ken, T. Dominic. O'Sullivan ''Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0''. Energy & Buildings, 8–22, 2018

[15] G. González, P. Urrútia, Alonso-Coello,'' Systematic reviews and meta-analysis: scientific rationale and interpretation, Rev. Española Cardiol. (English Ed.) **64** (2011) 688–696, 2011.

[16] M. Ha-Duong, J. Grubb, J-C Hourcade ''Influence of socioeconomic inertia and uncertainty on optimal CO_2-emission abatement''. Nature;**390**(6657):270–3, 1997.

[17] L. Haibo, Z. Jialing, Z. Kai, J. Muhammad ''An improved genetic algorithm optimization fuzzy controller applied to the wellhead back pressure control system''. Mechanical Systems and Signal Processing **142**, 106-708, 2020.

[18] A. Jonghoon; C. Soolyeon ''Anti-logic or common sense that can hinder machine's energy performance: Energy and comfort control models based on artificial intelligence responding to abnormal indoor environments''. Applied Energy **204**, 117–130, 2017. Doi: 10.1016/j.apenergy.2017.06.079

[19] R. Liviane., L. de Santana., G. Conde, S. da Silva, L. Francês, A. Rocha ''Comparative Analyses of Computational Intelligence Models for Load Forecasting: A Case Study in the Brazilian Amazon Power Suppliers'', 2009.

[20] M. Naderi, E. Khamehchi ''Fuzzy logic coupled with exhaustive search algorithm for forecasting of petroleum economic parameters'', Energy, **176**, 291–298, 2019.

[21] J. Mathur ''Development of a modified dynamic energy and greenhouse gas reduction planning approach through the case of Indian power sector in Mechanical and Process Engineering''. University of Duisburg-Essen; 2001.

[22] A. Mohammad, H. Mohammad, A. Hayati, A. Hasimah , M. Pauzi M, Faridah ''Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review''. Renewable and Sustainable Energy Reviews **70**, 1108–1118, 2017.

[23] D. Moher, A. Liberati, J. Tetzlaff, G. Altman '' Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement''. Energy, **8**, 336-341, 2010.

[24] MAH. Monda. ''Implications of renewable energy technologies in the Bangladesh power sector: long-term planning strategies in Institute of Agricultural Engineering/Center for Development Research (ZEF)'' University of Bonn; 2010.

[25] M. Mourshed, A. Quddus ''renewable energy RD&D expenditure and CO_2 emissions in 15 European countries''. International Journal of Energy Management; **3**(2):187–202, 2009.

[26] K. Moustris, K. Kavadias, D. Zafirakis, J. Kaldellis. ''Medium, short and very short-term prognosis of load demand for the Greek Island of Tilos using artificial neural networks and human thermal comfort-discomfort biometeorological data''. Renewable Energy **147**, 100-109, 2020.

[27] Q. Muhammad; K Abbas ''A review on artificial intelligence-based load demand forecasting techniques for smart grid and buildings'' Renewable and Sustainable Energy Reviews **50**, 1352–1372, 2015. Doi: 10.1016/j.rser.2015.04.065

[28] Q. Nguyen '' Long term optimization of energy supply and demand in Vietnam with special reference to the potential of renewable energy''. University of Oldenburg; 2005.

[29] C. Pablo., J. De Paz, J. Bajo, ''Agent-based tool to reduce the maintenance cost of energy distribution networks''. Knowledge Information System **54,** 659–675, 2018. Doi: 10.1016/j.rser.2018.04.054

[30] S. Pfenninger, A. Hawkes, J. Keirstead ''Energy systems modeling for twenty-first century energy challenges''. Renewable Sustainable Energy Review; **33**, 74–86. 2014. Doi: 10.1016/j.rser.2014.04.033

[31] V. Ramanathan, Y. Feng. ''On avoiding dangerous anthropogenic interference with the climate system: Formidable challenges ahead'' Production National Academic Science;**105**(38):14245–50, 2008.

[32] T.L Saaty ''The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation''. McGraw-Hill, New York, 1980

[33] M. Shahid., M. Aslam, A. Khan, ''An efficient model based on artificial bee colony optimization algorithm with Neural Networks for electric load forecasting''. Neural Computer & Application **25,** 1967–1978, 2014.

[34] S. Solomon ''Irreversible climate change due to carbon dioxide emissions''. Production National Academic Science;**106**(6):1704–9, 2009.

[35] L. Suganthi, A. Samuel ''Energy models for demand forecasting—A review''. Renew Sustainable Energy Review;**16**(2):1223–40, 2012.

[36] D. Tonelli, A. Hackam, Garg ''Primer on systematic review and meta-analysis'', Energy, **473**, 217–233, 2009.

[37] A. Valentina, N. Antonio, V. Roberto, V. Daniele, A. Rossano, D. Corrado, Z. Pietro ''A bottom-up spatially explicit methodology to estimate the space heating demand of the building stock at regional scale''. Energy & Buildings **206**, 109-581, 2020.

[38] R. Weron ''Modeling and forecasting electricity loads and prices: a statistical approach'' **403**. John Wiley & Sons; 2007.

[39] Z. Zhang, T. Gao, Y. Zheng, Q. Ma, M. Wang, X. Gao, ''A bibliometric analysis of biodiesel research during 1991–2015'', Journal of Material Cycles Waste Management. **20**, 10–18, 2018. Doi: 10.1007/s10163-016-0575-z

Manufacturing Industry in the Age of Fourth Industrial Revolution using a Dematel Approach

Ndala Yves Mulongo[*,1], Grace Kanakana-Katumba[2], Ernest Mnkandla[1]

[1]*School of Computing, Department of Information Systems, University of South Africa, Johannesburg, 2000, South Africa*

[2]*Faculty of Engineering and the Built Environment, Tshwane University of Technology, Pretoria West, 0183, South Africa*

A B S T R A C T

The concept of Fourth industrial revolution covers various digital technologies such as, Big data, cloud computing, artificial intelligence, the Internet of Things, and etc. But, most of South African business organisations have resisted it thus far – or at worst dismissed it – because they always considered that traditional ways worked perfectly. Manufacturing industry is in unknown territory, with little information about whether the projected economic consequences would be and how to survive in the midst of this global health emergency. To this end, Digital technologies have risen to save the day amid all this misery. Although, up to this point, most firms have resisted industry 4.0 innovations. In view of these, this study made significant contributions by firstly determining key challenges to the introduction of Industry 4.0 initiatives in manufacturing industry in the context of South Africa. Secondly, the causal correlations involving 12 hindrances are measured using the Grey-DEMATEL method. Finally, it highlights the concern that threats to disseminating technological advancement towards smart manufacturing from fully operating in the midst of such pandemic, if, Industry 4.0 revolution innovations still ignored.

1. Introduction

Fourth industrial revolution, the German Strategic Initiative, aims to build digital companies that deploy numerous innovative technologies such as big data analytics, Artificial Intelligence (AI), Machine learning; deep learning; additive manufacturing, cloud computing, robotic systems to meet cyber physical structures and machinery interfaces that lead towards an ecologically sound, financially and social competitive production model. To date, the Fourth Industrial Revolution, initially emerged in November 2011 via a report published by the German government, as a result of an action plan for the 2020 High Technology Strategy [1]. The industry 1.0, the industry 2.0 centered on large scale production, the third industrial age concentrated on computing [2]. And thus, the fourth industrial age has been the digital age for manufacturing something that emerged from the advanced cloud computing and programming of any of the production spheres [3]. The invention of combustion engine technology was a cruise missile. Fourth industrial revolution is focused on the development of digital production called a "intelligent manufacturing plant," meaning artificial communication across industries, process automation, industrial

activity efficiency and its integration, collaboration with consumers and distributors and the implementation of competitive business strategies [4]. The key feature of Fourth Industrial revolution is smart cyber-physical systems (CPS)-based networks [5]. Fourth industrial revolution combines the virtual and the physical environment with cyber-physical systems, which provides businesses with better productivity and efficiency [6]. the cyber- physical systems include intelligent equipment, data centers and structures that can independently share information, intrigue virtual control system of the manufacturing plant [7]. Further, fourth industrial revolution combines the utilisation of Big Data, IoT and AI in one platform and has a huge socioeconomic gain opportunity for businesses [8]. Internet of things influenced the way cyber-physical systems communicates, tracks, regulates and manages, enabling process and device alignment around the manufacturing industries and technology [9]. This leads to improved coordination and collaboration in a modern and informed way with each manufacturing facility, which revolutionizes production, boosts service delivery, improves distribution and facilitates the coordination of assets efficiently and at a lower cost [10], [11].

2. A Theoretical conceptual Fourth Industrial revolution model for Manufacturing Industry

Fourth industrial revolution could be divided, at best in dual different sections in accordance with their main purposes [12]. Actually, the emphasis on the Figure 1 below, is on- 'Front End Technology' of Fourth Industrial revolution, which takes a view of the development of production operations relying on innovative technologies (intelligent production) and of the distribution of goods (Intelligent Goods) [13], [14]. This also considers the manner in which commodities and goods are supplied (intelligent value chain) [15]. The modern way in which employees operate on the basis of emerging market (Intelligent work) [16]. We label "front- innovations" since the four "intelligent" aspects are organizational and consumer demands. Consequently, providing a digital end— framework for the supply chain of the businesses, as seen in the graphical arrow of Figure.1.
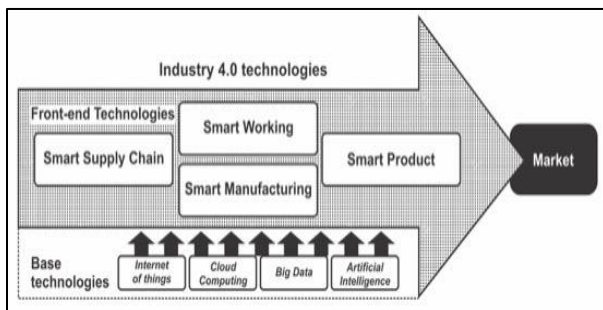

Figure 1: Industry 4.0 model adapted from [17], [18]

It should be mentioned that the core aspect of the front-development structure is the output of Intelligent Manufacturing whilst the other measures are connected to it. The front-end structure is based onto another surface in Fig. 1: 'dependent devices' including technology that include front-end technologies communication and information. The very last phase makes the Fourth industrial age definition, which separates it from preceding phases of industry. dependent technology enables for the incorporation of front- technology into a fully functional production mechanism [17], [18]. In the following paragraphs this paper will describe every level suggested under Figure. 1. The goal is to explain in which manner these innovations can be used by production companies and how they can adopt introduction trends.

2.1. Industry 4.0
2.1.1. Digital production process

Intelligent production corresponds to the concept of how to use data, patterns, algorithms and mobile device controls for the optimization of the value chain in the manufacturing of finished goods. Intelligent management is a concern in most emerging economies, including the Western; Asian, and the United states. Intelligent production system is the implementation of this mechanism production approach [19], [20]. It is also called Fourth industrial age. Quick and efficient data and widely interconnected networks remain critical, yet algorithms and simulations are needed to make the most of the data. The intelligent production revolution is said to be in three stages: factory automation and plant-wide integration, the Utilisation of digital manufacturing, and the development of new business strategies.

In a variety of manufacturing industries which mostly manufacture customized goods, smart production technologies are especially required. Customer products have to be very sensitive to market conditions (just like the food manufacturing) so that they can adapt their production to fluctuations in sales and quantities that are very much dependent on increasing market demand [21], [22]. The drug industry has often needed to become less flexible as healthcare demands may be more routine, although there is more demand onto pharmaceutical companies, which needs more flexibility and less inventory in the buying expenditures of medicines – mostly domestic organizations [23]. In the long term, as just the concept of customized medicinal products will become more practical, hence, a much more sensitive output is required for particular individuals or user units, often with smaller groups. The drug industry aims to put new drugs on the market sooner, so that they can have a peculiar advantage and a very rapid response to epidemic / pandemic vacuum [24], [25]. The economic environment may be more evolving and the reaction to evolving demands for product quality may be quicker and the production activities may be managed more effectively.

2.1.2. Digital Manufacturing and Intelligent Goods

Intelligent Manufacturing systems function as a core element of inner operation activities in the heart of the Fourth industrial revolution [26]. Although Intelligent Business model takes account of the organizational value generated by goods in connection with embedded consumer knowledge and data through the manufacturing system [27]. Both measurements consider developments that directly affect the manufactured goods. Smart Manufacturing recognizes material production systems, whereas Intelligent Goods recognizes product-. This paper, thus, believe that Digital Production system is the first and first goal of Fourth industrial revolution, although its extension is Intelligent Item .. This concept reflects the recent chronological growth of the Fourth industry revolution model, which originates in high tech manufacturing processes, and is linked to the firm's other systems [28], [29]. With regard to the Intelligent Manufacturing layer, in this paper, the relevant technologies are divided into six key objectives: (i) vertical integration, (ii) virtualisation, (iii) optimization, (iv) data integrity, (v) agility and (vi). electrical energy management [30], [31].

The vertical integration of Production line includes integrated Information technology mechanisms integrating all the firm's centralized structures from the production floor to the middle and upper managerial structures which make decision- activities less reliant on manual intervention [32]. The very first course of action throughout the production line is to digitize all physical items and variables with detectors, gyroscopes and microprocessor regulators to achieve vertical integration [33]. Data will then be collected in the production line for production management. Mass production operation processes receive data from Structured cabling in operational intelligence systems, offering the Enterprise Resource Planning (ERP) system with output status. While all circuits are implemented correctly, data collected about the sales demands often flows inversely (downwind), actually moving business capital into the production order field [34]. Vertical integration also offers greater clarity and flexibility over the supply chain and tends to enhance the decision- on the

production line. To make the goods more adaptable, Smart Manufacturing consists of interconnected devices on a production line via processor--- information sharing [35]. The processor information sharing includes of a hardware and software satellite system, which allows machinery to identify with one another and facilitates integration in assembly lines [36]. Such computation is advanced in automated production, that in addition to PLC- sometimes takes information across all virtualized items in the production line and afterwards replicates operational mechanisms, considering many variables which might influence the manufacturing processes. Particularly supported by Smart Production, an improved workflow [37]. Intelligent machines could process information with greater accuracy than previously and improve performance while being much less fatiguing [38].

Firms could concentrate on different requirements while prioritizing the deployment of the above- intelligent manufacturing technology [39]. Recent studies developments tend to show that the manufacturing sector has specific benefits to production efficiency and businesses must therefore structurally consider using such technology to reach a higher maturity and intelligence for Fourth industrial revolution [40]. It therefore indicates that Smart Manufacturing systems could be integrated and synergized with Fourth industrial revolution. The potent implementation of Internet of things into the intelligent manufacturing systems leads in the integration of the physical items of the manufacturing plant with the digital components, which includes digital knowledge, artificial intelligence into the cyber Physical Systems [41], which is among the synergies between these systems). it is thus presuming the subsequent assumption about the use of intelligent manufacturing technologies.

Further, Intelligent Items 'front- end technology includes intelligent elements that allow virtual functions and services to be provided via product lines. For this instance, the technical abilities required for various levels of intelligent item are considered [42]. The integrated detectors enable product communication with certain artefacts and mechanisms in a computer system. Actuators may have external product surveillance ability that enables consumers to recognize the item quality and the consumption specifications. Cloud- items with systems integration can sometimes be operated via virtual remote interface. Products will still have optimisation features with computational algorithm, as well as the quality of products based upon computational modelling that render adjustments possible. Things can robotically design themselves utilizing machine learning. Such capability expand the functionality of products in the first place and generate new avenues for manufacturing companies. Product tracking therefore offers valuable information to suppliers who would collect these data to detect material consumption trends for both the consumer preference and production of new items. Then it allows automated PSS, wherein producers are able to provide new product and even sell goods as a service [43]. Whereas other businesses will concentrate on the individual dimension of digital technology, for example, fourth industrial revolution definition offers smart goods for an end user to be linked and incorporated with both Smart internal and Smart external goods [44].

2.1.3. Smart Supply Chain

Several more compatible classes of Fourth industrial revolution front- innovations are the intelligent manufacturing process and intelligent function. This paper found them independently from the intelligent manufacturing and digital Products as both are designed to contribute positively to production and to the finished goods whereas the intelligent distribution system and the intelligent job measurements are designed to make the additional business operations effective. Beyond the plant, the smart distribution network integrates technology to enable plant collaboration with local retailers in order to maximize raw and final supplier product distribution, with an impact on operating expenses and turnaround times [9]. Intelligent Functioning deems the technology inside the plant, on the other hand, to help staff 'activities in order to make them more competitive and versatile to meet production specifications. Both Intelligent Distribution system and Intelligent Management are deemed to be a front end as they both contribute directly to the firm's operational efficiency. First is to share, in real time, information on production orders with manufacturers and warehouses, supported by horizontal convergence supported by Intelligent Supply Chain technology [45]. In view of the internal- logistics systems involving institutional quality control processes of products and auto-guided transportation [46]. Certain technological apparatus is required to join suppliers to outer operations [47]. This is what digital services accomplish, as they offer fast on-demand access to cloud-related information, merging distributors and factories [48].

Digital job technological innovations, on the other hand, seek to improve the factory workers' conditions to increase their productive capacity [49] by providing them with direct access to sensitive information on the factory lines [50]. Human beings and robots are therefore regarded as an intertwined socio- system in the Fourth industrial revolution definition [51]. Fourth industrial revolution not only allows the device of the activities using wireless devices, something that enhances the mechanism of policy-making and increases knowledge accessibility, several factors that also lead to Intelligent Management [52]. Digital devices could also be regarded as part of the intelligent job because they enable decision-making. Enhanced and immersive information technology are the only advanced technologies that build selective and full cyberspace [53]. In the field of supply chain servicing, the artificial intelligence quickens learning for employees with an immediate simulation of upkeep activities [54], while increasing the environment helps employees with engaging, actual-time feedback on the required job measures. Such technologies develop virtual design features for process improvement activities, allowing to identify faults throughout products use through requiring physical designs [54], [55].

2.2. Industry 4.0 base technologies

This paper also looks at a second layer of fourth industrial revolution innovations, which is called 'key technology,' as they endorse all the other aspects of the 'intelligent' mentioned above. The basic technological innovations include the modern ICTs [56], including Internet of things systems, compute cluster, big data and analytics [57]. Such systems are considered as critical factors for being active across all aspects and in different models

of these aspects. This paper exploits the attributes of the Fourth industrial revolution to render interoperability effective and then provide the insight of the latest production systems. Cloud computing means integrating sensors and information technology in such a wireless internet setting [58]. Recent developments on the Cloud have effectively made it possible to communicate many artefacts to achieve this idea. It has also been helped by the reduced costs of sensor technology in previous decades [58]; that allowed any type of entity to be sensed and connected to a wider system [58]. Internet services offer access to a different unit of artificial intelligence assets on request for the computer system [6]. Its system is capable of stowing data in a remotely accessible Internet service provider [58].

Internet services also allow various machines to be combined as the devices does not have to be geographically close and can exchange information and organize operations [1]. The integration of Cloud computing and Blockchain makes the communication between different equipment, storing massive amounts of data, resulting in large data processing. Big data includes of device and object data processing, including readouts [2]. Alongside data analysis, for example data mining and big data analytics, this is regarded as one of the key factors in the industry 4.0, and a primary source of a competitive edge for the years ahead. The most important thing is the knowledge that it will produce. Big Data is required to construct the firm's virtual doubles, and research allows improved diagnostic ability to detect production issues before production takes place [3]. Big data with algorithms will help the self-organization of manufacturing processes and automate decision-making processes across all industrial sector dimensions. The four above mentioned technological innovations – Internet of things, blockchain, big data analytics have different functions. Internet of things helps to address connectivity problems across all production line artefacts and processes, while cloud systems offer convenient access to data and services. Finally, the main features of Big Data and Analytics are integrated application of Fourth industrial revolution because machine knowledge relies on the vast volume of data collected (big data) and the potential for analyses using sophisticated technology (analytical techniques). [54] [58].

Table 1: Challenges on the implementation of industry 4.0 (author's literature review)

Challenges	Code	Authors
Huge capital cost for industry 4.0 technologies implementation	H	[6], [15], [19] ; [26]. [27], [28], [29]. [30], [32], [33].
Deficit of clear economic value	D	[7]; [38]; [40], [42], [43], [45], [47], [52], [53],
Challenges about value-chain integration	C	[6]; [15]; [19]; [39].
Risk of security infringes	R	[6]; [8], [13], [15]. [19]; [34].
Poor degree of technology development)	M	[17], [19], [34], [54], [55], [58].
Inequality	I	[46]

Interruption to existing employments	I	[24]; [40]; [44]; [46].
Dearth of standards, regulations and forms of certification	D	[45]; [46].
Deficiency of infrastructure	D	[8]; [23]. [45]
Shortage of digital skills	S	[6]; [19]; [25].
Challenges in ensuring data quality	C	[11]; [32]; [37].
Absence of internal digital culture and training	A	[6]

3. Research Methodology

Identification of obstacles in real cases is a complex issue, as there are challenges among them. Advanced multi-criteria decision-making (MCDM) approaches are therefore required to examine obstacles that take into consideration of limitations and contrasting priorities. Studies were conducted very widely among related MCDM strategies, ISM and AHP. DEMATEL technique was lately also used by researchers. In general, ISM and DEMATEL methods for study of dependency variables are better than those of AHP. DEMATEL approach gives also a higher great deal of control than an ISM approach. DEMATEL separates variables into classes of causes and effects, forming causal relationships. In addition, DEMATEL could be used with samples of small size. In the combination of grey system theory with DEMATEL approach, ambiguous and unreliable knowledge latent in human decisions is taken into consideration. In this study, the grey-DEMATEL approach is used to examine key challenges to industry 4.0 technologies in manufacturing sector. In previous studies such as internal obstacle. The following sections discuss in depth the DEMATEL method and the theory of grey systems.

3.1. Grey's approach

The Grey approach was first suggested by [18] as a "grey" concept of a grey set. A limited quantity of data used during the grey system would generate acceptable results, which seems to be of great benefit to the grey system over other systems [22]. The grey statistics were translated to crisp digits using a modified CFCS system using a three-stage process listed below.

For a grey number: $\otimes X_{ij}^k = \left[\underline{\otimes X_{ij}^k}; \, \overline{\otimes X_{ij}^k} \right]$

STAGE 1 : Normalization

$$\underline{\overline{\otimes X_{ij}^k}} = \frac{\left(\underline{\otimes X_{ij}^k} - \min_j \underline{\otimes X_{ij}^k} \right)}{\triangle_{min}^{max}} \tag{1}$$

$$\overline{\otimes X_{ij}^k} = \frac{\left(\overline{\otimes X_{ij}^k} - \min_j \underline{\otimes X_{ij}^k} \right)}{\triangle_{min}^{max}} \tag{2}$$

where $= \triangle_{min}^{max} = \max_j \overline{\otimes X_{ij}^k} - \min_j \underline{\otimes X_{ij}^k}$

STAGE 2 : Determination of a total normalized crisp value

$$Y_{ij}^k = \frac{\otimes \overline{X_{ij}^k}\left(1 - \otimes \overline{X_{ij}^k}\right) + \overline{\otimes X_{ij}^k} \times \overline{\otimes X_{ij}^k}}{1 - \otimes \overline{X_{ij}^k} + \overline{\otimes X_{ij}^k}} \tag{3}$$

STAGE 3 : Computation of final crisp values

$$Z_{ij}^k = \min_j \otimes \underline{X}_{ij}^k + Y_{ij}^k \; \triangle_{min}^{max} \tag{4}$$

3.2. DEMATEL method

The Battelle Memorial Institute initially conducted DEMATEL via its research institute in Geneva [1] [2]. This approach is designed specifically for the study of dynamic structural equations that respond causatively to complex factors [3]. DEMATEL's basic idea is a philosophy of diagrams, that mostly helps one to understand the cause and effect of the process by separating and comparing the problems associated with it [4]. This is fundamentally a computational modelling method that illustrates interdependence relations and causes values among important variables in the context of a diagram and the diagram of cause and effect [5]. With this approach, most of components are segregated into causal and affected classes and the conceptual relation between rates of system variables is analysed [58]. DEMATEL was used for race, poverty, environmental development and energy issues [51]. This paper uses a solution methodology with the integration of a grey method to determine challenges in an effective way with stages 1; 2; 3 as presented below adapted from [54] [58]. Steps involving DEMATEL are as follows:

STAGE 1: Initial relation matrix

The first step is to set up an initial relation matrix based on decision makers' rating over given criteria [54]. The form of the matrix is shown in equation (5).

$$B = \begin{bmatrix} 0 & b_{12} & b_{1n} \\ b_{21} & 0 & b_{2n} \\ b_{n1} & bn(n-1) & 0 \end{bmatrix} \tag{5}$$

where, b$_{ij}$ is denoted as the degree to which the criterion i affects the criterion.

STAGE 2: Set up normalized direct-relation matrix

The normalized direct relation matrix is obtained through equations (6) and (7). All elements in this matrix lie between 1 and 0 matrix lie between 1 and 0

$$K = \frac{1}{\max\limits_{1 \le i \le n} \sum_{j=1}^n a_{ij}} \tag{6}$$

$$X = K \times B \tag{7}$$

where, X ¼ Normalized direct relation matrix; K ¼ the normalization factor; A ¼ Initial relation matrix

STAGE 3: Set up total relation matrix

In this step we need to set up total relation matrix M. The normalized matrix is processed by the formula in equation (8) where I denote the identity matrix.

$$M = X(1 - X)^{-1} \tag{8}$$

STAGE 4: Obtain sum of sum of rows and columns

D denotes the sum of rows and R denotes the sum of columns. This should be calculated through equations (9) and (10):

$$D = \left[\sum_{j=1}^n m_{ij}\right]_{n \times 1} \tag{9}$$

$$R = \left[\sum_{i=1}^n m_{ij}\right]_{1 \times n} \tag{10}$$

STAGE 5: Set up causal diagram

A causal and effect diagram is obtained here by means of the dataset consisting of (D þ R, D R).

$$(D + R. D - R) \tag{11}$$

4. Research Framework

On the basis of the grey DEMATEL approach, a five-phased model is formulated to investigate the nature of various challenges and to determine the interdependencies between obstacles for the adoption of industry 4.0 technologies in manufacturing sector as shown in Figure 2 below. In fact, in this study, the DEMATEL approach was employed to examine the various obstacles to the implementation of industry 4.0 in manufacturing sector. Three industry practitioners' executives from X firm were handed over the questionnaires. The firm X is one of the four first round demonstration companies that is seeking ways to adopt and fully implement industry 4.0 technologies in their production line operations. Their yearly production accounts for more than 11% of the South African Gross Domestic Product In addition to three industry practitioners' executives, the questioners were handed to an academic research expert who has been active in research and development in manufacturing field over the past two decades. It was n agreed that manufacturing productivity needed to be improved, so each analyst was pleased to attend our questionnaires. In this context, our suggested approach assesses the obstacles to the introduction of industry 4.0 technologies that meet their requirements [54], [58]. An example of the application of the DEMATEL approach is a case study of the challenges to the effective introduction of fourth industrial revolution technologies in South African manufacturing firm. Firstly, a number of impediments were pinpointed after a thorough review of the literature and consultation with practitioners. Then surveys had been sent via emails to the selected industry practitioners executives from the local South African manufacturing industry.t When the answers were delivered, Grey-DEMATEL approach was used and tested, details of which are listed in the next section below. The methodology used in this paper has been extended to real South African manufacturing businesses, as these were aimed to accurately reflect the nature of emerging and advanced economies.
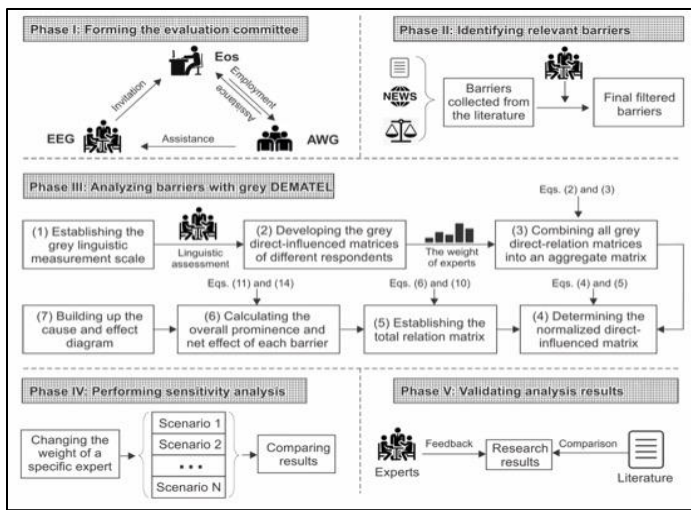
Figure 2: Proposed framework adapted from [54] [58]

The study conducted in this paper inspired businesses from such dynamic economies environment to decide where the problems they faced differentiated and where they collaborated. There are Five steps that explain the suggested methodology:

4.1. Creation of Evaluation committee

The process begins with the establishment of an initial assessment panel that mostly play a key role in the obstacle measurement. The assessment panel consists of three tasks: assessment administrators (EOs), the consultant analysis panel (EEG) and the support working committee (AWG) [54], [58].

4.2. Identification of potential obstacles

Obstacles to the implementation industry 4.0 initiatives are identified in a sustainable manner. The first is the array of obstacles from three sources: scientific literature, journal papers and government policies and regulatory regulations. The expert assessment group then extracts the compiled obstacles to obtain a specific list. Finally, the filtered impediments are recategorized across various aspects of industry 4.0 technologies and the definitive obstacle database is built [54], [58].

4.3. Using Grey DEMATEL to analyse the challenges

This section analyses the obstacles listed in stage II thoroughly. EEG specialists have to describe the relationship of any set of impediments with linguistic terms. The aggregate importance of each obstacle and the net impact is then determined by grey DEMATEL [54], [58].

4.4. Performing Sensitivity Analysis

This study provides a sensitivity analysis to assess the reliability of experiment findings, which could provide an overview into whether the outcomes achieved are affected by the possible biases of other consultants. A standard method of conducting sensitivity analysis is to adjust the weight of a single individual and then analyse the variations in the tests. This analysis applies this approach and establishes various scenarios based on the weights of various specialists [54], [58].

4.5. Confirmation of research results

Assessment organizers verify analytical findings in the last process with positive input from other specialists and the contrast with current literature.

Table 2: Grey linguistic scale for expert's responses.

Linguistic term	Grey numbers	Influence score
No influence	[0;0]	0
Very low influence	[0;0.25]	1
Low influence	[0.25;0.5]	2
High influence	[0.5;0.75]	3
Very high influence	[0.75;1.0]	4

Table 3: Linguistic scale for importance of expert's response.

Linguistic term	Grey numbers	Influence score
No influence	[0;0.3]	0
Very low influence	[0.3;0.5]	1
Low influence	[0.4;0.7]	2
High influence	[0.5;0.9]	3
Very high influence	[0.7;1.0]	4

Table 4: Direct-relation matrix for challenges on the implementation on Industry 4.0 by Expert 1

	H	D	C	R	M	I	I	D	D	S	C	A
H	O	1	3	2	2	3	3	3	2	1	0	3
D	2	O	3	3	3	2	2	3	3	2	4	4
C	3	3	O	2	2	1	0	0	4	2	3	1
R	3	4	1	O	1	0	2	3	0	1	4	3
M	3	2	2	3	O	2	2	3	4	1	2	4
I	3	2	3	3	4	O	2	3	3	4	4	2
I	4	3	3	4	2	2	O	3	3	4	4	2

D	4	3	2	2	3	4	1	O	1	4	3	3
D	4	4	2	4	4	4	3	3	O	2	2	1
S	3	0	3	2	4	4	0	0	1	O	2	2
C	4	0	1	1	3	3	3	2	0	4	O	3
A	3	1	2	2	3	3	3	4	4	4	1	O

Table 5: Direct-relation matrix for challenges on the implementation on Industry 4.0 by Expert 2

	H	D	C	R	M	I	I	D	D	S	C	A
H	O	1	2	2	4	4	2	1	0	0	0	2
D	3	O	4	4	4	1	1	3	3	4	2	3
C	4	4	O	1	1	3	1	1	3	2	2	2
R	2	3	3	O	3	2	4	2	4	2	3	2
M	4	3	3	2	O	1	1	4	3	2	1	4
I	3	3	4	2	2	O	3	1	2	4	4	3
I	3	4	4	3	1	2	O	4	4	3	3	4
D	2	1	0	0	2	4	2	O	3	3	2	2
D	3	3	1	3	3	2	4	4	O	1	2	3
S	4	4	4	3	2	2	1	1	3	O	3	4
C	2	2	2	3	3	4	2	3	4	3	O	2
A	4	3	3	2	2	0	1	0	3	3	2	O

Table 6: Direct-relation matrix for challenges on the implementation on Industry 4.0 by Expert 3

	H	D	C	R	M	I	I	D	D	S	C	A
H	O	2	2	3	3	3	1	1	4	4	1	2
D	4	O	4	4	4	1	1	2	2	3	2	2
C	3	4	O	1	1	0	2	2	3	3	4	3
R	4	4	3	O	2	2	3	3	3	2	2	2
M	1	0	0	4	O	3	3	2	2	4	3	3
I	4	4	4	3	3	O	3	2	2	1	0	2
I	3	2	2	3	1	1	O	4	4	4	4	0
D	1	0	0	2	3	3	3	O	1	3	3	2
D	3	3	3	2	2	1	1	0	O	4	3	4
S	4	2	2	3	3	3	3	3	2	O	1	0
C	2	2	3	3	2	2	3	2	1	3	O	2
A	4	2	3	3	4	4	2	0	1	2	1	O

Table 7: Direct-relation matrix for challenges on the implementation on Industry 4.0 by Expert 4

	H	D	C	R	M	I	I	D	D	S	C	A
H	O	2	2	3	3	4	4	2	3	2	2	1
D	3	O	3	3	3	2	2	3	3	2	4	4
C	2	2	O	2	3	1	0	0	4	2	3	0
R	1	3	3	O	1	0	3	3	3	2	4	2
M	1	3	4	3	O	2	2	3	4	1	2	2
I	2	3	2	3	4	O	2	3	3	4	4	3
I	3	1	0	0	3	1	O	1	3	3	2	3
D	3	2	4	4	4	4	1	O	1	4	3	3
D	2	1	3	4	4	4	3	3	O	2	2	2
S	4	2	3	3	3	3	0	1	4	O	3	3
C	3	2	4	1	3	3	3	2	0	4	O	4
A	2	3	4	0	0	1	1	2	3	3	2	O

Table 8: Average grey relation matrix for challenges on the implementation of Industry 4.0

	H	D	C	R	M	I	I	D	D	S	C	A
H	O	1.5	2.25	2.5	3	3.5	2.5	1.75	2.25	1.75	0.75	2
D	3	O	3.5	3.5	3.5	1.5	1.5	2.75	2.75	2.75	3	3.25
C	3	3.25	O	1.5	1.75	1.25	0.75	0.75	3.5	2.25	3	1.5

R	2.5	3.5	2.5	O	1.75	1	3	2.75	2.5	1.75	3.25	2.25
M	2.25	2	2.25	3	O	2	2	3	3.25	2	2	3.25
I	3	3	3.25	2.75	3.25	O	2.5	2.25	2.5	3.25	3	2.5
I	3.25	2.5	2.25	2.5	1.75	1.5	O	3	3.5	2.5	3.25	2.25
D	2.5	1.5	1.5	2	3	3.75	1.75	O	1.5	3.5	2.75	2.5
D	3	2.75	2.25	3.25	3.25	2.75	2.75	2.5	O	2.25	2.25	2.5
S	3.75	2	3	2.75	3	3	1	1.25	2.5	O	2.25	2.25
C	2.75	1.5	2.5	2	2.75	3	2.75	2.25	1.25	2.5	O	2.75
A	3.25	2.25	3	1.75	2.25	2	1.75	1.5	2.75	3	1.5	O

Table 9: Normalized direct-relation matrix for challenges on the implementation of Industry 4.0

	H	D	C	R	M	I	I	D	D	S	C	A
H	0	0.048	0.072	0.08	0.096	0.112	0.08	0.056	0.072	0.056	0.024	0.064
D	0.096	0	0.112	0.112	0.112	0.048	0.048	0.088	0.088	0.088	0.096	0.104
C	0.096	0.104	0	0.048	0.056	0.04	0.024	0.024	0.112	0.072	0.096	0.048
R	0.08	0.112	0.08	0	0.056	0.032	0.096	0.088	0.08	0.056	0.104	0.072
M	0.072	0.064	0.072	0.096	0	0.064	0.064	0.096	0.104	0.064	0.064	0.104
I	0.096	0.096	0.104	0.088	0.104	0	0.08	0.072	0.08	0.104	0.096	0.08
I	0.104	0.08	0.072	0.08	0.056	0.048	0	0.096	0.112	0.08	0.104	0.072
D	0.08	0.048	0.048	0.048	0.096	0.12	0.056	0	0.048	0.112	0.088	0.08
D	0.096	0.088	0.072	0.104	0.104	0.088	0.088	0.08	0	0.072	0.072	0.08
S	0.12	0.064	0.096	0.088	0.096	0.096	0.032	0.04	0.08	0	0.072	0.072
C	0.088	0.048	0.08	0.064	0.088	0.096	0.088	0.072	0.04	0.08	0	0.088
A	0.104	0.072	0.096	0.056	0.072	0.064	0.056	0.048	0.088	0.096	0.048	0

Table 10: Total relation matrix for challenges on the implementation of Industry 4.0

	H	D	C	R	M	I	I	D	D	S	C	A
H	0.1341	0.1418	0.2269	0.2416	0.1776	0.2589	0.2266	0.3574	0.3218	0.2834	0.2266	0.1669
D	0.1721	0.0896	0.2025	0.1666	0.1656	0.2499	0.1939	0.3464	0.1791	0.1927	0.1966	0.0772
C	0.2156	0.1962	0.1592	0.1243	0.0895	0.2256	0.1988	0.3108	0.0857	0.1564	0.2238	0.1201
R	0.1234	0.1449	0.0996	0.1406	0.1314	0.1898	0.1197	0.2725	0.1168	0.2357	0.2013	0.0659
M	0.0878	0 .0775	0.0896	0.1346	0.1265	0.1275	0.1893	0.2156	0.0878	0.2437	0.1681	0.0562
I	0.1718	0.1148	0.1375	0.1605	0.1899	0.2387	0.2103	0.1438	0.0664	0.2078	0.1727	0.0517
I	0.1653	0.1234	0.1789	0.0568	0.0491	0.1989	0.2355	0.2197	0.1021	0.1804	0.0878	0.0798
D	0.1944	0.1071	0.2272	0.1316	0.0729	0.1028	0.1916	0.1378	0.1199	0.1598	0.0981	0.0492
D	0.1883	0.1984	0.2165	0.1282	0.0238	0.0766	0.1335	0.1046	0.1017	0.2864	0.0543	0.0376
S	0.0848	0.0632	0.1078	0.0597	0.0346	0.0991	0.1405	0.2647	0.0607	0.2405	0.1174	0.1611
C	0.0654	0.0318	0.0415	0.0835	0.0412	0.0823	0.0827	0.2965	0.2211	0.2216	0.0507	0.2176
A	0.1128	0.0638	0.0825	0.1531	0.0743	0.1264	0.0848	0.2292	0.2493	0.2378	0.0362	0.1848

Table 11: Degree of prominence and net cause/effect values for challenges on the implementation of Industry 4.0

Challenges	R sum	D sum	R + D	R-D
H	1,7158	2,7636	4.4794	-1.0478
D	1.3525	2,2322	3.5847	-0.8797
C	1,7697	2,1060	3.8757	-0.3363
R	1,5811	1,8416	3.4227	-0.2605
M	1,1764	1,6042	2.7806	-0.4278
I	1,9765	1,8659	3.8424	0.1106
I	2,0072	1,6777	3.6849	0.3295
D	2,8990	1,5924	4.4914	1.3066
D	1,7124	1,5499	3.2623	0.1625
S	2,6462	1,4341	4.0803	1.2121
C	1,6336	1,4359	3.0695	0.1977
A	1,2681	1,6350	2.9031	-0.3669

5. Discussion and Conclusion

The concept of Fourth industrial revolution covers various digital technologies such as, Big data, cloud computing, artificial intelligence, the Internet of Things, and etc. Manufacturing industry is in unknown territory, with little information about whether the projected economic consequences would be and how to survive in the midst of this global health emergency. To this end, Digital technologies have risen to save the day amid all this misery. Although, up to this point, most firms have resisted industry 4.0 innovations. In view of these, this study made significant contributions by firstly determining key challenges to the introduction of Industry 4.0 initiatives in manufacturing industry in the context of South Africa. Secondly, the causal correlations involving 12 hindrances are measured using the Grey-DEMATEL method. Finally, it highlights the concern that threats to disseminating technological advancement in South Africa might keep manufacturing sector from fully operating in the midst of such pandemic, if, Industry 4.0 revolution innovations still ignored. To this end, this section examines the study results as presented in the preceding section. The results show that the obstacles Huge capital cost for industry 4.0 technologies implementation; Deficit of clear economic value; Challenges about value-chain integration; Risk of security infringes; Poor degree of technology development; and Absence of internal digital culture and training fall within the class of effect group. Since, such obstacles purport to be easily affected by all the other impediments, as their (R – D) values are negative.

On the other hand, impediments. Inequality; Interruption to existing employments; Dearth of standards, regulations and forms of certification; Deficiency of infrastructure; Shortage of digital skills; Challenges in ensuring data quality; and Absence of internal digital culture and training fit into the category of causal obstacles as their (R – D) values are positive. Further, causal obstacles have a strong influence towards other obstacles and should therefore be regarded with much attention and must be eliminated at the end of the day. Hence, it is critical that Management teams at the X South African manufacturing company should locate and prepare to eliminate such obstacles so that Industry 4.0 technologies can be implemented smoothly within their organizations. Because, these obstacles have a somewhat prominent effect (R) than impact (D). The higher the value of R-D, the significantly higher the impact. Furthermore, obstacles with strong prominence values (R + D) have a significant impact towards other impediments and are therefore influenced by other obstacles. These are the obstacles which need to be taken into consideration in the short and medium term because they are fundamental obstacles. The obstacle, Dearth of standards, regulations and forms of certification does have the highest rating (R+D) and therefore this obstacle shows the strongest causal relationship with other variables. Huge capital cost for industry 4.0 technologies implementation and Shortage of digital skills are ranked after. It is therefore obvious from the outcomes that Dearth of standards, regulations and forms of certification needs to be dealt with in the short to medium term. In conversations with specialists, Dearth of standards, regulations and forms of certification; Huge capital cost for industry 4.0 technologies implementation; and Shortage of digital skills were validated as the main obstacles. Further, fourth industrial revolution innovations 'underlying cause of obstacles is the "paucity of a digital strategy alongside resource shortages." Management teams should therefore creatively prepare to direct organization activities and invest in assets that could support the migration to Fourth industrial revolution. Trying to tackle fourth industrial revolution's underlying cause of challenges would allow businesses to succeed in the technological revolution. Such considerations are related to product growth and health concerns. Management teams should therefore emphasize on infrastructure development in support of technology mechanisms in order to prevent delays with Industry 4.0 technology rollout. It was also observed that fourth industrial revolution has been advanced in emerging countries at a company level depending on individual corporate projects instead of on institutional and organized measures to date. This makes marketplace and legislative conditions an important obstacle. Lastly, the removal of these obstacles may very well make a difference in manufacturing sector and assist the sector to respond to the ever-growing competition within the global market [58].

References

[1] A. Gabus, E. Fontela,, "World Problems, an Invitation to Further Thought within the Framework of DEMATEL". Battelle Geneva Research Centre, Switzerland, Geneva, 1972.

[2] A. Haddud, A. DeSouza, A. Khare, H. Lee, "Examining potential benefits and challenges associated with the Internet of Things integration in supply chains". Journal of Manufacturing. Technology Management **28** (8), 1055–1085, 2017. Doi: 10.1108/JMTM-05-2017-0094

[3] A. Majeed, T. Rupasinghe, "Internet of Things (IoT) embedded future supply chains for industry 4.0: an assessment from an ERP-based fashion apparel and footwear industry". International. Journal of Supply Chain Management **6** (1), 25–40, 2017.

[4] B. Buntz, "Top 10 Barriers for Adoption of the Internet of Things".2016. Retrieved July, 2018, from. http://www.ioti.com/security/top-10-reasons-people-aren-t-embracingiot

[5] B. Chang., C. Chang, H. Wu, "Fuzzy DEMATEL method for developing supplier selection criteria". Expert System with Application **38** (3), 1850–1858, 2011. Doi: 10.1016/j.eswa.2010.07.114

[6] C Bai, J. Sarkis, "A grey-based DEMATEL model for evaluating business process management critical success factors". International Journal of Production Economy. **146** (1), 281–292, 2013. Doi: 10.1016/j.eco.2013.07.146

[7] C. Schro€der, "The Challenges of Industry 4.0 for Small and Medium-Sized Enterprises". 2016. Friedrich-Ebert-Stiftung, Bonn.

[8] C. Su, D. Horng, M. Tseng, A. Chiu, K. Wu, H. Chen, "Improving sustainable supply chain management using a novel hierarchical grey-DEMATEL approach". Journal of Cleaner. Production. **134**, 469–481, 2016. Doi: 10.1016/j.jclepro.2015.05.080

[9] D Bogle.. "A perspective on smart process manufacturing research challenges for process systems engineers". Engineering 2017; **3**(2):161–5, 2017. Doi: 10.1016/J.ENG.2017.02.003

[10] D. O'Halloran, E. Kvochko, "Industrial internet of things: unleashing the potential of connected products and services", In: World Economic Forum, 2015.

[11] E. Brynjolfsson, "The productivity paradox of information technology Commun". Association for Computing Machinery **36** (12), 66–77, 1993.

[12] F. Kache, S. Seuring, "Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management". International Journal of Production Management. **37** (1), 10–36, 2017. Doi: 10.1016/j.pro.2017.02.037

[13] G. Tzeng, C. Chiang, C. Li, "Evaluating intertwined effects in e-learning pro- grams: a novel hybrid MCDM model based on factor analysis and DEMATEL". Expert System with Application. **32** (4), 1028-1044, 2007. Doi: 10.1016/j.eswa.2006.02.004

[14] H. Agermann, W. Wahlster, J. Helbig, "Recommendations for implementing the strategic initia- tive INDUSTRIE 4.0: Securing the future of German manufacturing industry"; Final report of the Industrie 4.0 Working Group. Forschungsunion., 2013.

[15] H. Kagermann, W. Wahlster, J.Helbig. "Securing the Future of German Manufacturing Industry: Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0". Final Report of the Industrie 4.0 Working Group, Forschungsunion im Stifterverband für die Deutsche Wirtschaft e.V., Berlin, 2013.

[16] H. Lee, G. Tzeng, W. Yeih, Y. Wang, S. Yang, "Revised DEMATEL: resolving the infeasibility of DEMATEL". Application. Mathematic Model. **37** (10–11), 6746–6757,2013. Doi: 10.1016/j.eswa.2013.02.037

[17] I. Lee, K. Lee, "The internet of things (IoT): applications, investments, and challenges for enterprises". Bus. Horiz. **58** (4), 431–440, 2015. Doi: 10.1016/j.bushor.2015.03.008

[18] J. Cohen., H. Kharas, "Using big data and artificial intelligence to accelerate global development".2018. Brookings Institution Accessed 04.07.19. https://www.brookings.edu/ research/using-big-data-and-artificial-intelligence-to-accelerate-global- development/.

[19] J. Davis , T. Edgar, J. Porter, J. Bernaden, M. Sarli "Smart manufacturing, manufacturing intelligence and demand-dynamic performance". Computing Chemical Engineering;**47**:145–56, 2012.

[20] J. Gong., F. You "Resilient design and operations of process systems: nonlinear adaptive robust optimization model and algorithm for resilience analysis and enhancement". Computing Chemical Engineering; **116**, 231–52, 2018.

[21] J. Litster "Design and processing of particulate products." Cambridge: Cambridge University Press; 2016.

[22] J. Ryan, R. Watson, "Research challenges for the internet of things: what role can OR play?" Systems **5** (1), 24, 2017.

[23] J. Srai, C. Badman, M. Krumme , M. Futran, C. Johnston "Future supply chains enabled by continuous processing—opportunities and challenges". May 20– 21, 2014 Continuous Manufacturing Symposium. Journal of Pharmaceutical Science;**104** (3):840–9, 2015.

[24] J.Shao, M. Taisch, M., Ortega-Mier, "A grey-Decision-Making trial and evaluation laboratory (DEMATEL) analysis on the barriers between environmentally friendly products and consumers: practitioners' viewpoints on the European automobile industry". Journal of Cleaner Production, **112**, 3185–3194, 2016.

[25] K. Chen, S. Yang, S Tan, H. Zhao, L. He, G. Zhang, Y. Yang,"The data allocation strategy based on load in NoSQL database". In: Applied Mechanics and Materials, **513**, Trans Tech Publications, pp. 1464–1469, 2014.

[26] K. Govindan,. R Khodaverdi., A. Vafadarnikjoo, "A grey DEMATEL approach to develop third-party logistics provider selection criteria". Industrial Management and Data System. **116** (4), 690–722, 2016.

[27] K. Schwab "The Fourth Industrial Revolution". 2017. Crown Business.

[28] l. Bacudio., M. Benjamin, R. Eusebio, S. Holaysan, M. Promentilla, K. Yu, K., Aviso,"Analyzing barriers to implementing industrial sym- biosis networks using DEMATEL". Sustainable Production Consumption **7**, 57–65, 2016

[29] L. Cui,. H. Chan, Y. Zhou, J. Dai, J. Lim "Exploring critical factors of green business failure based on Grey-Decision Making Trial and Evaluation Laboratory (DEMATEL)". Journal of Business Research, 98, 450–461, 2019.

[30] L. Dalenogare, G. Benitez, F. Ayala, G. Frank "The expected contribution of Industry 4.0 technologies for industrial performance". International Journal of Production Economics. https://doi.org/10.1016/j.ijpe.2018.08.019.

[31] M. Bogers, R. Hadar, A. Bilberg "Additive manufacturing for consumer-centric business models: implications for supply chains in consumer goods manufacturing." Technology Forecast Social Change; **102**, 225–39, 2016.

[32] M. Breunig, R. Kelly, R. Mathis, D. Wee, "Getting the Most Out of Industry 4.0".2016. Retrieved March **12**, 2018, from. https://www.mckinsey.com/business-functions /operations/our-insights/industry-40-looking-beyond-the-initial-hype.

[33] M. Hung, "IoT Implementation and Management - from the Edge to the Cloud" 2016. [WWW Document]. Gartner. URL. https://www.gartner.com/technology/research /internet-of-things/report/, 7.23.18.

[34] M. Löffler, A. Tschiesner "The Internet of things and the future of manufacturing".2013. McKinsey & Company Accessed in April 2019. https://www.mckinsey.com/business- functions/digital-mckinsey/our-insights/the-internet-of-things-and-the-future-of- manufacturing.

[35] M. Tseng, "A causal and effect decision making model of service quality expectation using grey-fuzzy DEMATEL approach". Expert Syst. With Appl. **36** (4), 7738–7748, 2009. Doi: 10.1016/j.eswa.2008.09.011

[56] N. Khan, I. Yaqoob, I. Hashem, Z. Inayat, M. Ali, W. Kamaleldin, W. "Big data: survey, technologies, opportunities, and challenges". Science World Journal. 2014.

[37] P. Inman, "Coronavirus 'could cost global economy \$1.1tn in lost income". The Guardian Wed 19 Feb 2020.

[38] P. Inman, "Will the coronavirus outbreak derail the global economy? The Guardia. 2020.

[39] Q. Zhou, W. Huang, Y. Zhang, "Identifying critical success factors in emergency management using a fuzzy DEMATEL method". Safety. Science. 49 (2), 243-252, 2011. Doi: 10.1016/j.ssci.2010.08.005

[40] R. Baxendale, R. Braatz, B. Hodnett, K. Jensen., M. Johnson, P. Sharratt "Achieving continuous manufacturing: technologies and approaches for synthesis, workup, and isolation of drug substance". May 20–21, 2014 Continuous Manufacturing Symposium. Journal of Pharmaceutical Science ;**104**(3):781–91, 2015. Doi: 10.1016/j.ssci.2015.03.104

[41] R. Geissbauer,. J. Vedso, S.Schrauf,. "Global Industry 4.0 Survey, Industry 4.0: Building the digital enterprise, PWC."2016.

[42] R. Geissbauer,. S. Schrauf,. V. Koch,."Industry 4.0: Opportunities and Challenges of Industry 4.0. 2014.

[43] R. Hutt, "The economic effects of the COVID-19 coronavirus around the world. The World" 2020.

[44] R. Rajesh, V. Ravi, "Modeling enablers of supply chain risk mitigation in electronic supply chains: A Grey–DEMATEL approach". Computers & Industrial Engineering, **87**, 126–139, 2015. Doi: 10.1016/j.comp.2015.03.087

[45] R. Sousa, N. Shah, L. Papageorgiou. "Supply chains of high-value low-volume products". Process systems engineering: volume 4: supply chain optimization. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA; 2008. p. 1–27, 2008.

[46] R. Sousa, S. Liu, G. Papageorgiou., N. Shah "Global supply chain planning for pharmaceuticals". Chemical Engineering Research Design 2011;**89**(11):2396–409, 2011. Doi: 10.1016/j.chem.2011.03.089

[47] R.Y. Zhong, X. Xu, E. Klotz, "Newman, Intelligent Manufacturing in the Context of Industry 4.0: A Review", Engineering, vol. 3, pp. 616–630., 2017.

[48] S. Byrn, M. Futran, H. Thomas, E. Jayjock, N. Maron, R. Meyer "Achieving continuous manufacturing for final dosage formation: challenges and how to meet them". May 20–21, 2014 Continuous Manufacturing Symposium. Journal of Pharmaceutical Science;**104**(3):792–802, 2015. Doi: 10.1016/j.phar.2015.03.104

[49] S. Kamble, A. Gunasekaran, R. Sharma, "Analysis of the driving and dependence power of barriers to adopt industry 4.0 in Indian manufacturing industry".2018. Industrial Internet, PricewaterhouseCoopers. Freudenberg IT. Available at: https://www.pwc.nl/en/assets/documentAhlers, T, 2015. Industrie 4.0: The Big Unknown?

[50] S. Lohr, "The Age of Big Data", p. 11. New York Times, 2012.

[51] S. Luthra, K. Mangla, R. Shankar, C. Prakash Garg, S. Jakhar, "Modelling critical success factors for sustainability initiatives in supply chains in Indian context using Grey-DEMATEL". Production Planning & Control, **29**(9), 705–728, 2018.

[52] S. Patil, R. Kant, "Predicting the success of knowledge management adoption in supply chain using fuzzy DEMATEL and FMCDM approach". International Journal of Business performance and Supply Chain Modelling **6** (1), 75-93, 2014.

[53] T. Dawson,. "Industry 4.0 – Opportunities and challenges for smart manufacturing." 2014. http://blog.ihs.com/q13-industry-40-opportunities-and-challenges-for-smart-manufacturing.

[54] W. Yunna., H. Yong., L. Xiaoshan L. Lingwenying, K. Yiming K., "Identifying and analyzing barriers to offshore wind power development in China using the grey decision-making trial and evaluation laboratory approach". Journal of Cleaner Production **189**; 853-863, 2018.

[55] Worldometer. 2020. Available: https://www.worldometers.info/coronavirus/.

[56] X. Xia, K. Govindan, Q. Zhu, "Analyzing internal barriers for automotive parts remanufacturers in China using grey-DEMATEL approach". Journal of Cleaner Production, **87**, 811–825, 2015.

[57] Y. Chou,, C. Sun, H. Yen, "Evaluating the criteria for human resource for science and technology (HRST) based on an integrated fuzzy AHP and fuzzy DEMATEL approach". Appl. Soft Comput. **12** (1), 64–71, 2012.

[58] Y. Wu, Y. Hu, X. Lin, L. Li, Y. Ke, "Identifying and analyzing barriers to offshore wind power development in China using the gray decision-making trial and evaluation laboratory approach". Journal of Cleaner Production, **189**, 853-863, 2018.

A S T E S

Enhancing Decision Trees for Data Stream Mining

Mostafa Yacoub[1,*], Amira Rezk[1], Mohamed Senousy[2]

[1]*Faculty of Computers and Information, Information System Department, Mansoura University, Mansoura, 35511, Egypt*

[2]*Faculty of Management Sciences, Computer and Information system Departments, Sadat Academy for Management Sciences, Cairo, 00202, Egypt*

A R T I C L E I N F O	A B S T R A C T
	Data stream gained obvious attention by research for years. Mining this type of data generates special challenges because of their unusual nature. Data streams flows are continuous, infinite and with unbounded size. Because of its accuracy, decision tree is one of the most common methods in classifying data streams. The aim of classification is to find a set of models that can be used to differentiate and label different classes of objects. The discovered models are used to predict the class membership of objects in a data set. Although many efforts were done to classify the stream data using decision trees, it still needs a special attention to enhance its performance, especially regarding time which is an important factor for data streams. This fast type of data requires the shortest possible processing time. This paper presents VFDT-S1.0 as an extension of VFDT (Very Fast Decision Trees). Bagging and sampling techniques are used for enhancing the algorithm time and maintaining accuracy. The experimental result proves that the proposed modification reduces time of the classification by more than 20% in more than one dataset. Effect on accuracy was less than 1% in some datasets. Time results proved the suitability of the algorithm for handling fast stream mining.

1. Introduction

Recently, information played a major role in our world. Subsequently, the process of extracting knowledge is becoming very important. New applications that depend on data streams became more popular with time. Stream data are clear in sensors, telephone call records, click streams, social media, and stock market.

Contrary to traditional data mining, which analyses a stored data set, the stream mining analyses a data stream which cannot be saved as it's infinite and needs expensive storage capabilities. Data streams arrive continuously and with fast pace, this prevents multiple passes of the data. So, processing time is more constrained in data streams.

Classification is a mining technique used to build a classification model based on the training data set which used to predict the class label of a new undefined data. Decision trees, neural networks, Bayesian networks, and Support Vector machines (SVM) are considered the most effective methods of classification. Decision trees are data structures organized hierarchically by splitting input space into local zones to predict the dependent variable.

Decision trees are hierarchical data structures for supervised learning by which the input space is split into local regions to predict the dependent variable [1]. It is classified as greedy algorithms which try to find a decision at each step of small steps. Decision trees consist of nodes and edges (branches). Root node has no incoming edge. Leaves or terminal nodes have no outgoing edges. All other nodes – besides root – have exactly one ingoing edge. Internal or test nodes are the nodes with outgoing edges. Each internal node splits the instance space into two or more instance sub-space. These splits are done according to a specific splitting discrete function of attribute values (inputs). Classes are assigned to leaf nodes.

Decision trees are characterized by simplicity, understandability, flexibility, adaptability and higher accuracy [2], [3]. The ability to handle both categorical and continuous data is an important advantage of decision trees. So, there is no need to normalize the data before running the decision tree model, that means fewer preprocessing processes. Being easier to construct and understand is another important factor for preferring decision

trees over other data mining techniques. In addition, decision trees are interpretable as it can be expressed as a logical expression. Missing values in data are considered issues need to be handled before running data mining techniques in order not to affect the results. Decision trees can handle data with missing values successfully.

Traditional decision tree learners like ID3 (Iterative Dichotomiser 3) and C4.5 (Classification 4.5) have problems in handling data streams. It presumes that the whole training examples can be stored concurrently in main memory, which is not valid in data streams [4].

Very Fast Decision Trees (VFDT) was introduced by Domingos and Hulten in 2000[5]. VFDT uses the Hoeffding bound for node splitting and creating Hoeffding trees. The basis of Hoeffding trees is "a small sample can often be enough to choose an optimal splitting attribute". Hoeffding bound gives a mathematical support to that basis quantifying the number of examples needed to estimate some statistics within a prescribed accuracy [6].

According to Hoeffding bounds, with probability $1 - \delta$, the true mean of r is at least $\bar{r} - \varepsilon$, where

$$\varepsilon = \sqrt{\frac{R^2 \ln(\frac{1}{\delta})}{2n}} \quad (1)$$

In equation (1), r represents continuous random variables whose range is R. $\bar{r}$ is the observed mean of the samples after n independent observations. [7]. The VFDT defines the two attributes t_1, t_2 with highest information gain G_{t1} and G_{t2}. If $\triangle G = G_{t1} - G_{t2}$ is higher than ε (equation 1), then G_{t1} is the best split attribute with probability of $1 - \delta$ and the split is done. (Algorithm 1: VFDT)

In VFDT, leaves are replaced with decision nodes recursively. Statistics about attributes values are saved in each leaf. Based on these statistics, a heuristic function calculates the value of split tests. Each new instance passes from root to a leaf. At each leaf, attribute evaluation is done and follow the branch according to evaluation result. An important step must be done, which is updating the enough statistics of the leaf [8].

VFDT can address the research issues of data streams such as ties of attributes, bounded memory, efficiency and accuracy[9]. VFDT is known for having decent memory management. It can save memory by deactivating less promising leaves when memory reaches a limit then it turns back to normal when memory is free[10]. Also, it monitors the available memory and prunes leaves (where sufficient statistics are stored) depending on recent accuracy [11], [12].

The rest of this paper will discuss the related work in section two, the proposed modification on VFDT in section three, the evaluation of the proposed modification in section four and finally the conclusion and future work in section five.

Algorithm 1: VFDT

Result: very fast decision tree
begin
 Let T be a tree with a one leaf (the root)
 for all training examples **do**

  Update sufficient statistics in l
  Increment n_1, the number of examples seen at l
  if n_1 mod n_{min} = 0 **and** all examples seen at l not all same class **then**
   Compute $\bar{G}_l(X_i)$ for each attribute
   Let X_a be the attribute with highest $\bar{G}_l$
   Let X_b be the attribute with the second highest $\bar{G}_l$
   Compute Hoeffding bound $\varepsilon = \sqrt{\frac{R^2 \ln(\frac{1}{\delta})}{2nl}}$
   if $X_a \neq X_\Phi$ **and** $(\bar{G}_l(X_a) - \bar{G}_l(X_b) > \varepsilon$ **or** $\varepsilon < \tau)$ **then**
    Replace l with an internal node that splits on X_a
    for all branches of the split **do**
     Add new leaf with initialized sufficient statistics
    end
   end
  end
 end
end

---

2. Related Work

Although decision trees have more than accepted results in data stream mining, there have been many trials of modification to enhance results. For being one of the noticeable algorithms in decision trees, VFDT has share in these studies. Following studies present VFDT modifications to achieve higher accuracy, less time, or both. Next section summarizes these studies followed by a table to show impact on time and accuracy.

2.1. Bagging

In [13], the author proposed VFDTc and VFDTcNB, which can include and classify new data online with a one scan of the data for medium and large size datasets. VFDTc can deal with numerical attributes heterogeneous data, while VFDTcNB can apply naive Bayes classifiers in tree leaves and reinforces the anytime characteristic. In [14], the authors presented GVFDT, an employment of the VFDT used for creating random forests that use VFDTs for GPUs data streams. This technique takes advantage of the huge parallel architecture of GPUs. Furthermore, GVFDT algorithm reduces the communication between CPU and GPU by constructing the trees inside the GPU.

2.2. Adaptability

In [15], the authors proposed Strict VFDT in two versions; SVFDT-I and SVFDT-II. Both are seeking reducing tree growth and decreasing memory usage. Both algorithms produce trees much smaller than those produced by the original VFDT algorithm. Testing them on eleven datasets, SVFDT-II produced better accuracy than the SVFDT-I, together with significantly reducing tree size.

In [16], the authors presented ODR-ioVFDT (Outlier Detection incremental optimized VFDT) as an extension of VFDT to handle outliers in continuous data learning. The new algorithm was applied onto bioinformatics data streams–loaded by sliding windows – to diagnose and treat disease more efficiently. The ODR model chooses the outlier, which is stored into misclassified database. Clean data will be passed through ioVFDT classifier for decision tree building. The lower performance will send response to outlier and classifier model, the model update will be needed. In

[17], the authors proposed an optimization of VFDT algorithm to decrease the effect of concept drift by utilizing sliding windows and fuzzy technology. Results showed improvements in accuracy results.

Table 1: Summary of related work

Title	Year	Algorithm Name	Algorithm Idea	Time Results	Accuracy Results
Speeding up Very Fast Decision Tree with Low Computational Cost	2020	IMAC (Incremental Measure Algorithm Based on Candidate Attributes)	The algorithm calculates the heuristic measure of an attribute with lower computational cost. Possible split timing is found by selecting subset of attributes precisely.	Decreased in most datasets except two with minor increase	No loss in some datasets and minor loss of accuracy in few datasets
A VFDT algorithm optimization and application thereof in data stream classification	2020	Optimized VFDT	an optimization of VFDT algorithm to decrease the effect of concept drift by utilizing sliding windows and fuzzy technology	Lower Time	Higher Accuracy
Enhancing Very Fast Decision Trees with Local Split-Time Predictions	2018	OSM (One-sided minimum)	replaced the global splitting scheme with local statistics to predict the split time which leads to lower computational cost by avoiding excessive split tries.	Decreased run-time	Same accuracy
Strict Very Fast Decision Tree: a memory conservative algorithm for data stream mining Victor	2018	Strict VFDT: SVFDT-I & SVFDT-II	Both are seeking reducing tree growth and decreasing memory usage. Both algorithms produce trees much smaller than those produced by the original VFDT algorithm.	Decreased in 3 datasets, and higher in the other 8 datasets	Decreased in 5 datasets, same accuracy in 3 datasets, and higher accuracy in 3 more
Robust High-dimensional Bioinformatics Data Streams Mining by ODR-ioVFDT	2017	ODR-ioVFDT	The ODR model chooses the outlier, which is stored into misclassified database. Clean data will be passed through ioVFDT classifier for decision tree building. The lower performance will send response to outlier and classifier model, the model update will be needed.	Higher in all datasets	Higher in all datasets with small percentage
Random Forests of Very Fast Decision Trees on GPU for Mining Evolving Big Data Streams	2014	GVFDT: Very Fast Decision Trees for GPU	This technique takes advantage of the huge parallel architecture of GPUs. Furthermore, GVFDT algorithm reduces the communication between CPU and GPU by constructing the trees inside the GPU.	Lower time in the three datasets	Lower Accuracy in two datasets and same accuracy in one.
Accurate Decision Trees for Mining High-speed Data Streams	2003	VFDTc & VFDTcNB	VFDTc: can deal with numerical attributes. VFDTcNB: apply naive Bayes classifiers in tree leaves	Decrease with more than 50%	Increase by 2% (average)

2.3. Split Function

In [18], the authors replaced the global splitting scheme with local statistics to predict the split time which leads to lower computational cost by avoiding excessive split tries. Results showed decreased run-time with no loss in accuracy. In [19], the authors introduced IMAC (Incremental Measure Algorithm Based on Candidate Attributes) an online incremental algorithm with a much lower computational cost. The algorithm calculates the heuristic measure of an attribute with lower computational cost. Possible split timing is found by selecting subset of attributes precisely. The algorithm showed faster and more accurate results by decreasing split attempts with much lower split delay.

Table 1 summarizes efforts in this area, but the time still a challenge that face the algorithms that applied to the stream data. All mentioned studies achieved better time results except on

research. From accuracy side, only three studies achieved higher accuracy and another two achieved less accuracy. So, this paper will try to propose a modification to reduce the time of the decision tree in stream data.

3. The proposed VFDT-S1.0

The proposed VFDT-S1.0 aims to modify the original VFDT algorithm to reduce the time of classification. The idea of the modification is based on two main factors. First is bagging more than one algorithm to improve performance and second factor is using random sampling with fixed percentage from the whole data.

Algorithm 2: VFDT-S1.0

Result: M: Model with the highest accuracy
begin
Load Data Stream S
For every record in S:
Delete record if contains null value
Let $S_{train} = S * 0.8$
$S_{test} = S - S_{train}$
S_{train} = SimpleRandomSample(S_{train})
HT=HoeffdingTree(S_{train})
HT_{Pred}=Predict(HT,S_{test})
HT_{Acc}=mean(HT_{Pred} , S_{test}Class)*100
HOT=HoeffdingOptionTree(S_{train})
HOT_{Pred}=Predict(HOT,S_{test})
HOT_{Acc}=mean(HOT_{Pred} , S_{test}Class)*100
HAT=HoeffdingAdaptiveTree(S_{train})
if (HTAcc > HOTAcc **and** HTAcc > HATAcc) **then**
   | M = HT
else
   if ($HOT_{Acc} > HT_{Acc}$ & $HOT_{Acc} > HAT_{Acc}$) **then**
     | M = HOT
   else
     if ($HAT_{Acc} > HT_{Acc}$ & $HAT_{Acc} > HOT_{Acc}$) **then**
     M = HAT
     end
   end
end
end

The three algorithms are run sequentially to find the one with more accurate results. Accuracy is measured for the three models generated by the three algorithms. The algorithm with highest accuracy is used on the rest of data.
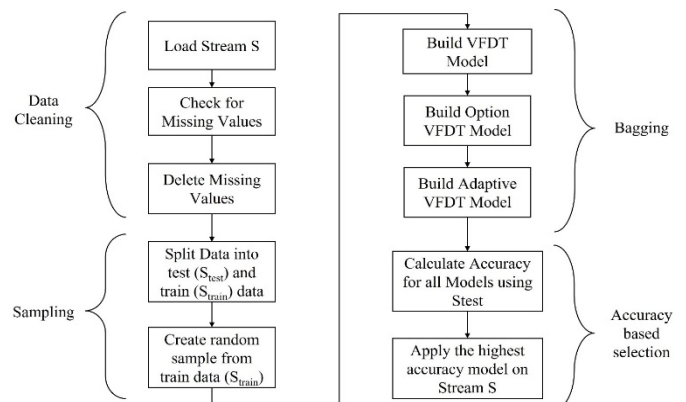

Figure 1: VFDT-S1.0 Framework

Sampling is used to compensate using three different algorithms sequentially. Using sampling in data streams has been discussed in many studies. Three sampling techniques related to data streams are reservoir sampling, AMS-sampling, and Sliding window sampling. In [20], random sampling was used to challenge time constraint. As shown in figure 1, the three algorithms were trained using the same sample. As we choose the best accuracy of the three to use and compare with original VFDT algorithm. Figure 1 displays VFDT-S01 framework, explaining the four basic stages of it.

4. Implementation and Evaluation

To examine the proposed algorithm, it is tested and compared to the original VFDT algorithm. Coding and evaluation were done using Java and R languages working on Microsoft Windows 10 environment on core i5-5200U processor machine. Source code of algorithms is written in Java in Massive Online Analytics (MOA) tool, employing MOA codes in R is done by using RMOA package. RMOA is connecting R with MOA to build classification and regression models on streaming data.

The test is done using 7 different real classification datasets; covType[21], Airlines[22], KDD99[23], Elecnorm[24], MplsStops[25], Chess[26], and Income[27]. Table 2 summarizes the seven datasets and comparing them according to number of instances, attributes, and classes.

Table 2: Sample Table

Dataset	Number of Instances	Number of attributes	Number of Classes
covType	581,012	55	7
Airlines	539,383	8	2
KDD99	494,020	42	23
Electricity	45,312	9	2
MplsStops	51,920	15	2
Chess	28,056	7	18
Income	48,842	15	2

Each dataset was divided into training and test set. Training set is 80% the whole data and the reminder was the test set for prediction. Both algorithms were tested using the same test set to get more accurate comparison results. Accuracy was calculated as number of true predictions divided by test set size.

Time was calculated by using built-in time function in R at the start and end of code. Both accuracy and time were calculated as an average of three runs of both algorithms on every dataset.

Table 3 compares the proposed VFDT-S1.0 and VFDT based on the accuracy and time. Also shows that the original algorithm achieves higher accuracy in all seven datasets.

Table 3: Algorithms Comparison

	VFDT		VFDT-S1.0		Difference Percentage	
Data set	Accuracy%	Time (sec)	Accuracy %	Time (sec)	Accuracy %	Time %
CovType	72.86 %	816.00	69.95 %	620.74	-4.00 %	-23.93 %
Airline	65.06 %	635.10	60.93 %	539.86	-6.34 %	-15.00 %
KDD99	99.79 %	638.39	99.55 %	492.65	-0.24 %	-22.83 %
Elec.	77.11 %	52.70	76.38 %	45.11	-0.94 %	-14.40 %
MplsStop	79.53 %	20.12	77.91 %	18.60	-2.04 %	-7.54 %
Chess	33.70 %	29.71	32.29 %	27.06	-4.18 %	-8.92 %
Income	83.94 %	53.18	81.92 %	46.51	-2.40 %	-12.53 %

Differences between VFDT accuracy and VFDT-S1.0 accuracy varies from 0.24% at KDD99 dataset to 4.13% at Airline dataset. Figure 2 displays the accuracy between the two algorithms.
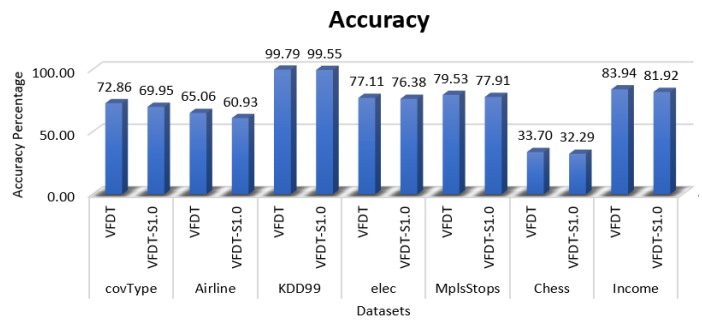

Figure 2: Accuracy Comparison on all datasets

Figure 3: Time Comparison on datasets (covType, Airline and KDD99)
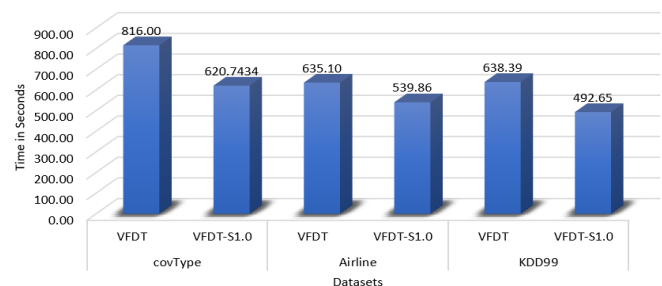

Figure 4: Time Comparison on datasets (elec, MplsStops, Chess, and Income)

Figure 3 represents processing time of both algorithms on largest three datasets and figure 4 displays time on the reminder datasets. Time was always better with VFDT-S1.0 at all datasets. 1.52 seconds was the minimum difference between two algorithms on MplsStops dataset. CovType dataset had the major difference with 195.26 seconds. At KDD99 dataset, which had the highest accuracy difference, the time was less by 145.74 seconds.

5. Conclusion

This paper proposed the VFDT-S1.0; a modified VFDT algorithm that uses bagging techniques to achieve most possible accuracy. In time factor, we used random sampling to achieve better processing time. We tested the new algorithm using seven real classification datasets and compared results with VFDT algorithm. Improvements have been noticed in time as VFDT-S1.0 took much less time with all datasets. Biggest time difference was 24% in CovType dataset. In KDD dataset the time dropped by 23% with -0.2% in accuracy. This time difference shows potential for scaling VFDT. As it can be processed by much lower processing resources. Also, the ability to handle very fast data streams with dependable accuracy.

6. Future Work

In future work, tree size, Kappa, sensitivity, and specificity will be measured for both algorithms. Accuracy can be enhanced with bagging more models and choosing a sample with the same class representation in dataset. Also, parallel processing is considered for much time improvement. Change detection techniques are going to be added to deal with concept drifts.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] E. Alpaydın, "Introduction to machine learning," Methods in Molecular Biology, **1107**, 105–128, 2014, doi:10.1007/978-1-62703-748-8-7.

[2] Z. Çetinkaya, F. Horasan, "Decision Trees in Large Data Sets," International Journal of Engineering Research and Development, **13**(1), 140–151, 2021, doi:10.29137.

[3] S. Moral-garcía, J.G. Castellano, C.J. Mantas, A. Montella, J. Abellán, "Decision Tree Ensemble Method for Analyzing Traffic Accidents of Novice Drivers in Urban Areas," 1–15, 2019, doi:10.3390/e21040360.

[4] F.M.J.M. Shamrat, R. Ranjan, A. Yadav, A.H. Siddique, S. Engineering, C. Neusoft, C.C. Officer, "Performance Evaluation among ID3 , C4 . 5 , and CART Decision Tree Algorithms," International Conference on Pervasive Computing and Social Networking, 2021.

[5] P. Domingos, G. Hulten, "Mining high-speed data streams," Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '00, 71–80, 2000, doi:10.1145/347090.347107.

[6] M. Yacoub, A. Rezk, M. Senousy, "Adaptive classification in data stream mining," Journal of Theoretical and Applied Information Technology, **98**(13), 2637–2645, 2020.

[7] W. Zang, P. Zhang, C. Zhou, L. Guo, "Comparative study between incremental and ensemble learning on data streams: Case study," Journal of Big Data, **1**(1), 1–16, 2014, doi:10.1186/2196-1115-1-5.

[8] J. Gama, P.P. Rodrigues, An Overview on Mining Data Streams, Springer-Verlag Berlin Heidelberg: 38–54, 2009, doi:10.1007/978-3-642-01091-0.

[9] C.C. Aggarwal, Data streams: Models and Algorithms, 1st ed., Springer-Verlag US, 2010, doi:10.1007/978-0-387-47534-9.

[10] E. Ikonomovska, J. Gama, S. Džeroski, "Learning model trees from evolving data streams," Data Mining and Knowledge Discovery, **23**(1), 128–168, 2011, doi:10.1007/s10618-010-0201-y.

[11] A. Muallem, S. Shetty, J.W. Pan, J. Zhao, B. Biswal, "Hoeffding Tree Algorithms for Anomaly Detection in Streaming Datasets : A Survey,"

Journal of Information Security, **8**(4), 339–361, 2017, doi:10.4236/jis.2017.84022.

[12] D.H. Han, X. Zhang, G.R. Wang, "Classifying Uncertain and Evolving Data Streams with Distributed Extreme Learning Machine," Journal of Computer Science and Technology, **30**(4), 874–887, 2015, doi:10.1007/s11390-015-1566-6.

[13] J. Gama, R. Rocha, P. Medas, "Accurate Decision Trees for Mining High-speed Data Streams," Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 523–528, 2003, doi:10.1145/956750.956813.

[14] D. Marron, A. Bifet, G. De Francisci Morales, "Random forests of very fast decision trees on GPU for mining evolving big data streams," Frontiers in Artificial Intelligence and Applications, **263**, 615–620, 2014, doi:10.3233/978-1-61499-419-0-615.

[15] V. Guilherme, A. Carvalho, S. Barbon, "Strict Very Fast Decision Tree : a memory conservative algorithm for data stream mining," Pattern Recognition Letters, 1–7, 2018.

[16] D. Wang, S. Fong, R.K. Wong, S. Mohammed, J. Fiaidhi, K.K.L. Wong, "Robust high-dimensional bioinformatics data streams mining by ODR-ioVFDT," Scientific Reports, **7**, 1–12, 2017, doi:10.1038/srep43167.

[17] S. Jia, "A VFDT algorithm optimization and application thereof in data stream classification A VFDT algorithm optimization and application thereof in data stream classification," Journal of Physics: Conference Series, 1–7, 2020, doi:10.1088/1742-6596/1629/1/012027.

[18] V. Losing, H. Wersing, B. Hammer, "Enhancing Very Fast Decision Trees with Local Split-Time Predictions," IEEE International Conference on Data Mining (ICDM), 287–296, 2018, doi:10.1109/ICDM.2018.00044.

[19] J. Sun, H. Jia, B. Hu, X. Huang, H. Zhang, H. Wan, X. Zhao, "Speeding up Very Fast Decision Tree with Low Computational Cost," International Joint Conferences on Artificial Intelligence, 1272–1278, 2020.

[20] E. Ikonomovska, M. Zelke, Algorithmic Techniques for Processing Data Streams, 2013.

[21] J A Blackard D J Dean, "Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables," Computers and Electronics in Agriculture, **24**, 131--151, 1999.

[22] E. Ikonomovska, Airline, 2009.

[23] S.D. Hettich, S. and Bay, The UCI KDD Archive, 1999.

[24] M. Harries, Electricity, Aug. 2019.

[25] M. GIS, Police Stop Data, 2017.

[26] M. J, Chess Game Dataset, 2017.

[27] W. Liu, Adult income dataset, 2016.

The Effect of Obstacle Design Architectures on Randomly Ranging AGVs in a Shared Workspace

Indravash Chowdhury, Ravinder Singh, Christopher Kluse, Mohammad Mayyas[*]

Bowling Green State University, Engineering Technologies, Bowling Green, 43043, USA

A R T I C L E   I N F O

A B S T R A C T

As mobile robots or Automated Guided Vehicles age, and their navigation components wear down, downtime may increase, and therefore robots may lose their path planning capabilities. In such blind-navigation scenarios, we assume robot(s) is programmed to undergo a default random-walk locomotion in a two-dimensional confined workspace with spatially populated obstacles. The robot(s) has only one capability that is to identify its destination. Extensive simulations are conducted to evaluate the cost of sending such robots, measured in number of steps, to a safe "home" position. Specifically, the contribution of this paper is to provide guidelines on how such blind-navigation mission is accomplished under two conditions: (1) Robot(s) locomoting within static obstacles configured in desired patterns, and (2) Robot(s) locomoting within randomly and freely ranging objects (robot or human). The research compares the efficiencies of randomly walking objects to workspace size, the number of obstacles and their dynamics, which provide engineers an understanding of how obstacle design may reduce the risk of collision.

1. Motivation

According to the Occupational Safety and Health Administration (OSHA), industry related injuries indicate that from 2015 to 2019, 76% of injuries (4 injuries total) were severe and 3 fatalities occurred throughout the USA [1]. These injuries are attributed to autonomous cars, mobile cars, and AGVs. OSHA states that there were 30 casualties in human-robot workspaces from 1987 to 2019. Due to a significant increase in robot installation worldwide, and the increasing amount of interaction, it has become important to investigate the robot(s) locomotion within workspace, examine the probability of accidents, and determine a solution to minimize unsafe interaction. Automated Guided Vehicle systems (AGVs) provide manufacturing and logistic facilities with added flexibility thereby increasing the overall rate of production. These AGVs are outfitted with appropriate safety equipment and sensors to make them safe while in operation [2]. The main advantage of AGVs are that "*if properly guided*", they do not injure people and/or damage infrastructures. It is generally accepted that AGVs are safer than manual equipment such as conventional forklifts. The introduction of collaborative AGVs or robots has initiated the concept of fencing, and currently, in this advanced technology era, the combined work effort between humans and robots has become efficient and productive. However, the coexistence of autonomous mobile robots within themselves and

most often in the presence of humans have made safety a priority in a shared workspace. This study is concerned with a worst-case scenario when a robot loses its path planning capability due to a communication or processing malfunction. The locomotion type discussed in this research assumes random walk for both robot-robot and human-robot interaction. This paper provides literature review on automatic guidance and safety method. The concept of Random Walk is then presented to application of robot locomotion, with a goal to understand efficiency to navigate in absence of a real-time localization system. Extensive computational simulations are performed for several robot interaction with structured and random obstacle in confined area, and with objective to study navigation tradeoff under random walk.

2. Automatic Guidance

AGVs can be used in external and internal environments such as manufacturing, distribution, transhipment and (external) transportation areas [3]. AGVs also play an important role in the transportation of materials that are associated with manufacturing processes. About 20,000 AGVs are utilized for industrial applications [4]. The use of the automated guided vehicle has increased significantly since their introduction. According to [5], the human-machine interfaces facilitate the users to perform coordinated logistical activities within industrial environments. For example, PMD-cameras (Photonic

*Corresponding Author: Mohammad Mayyas, Email: mmayyas@bgsu.edu

Mixer Device camera) are typically used in sensing the obstacles within the path of an AGV. The PMD-camera acts as an interactive technique that records the video images of various objects and scenes in both 2-D and 3-D greyscales and deep-data images, respectively. In addition, the same technique is used when there is a need to monitor the area where an industry robot is working [6]. According to [7] **,** human-robot interface systems design considers the formation of a multi-agent framework, thus taking care of safety and possible conflict. Such interfaces can mimic social behaviors, uncertainties, and norms. Regardless of the constraints and shortcomings, Sabattini and fellow researchers proposes that AVG systems can be improved for factory logistics [8]. They agree on the means through which more of the human activities can be interpreted through artificial machine intelligence. The main goal is to act more like humans, with advanced efficiencies, and minimize human errors.

It is important to provide insight into the research regarding AGVs routing via transhipment and transportation along with the scheduling of various material handling equipment in large AGV systems since research in these areas are limited. The integration of routing and scheduling is a challenging problem. Therefore, it is important to conduct a study in which these characteristics should be addressed. The advantage of multi-objective problems in scheduling plays a significant role in scheduling the operations and production planning in a flexible manufacturing system. For example, *Heuristic based approaches* include evolutionary algorithms adopted to address the multi-objective scheduling problems in a flexible industrial environment by analysing the vehicle and machine scheduling features in FMS, moreover, it addresses the mutual problem for the minimization of the mean flow time as well as for the mean tardiness objectives [9]. In contrast, differential evolution has been used as one of the most influential tools and [10] depicts the synchronised scheduling for material handling systems. To minimize the cycle time, a combination of vehicle assignment heuristics and machine selection heuristics are very useful. In *Artificial Intelligent based approaches* [11], the authors explain that the total cost can be minimized by generating the optimum schedule including simulated annealing (SA), genetic algorithm (GA), swarm algorithm (PSA), and memetic algorithm (MA). Additionally, these mechanisms help to overcome the idle time. While solving the optimization problem, the objective is often achieved by involving wide-ranging computational resources. *Heuristic Approaches* are employed to overcome the routing problems by minimizing transportation time [12].

Countless inventions have been developed to accomplish the "guided" portion of AGV operation. One such patented invention enables an autonomous vehicle to adjust its pre-programmed path in order to avoid obstacles [13]. This type of system works with several different types of guided vehicles from single cargo units to ones that tow a train of connected carts. Typically, a system is comprised of a computer that receives an input signal from a detection sensor such as a laser scanner. The computer, rather than stopping the vehicle when an obstacle is detected, calculates a safe offset from its normal path and navigates the vehicle around the obstacle according to this temporary safe path. This type of collision avoidance is preferable because the vehicle navigates safely without having to stop.

Avoidance between multiple AVGs was developed in [14]. This method is executed using several steps. First, the vehicle's computer tracks the position of all other vehicles in its detection area at regular intervals. Next, the computer calculates the trajectory of each surrounding vehicle from the collected velocity and directional data and predicts where the vehicle and the other vehicles will collide. The computer then sends a warning signal to the other vehicles and maneuvers itself to avoid the predicted collision. This type of system is useful because it allows an AGV to avoid dynamic obstacles that are constantly moving.

Another way to avoid dynamic collisions is to allow each AGV to know where each other are located before it is assigned a path. This is achieved by a host computer that sends signals to all AGVs in the operating area that tells the AGV whether to move along a path or avoid the path [15]. The computer assigns a path for the subsequent vehicle, checks if there are other vehicles moving on that path, and sends the next vehicle along that path if no other vehicles are in the way. This method is proactive since it assigns clear paths before the vehicle starts to move. While steering clear of obstacles is important, the vehicle must be able to follow a pre-determined path. Kondo has invented a system that offers improved controls for the path of an automated guided vehicle [16]. This system is composed of several magnetic sensors underneath the vehicle that detects magnetic tape on the floor within the operating area. The main sensor is positioned in the middle of the vehicle, and several small sensors are positioned to the left and right of the main sensor. The side sensors allow the vehicle to better detect and adjust accordingly when it deviates left or right from the guide tape. Accurate movement and paths are essential for AGVs to work properly and efficiently; thus, having more sensors provides the ability to calculate small movements and deviations.

Another method to determine a vehicle's location was developed in [17]. Unlike the previous method that was guided by the floor, this method controls vehicle navigation by using the building's ceiling. This is beneficial when light conditions vary across the vehicle's operating area. The vehicles locate themselves by taking an image of the ceiling that are covered with numerous unique markers. The markers have areas that are infrared reflective and areas that are non-reflective. Based on the shape of the markers and the orientation of the reflective and non-reflective sections, the vehicle's computer calculates its position and orientation in the workspace. This method has several benefits. As mentioned previously, this system works in difficult lighting conditions because the ceiling markers reflect unique patterns of light sent by the vehicle. This system also allows for a vehicle to know its orientation in the space as well as its current location.

3. Collision Avoidance and Safety

To ensure a safe human-robot or robot-robotic environment before installing a mobile robot, a safety assessment of the workspace is considered a best practice. There has been a significant amount of work conducted relative to safety in shared interaction. In [18], the authors surveyed various methods that have been used for the same human-robot interaction. The survey covered areas by addressing safety through control, robotic

motion planning, prediction, and psychological factors. The survey also includes pre-collision and post collision methods that have been analyzed by predicting natural human action and motion. Other methods include operator safety on an assembly line for collaborative robots [19], hazard operability analysis to identify the hazards, risk assessment in a human-robot collaborative environment [20], and human-robot interaction guided by operator gesture [21]. In [22], the authors presented an algorithm for autonomous path planning and motion coordination for multiple AGV's to determine the shortest feasible path by using reliable detection and prioritized conflict resolution in the shared workspace.

In [23], the authors presented the essentials for the safe operation of AGVs. Part of this was an analysis on the merit of what was then, modern optic and ultrasonic technologies for sensing. Today, many of the problems pinpointed by Miller & Subrin have been addressed however, they still warrant consideration. The primary concerns for both technologies are false positives and inaccuracy when used on curved paths. With modern lenses and filtering technologies, these problems are minimized using optic technology. Safety is an important and complex issue, and there are many ways to solve safety problems. Some researchers have advocated for the use of a full system to represent a reality-based process for solving these problems, beginning with discovery, and commencing with continuous improvement. According to [24], there are four "metaprinciples"; (a) inventorize, (b) capacitate, (c) prioritize, and (d) integrate. *Inventorize*, in this case, means to catalog anything that is considered a safety issue. This includes the nature and severity of systems either planned or existing. *Capacitate* entails determining how certain features of the system (as planned/built) can be used to help minimize those issues, as well as improving the efficacy of those features when used in this way. *Prioritize* simply means that the most important improvements are given top priority. *Integrating* extends the envelope of responsibility beyond the product itself, through education of the workforce and management who are involved in the use of the product. Extensive research was performed in *Development of a Methodology for the Evaluation of Active Safety using the Example of Preventative Pedestrian Protection* [25]. Using millions of simulations, Helmer provided a great deal of data on the efficacy of three basic systems used in preventative safety measures. These systems are: (a) warnings, (b) brake assist, and (c) automatic braking. The research examined the main factors that lead to the results of each of the three systems.

To comply with safety rules, AGVs must include some safety sensors and devices to prevent risks. AGVs should contain sensors in the direction of travel covering the maximum moving width and length provided to prevent contact between the load and any possible obstacle. In Industry 4.0, robot or human risk mitigation through various safety measures has become a paramount concern. Laser scanners have been recently used in obstacle recognition and personnel safety. Other sensing methods include cameras, radar, and ultrasound sensors where the latter has been favored since the late 80's [26]. The main AGV active safety devices and AGV safety sensors are collision avoidance system, contact bumpers emergency, stop buttons, and safety PLC's. The collision avoidance system on the AGV can apply several different laser sensors that are set up on the front,

rear, side, and upper locations of the AGV. When the AGV is moving on a guided path, this system will detect an obstacle (such as a human) in any of the coverage locations. When the obstacle is within the warning field of any of these sensors, the AGV will decelerate to a slower speed in anticipation of a full stop. If the obstacle is still detected within the protective field of the sensor, the AGV will apply its brake to initiate a complete stop before contact is made. The AGV will resume automatic operation approximately three seconds after the obstacle is removed from the protective field [27].

The safety area must be designed depending on many factors, which include the surrounding area, vehicle speed, payload, and floor conditions. Every point of the path followed by an AGV must have its own safety area to ensure that the time and distance to stop the vehicle is sufficient to avoid contact with obstacles. According to recent information about AGV accidents demonstrates that even with onboard sensors these vehicles did not detect nearby workers [1]. Additional complications involving operator safety will occur as AGVs are employed in open environments or when robot arms are added to AGV platforms. These safety risks can be alleviated with the use of new technologies, such as three-dimensional sensors and algorithms that provide industrial vehicles with information about potential collisions in their paths. However, the capabilities of new sensors need to be fully explored and their strengths and weaknesses understood before they can be incorporated onto AGVs that are deployed on the production floor.

Researchers have developed various algorithms to improve safety of robotic movement through motion planning techniques and sensing function. In [28], the authors proposed an innovative method for extrinsic calibration of sensors to track arbitrary obstacles around the robot. In [29], the authors utilized KinectV2 depth cameras to track moving obstacles within workspace. Alternatively, optimizing the layout of obstacles in workspace can be essential for ensuring worker safety. The authors presented a geometrical analysis of a free workspace for a collaborative robot. The method uses an octree model of the free workspace for geometrical analysis to characterize the robots and obstacles using Constructive Solid Geometry (CSG) and Computer Aided Design (CAD) [30]. In [31], the authors introduced task specific interaction for both robots and human to determine suitable location for human robot interaction tasks. In [32], the authors proposed a risk-based framework for human-robot interface through safety control systems which includes proximity detection, collision avoidance systems and docking & compliance control.

Safety measures with probabilistic approach are imminent to avoid these collisions. There are ranges of approaches to calculate the probability of collisions in human-robot distributed workspace. The analyses of the probability of risk due to intrusion uses several factors such as the safety function of the robot, human and robots, relative velocity, and position [33]. Lambert, Gruyer and Saint Pierre [34] presented a probabilistic, Monte Carlo algorithm to reduce risk from Gaussian integrals with vehicle volume and obstacle volume as considerations [34]. J. van den Berg, P. Abbee, and K. Goldberg [35] proposed an approach which decreases the probability of collisions based on

linear quadratic Gaussian-motion planning. The algorithm uses rapid a random trees model to minimize the collision by sensing the uncertainty. Applying sampling-based algorithms such as the Monte Carlo sampling strategy model, collision probabilities are estimated by extrapolating the ratio of the number of collision free of simulated executions [36].

Several collision avoidance approaches have been proposed in the field of robotics research. In [37], the authors developed a human-robot collision avoidance approach through a potential field method where the repulsive command for the robot is generated to avoid obstacles. The distance between the robot and the moving obstacle is calculated and the repulsive vector is used to control the generic motion of the robot. Path finding methods such as artificial potential field, particle swarm optimization, and ant colony optimization are also used to foster safe mobile robot path planning. The authors used the artificial bee colony algorithm on motion planning of multiple mobile robots in a dynamic environment [38]. In [39], the authors indicated that most robots in manufacturing plants follow the sense and avoid approach with a camera mounted on a robot which computes the distance if an obstacle comes within a minimum distance, then the sense and avoid algorithm is applied to stop the robot.

Researchers have also studied safety measures through collision detection algorithms. For example, multi robot collision avoidance based on a velocity obstacle paradigm is used in a dynamic workspace to reduce collision probability [40], [41]. Collision probabilities of high definition robots in a dynamic environment are used with Gaussian distribution where the obstacles are represented in the workspace as a probability distribution [42]. In [43], the authors introduced and implemented an obstacle-dependent Gaussian potential field method for autonomous vehicles to avoid obstacle collision [43]. Mathematical models are present to facilitate the optimization of motion planning in a dynamic workspace. Examples of models based on analytical models include Gaussian distribution [34], [35], [42], random tree [35], and Monte Carlo sampling strategies [34], [36]. Current research [35], [36], [38], [44]-[46] indicates that the focus is primarily on the improvement of robotic path planning using different mathematical models. However, there is no substantial evidence of any mathematical form that can be used to calculate a process of creating random motion such as Simple Random Walk (SRW) for studying the effect of static and dynamic obstacles.

In recent research, the Random Walk model has been used in robot motion planning such as the swarm robotics area exploration strategy where the random walk methods, based on Brownian motion, and the levy flight method, are used as the direction moved by a robot after each step is stochastic[46]. New approaches based on Monte Carlo random walk such as bidirectional and optimizing planner Arvand are used to improve motion-planning performance by using simple random walk. The method uses the distance measurement between the goal point and the sampled states [44]. Fast random walk approach that determines diverse and efficient paths from different Homotopy class have also been proposed to improve robotic path planning [45].

4. Simple Random Walk (SRW)

Random Walk is a stochastic process where the path consists of a succession of random steps. Conditional probabilities are used in formulating the probability expression and commonly used in complex situations that make the use of a single simple probability calculation inaccurate. SRW is a stochastic process which involves the collection of identical independent random variables where each of those random variables represents the next move [47]. In a rectangular gridded workspace, the AGV has a 25% chance of moving in any of the four possible directions, at any given grid point position other than the boundaries. When repeated n times, the probability becomes $(25\%)^n$. For example, for a robot moving in a grid workspace, at any given position, the robot has the chance of moving along each of the adjacent grid points. In one-dimensional SRW, the walk state X_1 jumps left or right equally likely at each time along the X-direction. Similarly, in two-dimensional SRW, each step will have an equal probability bounded by a rectangular wall defined by two-end corners $\langle 0,0 \rangle$ and $\langle a,b \rangle$. A step is defined by the displacement unit made by AGV in x and y coordinates. The average steps per walk are used to account for the uncertainties in the robotic motion and to avoid outliers. The average step of an AGV roaming in a finite plane with obstacles, with an initial point "i" and end point "f", has no analytical form. Therefore, a numerical simulation has been being adopted in this research to investigate several case studies related to the ability of the AGV to succeed in completing a mission in the presence of obstacles. Pseudo code algorithms will be used to analyze the probability and the success to arrive at the desired location simulated for different environments

5. Scope of Research

While the previously aforementioned path planning methods require navigation, a capability based on on-board and distributed sensors that can realize and analyze the surrounding environment in real-time, however this paper presents a failsafe method for controlling the path of randomly moving robots or humans during worst case scenario particularly when minimum sensing capabilities are present. The sensing capabilities assumed in this paper are first, capability of sense and avoid, and second, capability to know final state upon arrival but not during the intermediate states. It is assumed throughout the paper that human or robot motion exhibit random walk locomotion. The social behavior model of a randomly walking worker and coexisting with other objects in a workplace is assumed in this research to follow simple random walk scheme if the worker stops upon the arrival to a predefined destination. This method expands on the Random Walk theory to include simulation of freely ranging robots in confined areas comingling with static and dynamic obstacles. This modified process uses random step to steer AGV randomly until it reaches the end point. Directing robot performance to go to a safe home position without any guidance is currently not determined. One primary research question is to understand how the distribution of obstacles within a workspace could affect the ability of robot to achieve a simple task, and how marching locomotion could help in the design and layout for a safe working environment.

The following section introduces the 'Obstacle Design' approach method. This method is based on probabilistic approach through simulation on the effect of obstacles using the Simple Random Walk (SRW) model in a shared workspace.

6. Method: Navigation based SRW

When the navigation systems on an AGV or a robot fail, the best the users can hope for is to send a signal to shut it down or direct it to a safe home position. The use of simple and robust sensors, such as inductive sensors, could be a viable solution to report whether a free ranging AGV has arrived at a desired state. This could be analogous to having a blind person randomly navigates his or her way in a room to an exit door. These free marching walks, whether applied to a person or robot, can be modeled by the SRW principle that randomizes steps during a journey to a desired location. This sensor-less roaming method, when combined with obstacle avoidance, could be an alternative to the navigation system when the communication or processing malfunction. On the other hand, with the increasing installation of AGVs in manufacturing plants, and with demand for inclusion of collaborative human robot interaction, it is desired to understand the risk of such collaboration at worst-case scenario. That is workers randomly and carelessly roaming around AGVs. The question that one could ask is, what is the probability of an accidence within given manufacturing workspace area, and what should be done better to improve the facility layout in such a way that if a layout is strategically reconfigured the accident probability reduces.

The relationship between robots and obstacles will be examined through an obstacle design approach. In a real-time workspace, AGV may malfunction and lose its navigational capability and that may lead to robot- robot collisions or human injuries and/or fatalities. To avoid these undesirable events, the relationship between the variables in the workspace must be studied and evaluated. This research develops an environment and a model for robot or human stepping randomly around several obstacles within a specified boundary. An array of static and dynamic obstacles will be populated randomly and deterministically within the workspace. This is to introduce various obstacle design approaches and then evaluate the average number of steps taken by the robot to reach its final position.

The proposed simulated algorithms attempt to answer the following three research questions:

- How does the distance *D,* measured between an initial and a final position, relate to the success of the robot to reach its final position when *m* number of obstacles are placed randomly in a confined workspace?
- Would such success be improved if the obstacle were determinately arranged in pattern such as L-shape, passage, and Zigzag?
- What is the success of the robot in the workspace where obstacles are undergoing Markov chain motion or random walk?

6.1. Workspace Conditions

In the simulation examples that follow, the workspace is defined by a rectangular area whose size is (*a*, *b*) as shown in Figure 1. *n* is the number of swarming robots that are comingling with *m* number of stationary obstacles and with *h* number of obstacles in motion. All robots and obstacles in motion or "dynamic obstacles" follow SRW properties. The location of static obstacle does not change with time. Let $s(t)$ define the step count at time or discreet event *t*. The location of an object in the Cartesian coordinates can then be defined by the value of their components for a given step count $(x(s(t)), y(s(t)))$. For example, robot *k* and dynamic obstacle *w* march simultaneously under SRW algorithm. The walk stops when all robots in a subset reach their desired locations. Let us say that the subset has one robot at zero step with initial position $(x(s(0)), y(s(0)))$, and then it requires $s(t_f)$ steps to reach the final location $(x(s(t_f)), y(s(t_f)))$. During one walk journey, a robot could collide with a workspace boundary, an obstacle, or another robot.

Multiple walks will be simulated to average the number of steps it takes a robot to reach destination starting from the same initial condition. In one walk, a success is defined as the average number of steps taken by the robot to reach its final position. However, not all walks could lead to success. Therefore, a cut-off step of 10^6 is applied to limit the maximum number of steps within one walk.
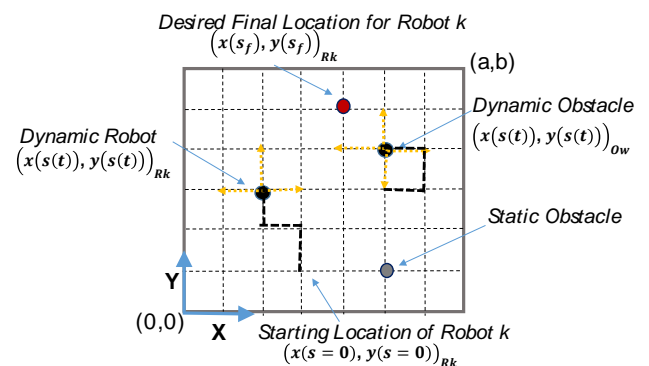

Figure 1: Robots & obstacles marching in confined workspace with stationary obstacles.

It is assumed that the robot has a defense mechanism such as a simple obstacle avoidance bumper switch that instructs the robot to try to move away by repeating the step again hoping it does not land on a boundary or another object. Such overlap is counted as a step and it will weight on the robot's success to find its desired location. On the other hand, obstacles are assumed to have least intelligence; therefore, they do not change the course of motion when overlaps occur. An example of that could be a worker wondering in workspace and not paying attention to surrounding.

In practice, AGVs or in general, most programmable mobile robots are equipped with basic obstacle avoidance sensors such as low range ultrasonic sensors or spring-loaded switches. The mobile robot could know its final position by means of magnetic, color, inductive, or capacitive sensors mounted on the floor. The examples discussed in the following simulation subsections assume that all navigation sensors are turned off except for the obstacle detection sensor and final state sensor. In this paper,

simulations of AGVs are studied in several environments defined in the following subsections.

6.2. Random Walk Roaming amidst Static Obstacles

The static obstacles are defined as non-moving objects that are randomly and deterministically structured within the workspace. The obstacles(s) are placed in such a way to mimic various kinds of obstacle arrangements. Random Walk amidst an array of static obstacles is implemented by using Algorithm 1.

Algorithm (1): Random walk of a robot within confined environment with static obstacles.

Algorithm (1) – Static Obstacles

S = 0
B ← workspace boundaries
O ← randomly placed obstacle(s) positions or deterministic boundaries
C ← cutoff
W ← number of journeys robot takes from (X_i, Y_i) to (X_f, Y_f)
For w = 1:W
        $(X_r, Y_r) ← (X_i, Y_i)$
        While $(X_r, Y_r) ≠ (X_f, Y_f)$
            S = S + 1
            (R_x, R_y) ← randomly generated ± 1
            $X_r = X_r + R_x$
            $Y_r = Y_r + R_y$
            If $(X_r, Y_r) = B$
                robot pushes back into workspace
            If $(X_r, Y_r) = O$
            robot takes a step back or forward away from O
            Or, robot pushes back into the obstacle region
            If S >= C
                loop breaks
Avg_Step = mean S over all w

Algorithm (2): Random walk of a robot within confined environment with dynamic obstacles.

Algorithm (2) – Dynamic Obstacles
S = 0
B ← workspace boundaries
O ≡ (X_{oj}, Y_{oj}) ← j obstacle(s) performing independent random motion
C ← cutoff
W ← number of journeys robot takes from (X_i, Y_i) to (X_f, Y_f)
For w = 1:W
        $(X_r, Y_r) ← (X_i, Y_i)$
        O ← random initial positions to all obstacles
        While $(X_r, Y_r) ≠ (X_f, Y_f)$
             S = S + 1
            (R_x, R_y) ← randomly generated ± 1
            $X_r = X_r + R_x$
            $Y_r = Y_r + R_y$
            For j ε O
                (Q_{xj}, Q_{yj}) ← randomly generated ± 1
                $X_{oj} = X_{oj} + Q_{oj}$
                $Y_{oj} = Y_{oj} + Q_{oj}$
            If $(X_r, Y_r) = B$
                robot pushes back into workspace
            If O = B
                obstacle(s) pushes back into workspace
            If $(X_r, Y_r) = O$
                robot takes a step back or forward away from O
            If S >= C
                loop breaks
Average Steps = mean S over all w

6.3. Random Walk Roaming amidst Dynamic Obstacles

The dynamic obstacles are defined in this research as moving objects under constant change in motion. The obstacles

can move either on a preprogrammed path or on a random walk that abides by different robot-robot and human-robot interactions. Algorithm 2 describes the boundary conditions for dynamic obstacle. The Dynamic Obstacles are simulated for a Markov Chain obstacle and Random Walk obstacle. The *Markov Chain obstacle design* is based on the *Markov Property* which states that "Future is independent of the past given the present" [48]. Mathematically this statement can be expressed as

$$P_r(S_{t+1} | s_t) = P_r(S_{t+1} | s_1, \ldots, s_t) \qquad (1)$$

where in the above equation, s_t or $S(t)$ denotes the current state and $S(t + 1)$ or S_{t+1} denotes the next state. The transition state of the Markov Chain obstacle is entirely independent of the past states. An obstacle such as (ow) in the workspace (a,b) in Figure 1 is based on Markov Chain model. In this obstacle design, the obstacle changes its positions on n fixed number of states with the probabilities from a fixed transition matrix. The Markov Chain in this study is assumed to be known absorbing. On the other hand, the Random Walk Obstacle design assumes a free-ranging AGV roaming between random obstacles. In this study, the robot such as (Rk) and obstacle(s) such as (Ow) both are moving on a Random Walk model. The obstacles are independent entities moving around the workspace (a,b) generating random initial positions with no destinations. The random walk concept is related to the Hamiltonian cycle. A Hamiltonian cycle is a path along a set of points that passes through every point exactly once before returning to the original start point. In the context of an AGV, a Hamiltonian cycle would allow the vehicle to pass through every point on a grid that defines its working zone. Generating a Hamiltonian cycle ensures that the robot reaches the end-point, but at the cost of moving efficiently. While this method will not be used directly, the results of the random walk can be compared to the Hamiltonian cycle of the operating space.

6.4. Probability of Presence

The "*Probability of Presence*" of a robot for a specified obstacle region are also simulated to present long-term chances of a robot to be in a certain position. To pursue a long-term probability of presence, the robot is performing infinite SRW with no destination within the obstacle region [49]. The theorem stipulates that the long-term stationary probability matrix *P* of random walk be computed from a matrix degree of freedom *D*

$$P = \left(\frac{1}{\sum D_{ij}}\right) \qquad (2)$$

where D_{ij} is the number of directions the robot can move when at the position (i,j) and $D_{ij} = 0$ for (i, j) at obstacle's position [50]. It is important to note that here the *P* is an invariant or stationary distribution vector written in form of a matrix since we have a 2D workspace grid.

6.5. Proposed Obstacle Designs

The proposed approach is to ensure safety through probabilistic "*Obstacle Design*" approach in a workspace and are composed of mainly six types of structured static obstacle designs including: Diagonal Design, Passage Design, L-shaped Design, Zig Zag Design and Dynamic Obstacles Design. The dynamic obstacle contains two types (i) Markov Chain Obstacle

(ii) Obstacles performing Simple Random Walk. The main emphasis behind making different obstacle designs is to cover the geometry of real-world workspaces and study their tradeoff. In this paper, our hypothesis is that various obstacle designs have a significant effect on the success of a freely moving AGV.

Both Algorithms (1) and (2) simulate the step movements of AGV randomly within workspace. The AGV pushes back when it hits a static or dynamic obstacle (such as another AGV) or workspace boundary and recalculate its walk again. These steps are still counted in the measuring the efficiency of the AGV. Algorithm (1) deals with structured static obstacles that are configured through various design approaches. Meanwhile, Algorithm (2) is deployed for randomly moving dynamic obstacles.

7. Simulation: Randomly Walking AGV within Structured Static Obstacles and in Confined Workspace

7.1. Diagonal Obstacle Design

The first obstacle design examines the relationship between a free ranging AGV and any m number of obstacles arranged diagonally. Let the linear accumulative distance, D index be defined by,

$$D = \sum_{j=1}^{m} \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} + \sqrt{(x_j - x_f)^2 + (y_j - y_f)^2} \quad (3)$$

where (x_j, y_j) is the location of the ij^{th} obstacle. The simulation case study examines 22 obstacles arranged diagonally between initial position at (x_i, y_i) at (1,1) unit, and final desired location (x_f, y_f) at (25,25) unit. The Average Step (AS) corresponding to, D, is simulated for every $m \in \{1,2,\dots 22\}$. The robot locomotion is set to take place within a confined workspace area of (50×50) unit square, as shown in Figure 2. Algorithm (1) is deployed to compute the average number of successful steps it takes the robot to travel from the initial point to the final point. According to the 'Law of Large numbers' theorem [51], it is sufficient to use 300 random "walk" in order to compute the average number of successful steps.
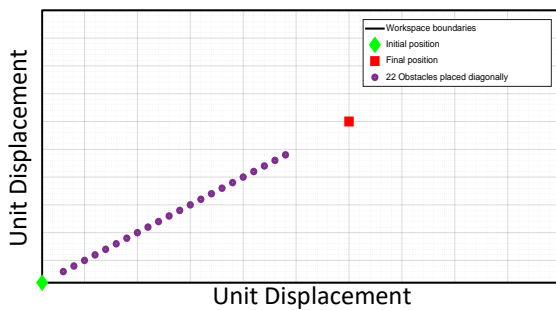
Figure 2: Diagonal design with m number of obstacles are alighned diagonally between initial and final locations.

The relationship between the diagonal obstacles and the average successful steps for every m diagonal obstacles are computed and plotted in Figure 3. The results show that the more linear obstacles are added between the start and final positions, the less steps will require AGV to find the desired location.
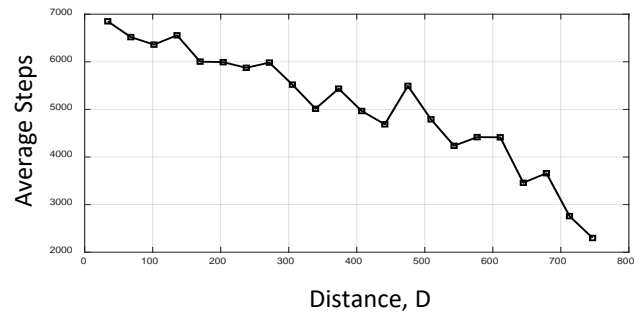
Figure 3: The effect of installing diagonal obstacles on the success of a ranging AGV to finds its destination.

The probability of occurrence, i.e. the long-term probability in equation (2), is simulated for workspace area of size (20×20) unit square with eight obstacles located between initial point (1,1) unit, and final point (10,10) unit. The probability of occurrence can help predict the robot path and warn the facility workers about which area to avoid during work. It can be easily concluded from the results in Figure 4 that the probability of occurrence of robot colliding with nearby obstacle increases. In addition, the chance to find the AGV elsewhere in the workspace is uniformly minimum. As a result, an AGV exhibiting a random walk in a confined area can find its destination faster by creating a physical or virtual diagonal obstacle within proximity of desired final location. In general, one advantage of installing the obstacle is due to the high probability of finding a robot nearby the boundary.
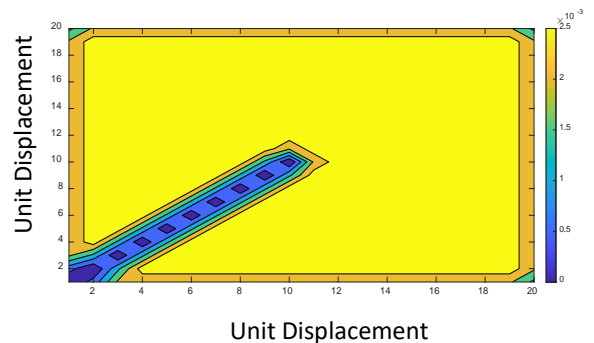

Figure 4: Probability of occurrence for the diagonal obstacle design.

7.2. L-Shape Obstacle Design

The second case study deals with an obstacle design configured in L-shape with fixed width, W, as suggested in Figure 5. The length of the L-shape corresponds to the maximum linear distance traveled between two points within the workspace. The initial and final desired position of a free ranging AGV are assumed at (1,1) and $(a - 1, b - 1)$, respectively. The AGV stays within the L-shape area. Algorithm (1) is executed to determine the average number of successful steps for various width.
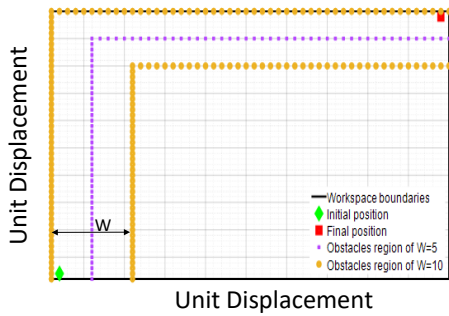
Figure 5: Example of 'L- shaped' obstacle containining the initial and desired locations.

Multiple simulations are carried out for widths ranging from $W = [1,20]$ and at step size, ΔW, of one unit. Consider three set of workspaces sizes (50×50), (75×75) and (100×100) unit square. The average success vs. width of L-shape is computed for workspace size, and all curves are plotted in Figure 6.a. Let the "*efficiency*" be defined by the average success normalized to the area size of the workspace. The average steps/enclosed L-shaped obstacle area is plotted in Figure 6.b.
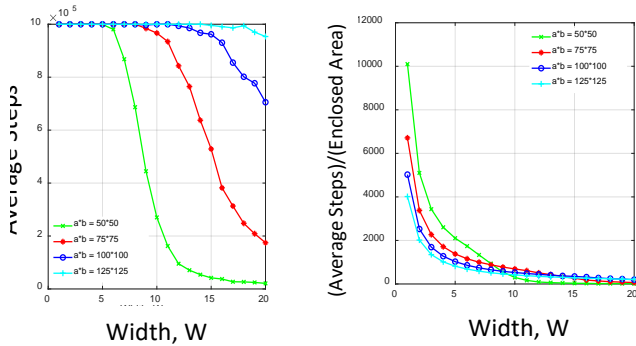
Figure 6: L-shape obstacle design.: (a) Success simulated for several widths and workspaces sizes. (b) Effeciency of successful search.

Let a cutoff threshold of 10^5 define the maximum number of steps before the AGV stops searching. Figure 6.a shows a rapid decay relationship, where it takes an AGV less steps to succeed in smaller workspace. The normalized plot in Figure 6.b implies that the exponential decay relationship becomes independent of the workspace size when the width size is large. However, the overall efficiency of search is better for the large workspace size within $W \in[1 \sim 7]$ unit.

7.3. Passage Obstacle Design

Passage obstacle design refers to rectangular area that confines the movement of AG, as suggested in Figure 7. It is assumed that the starting and final positions are always located within the passage area, and distance between these points defines the passage length. In this case study, the intent is to determine how the width of the passage, *W*, relates to the success of the AGV to find final position. Let the starting and final position be at (1,1) unit, and (a-1, b-1) unit, respectively. Algorithm (1) is deployed to study this relationship for various workspaces sizes, including (50×50), (75×75), (100×100), and (125×125) unit square. The simulation is also carried out for width $W = [1,20]$unit, and at step size, ΔW, of 1 unit. It can be

observed from the results shown in Figure 8.a that the size of the workspace has little significance to the average success. The average success drops significantly for widths ranging from 1 to 5 unit. This sharp exponential decay could explain why all workspaces have almost identical efficiency, as observed in Figure 8.b.
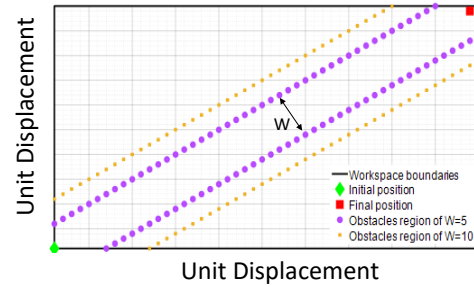

Figure 7: Passage Design is created encompassing the initial and desired location of AGV.
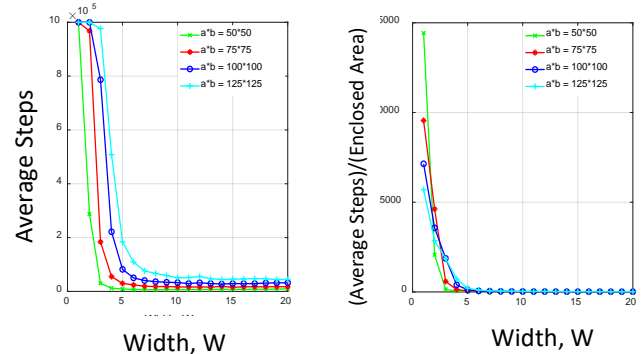

Figure 8: Passage obstacle design: (a) Success simulated for several widths and workspaces sizes. (b) Effeciency of successful search.

A more in-depth simulation is performed to examine if the curves in Figure 8.a reach steady state. To make the analysis more robust, a higher walk value of 9600 is chosen, as it is highly likely to converge to the true value. Across different workspace sizes, it is observed that the average steps concave up after attaining its lowest possible value. It is evident from Figure 9 that the lowest value corresponds to an optimum width at which the average number of steps is minimum. It is, however, important to note that the increase in the optimum width is very small with respect to the increase in workspace sizes. It is also observed that the optimal width shifts up as workspace size increases.
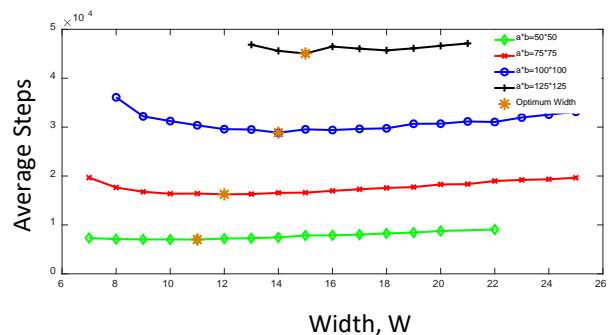

Figure 9: Optimum widths for passage design computed at different workspace sizes.

7.4. Defected Passage Obstacle Design

Consider *p* number of consecutive obstacles being removed from the passage design. Let the defects be present in the "lower side" or "upper side" of the passage as shown in Figure 10. These two configurations are asymmetric along the diagonal line that connects the initial to final position. However, the large area that is open to the passage is equivalent in both case, and therefore, simulations will be carried out separately for each case.
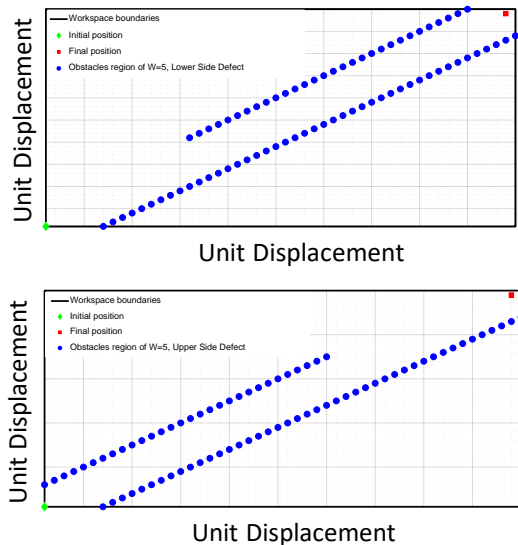

Figure 10: Example of defect on one diagonal side of the passage: (a) Lower side defect (b) Upper side defect.
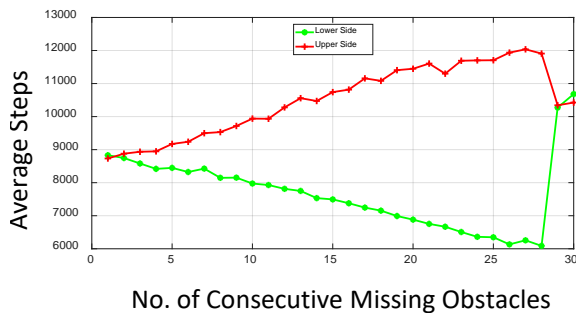
Figure 11: AGV success in the realm of defective obstacle design.

The simulations are carried out for "upper side" and "lower side" defected passage within the workspace of (50×50) unit square. The goal is to plot the average success against the number of obstacles removed. The order of the obstacle removed in the "lower side" case is chosen to begin from the lower corner of the workspace, i.e. removal begins closer to the initial position. The "upper side" case, however, begins from the upper corner of the workspace, i.e. removal begins from the final position. The maximum number of obstacles in each side is 28 units. It is clear from the simulation results in Figure 11 that in the case of defects in the "lower side", the average success improves (or linearly decreasing) as the defects (or the number of sequentially removed obstacles) increase. Whereas an exactly opposite effect is observed for the defect in the "upper side". The slopes obtained from the linear regression for defects [1 27] are equal in magnitude, but opposite as observed in Figure 11. This implies

that when *k=1*, the average success is identical whether the obstacle is removed from upper side or lower side. However, the average success is farthest apart when only one obstacle is left close to the "upper side" or "lower side". Interestingly, placing only one obstacle near to the final position would significantly improve the success of the AGV to find its final position. However, at *p*=28, the average steps between lower and upper sides drops drastically and becomes identical simply because the entire obstacles in the diagonal side is removed.

7.5. Zig-zag Obstacle Design

Zig-zig obstacle, such as example in Figure 12, could be utilized across the workspace to mimic certain complexities of an industrial workspace. The average success of an AGV performing SRW is studied against widths by using Algorithm (1). The width, W, defined in Figure 12, is varied between [1,30] unit, with increment, ΔW, of 1 unit. Various workspaces sizes were tested including (50×50), (75×75), (100×100), and (125×125) unit square.
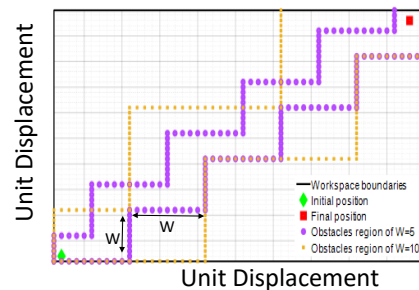

Figure 12: Example of zigzag obstacle design.

The simulation result shown in Figure 13.a depicts a fluctuating decay of the average success as the width increase. In addition, the average success curves, corresponding to each workplace size, are significantly spread apart from each other. These effects could be related to irregularity of the obstacle design, especially because the interior search area is sensitive to the width. On the other hand, the normalized plot in Figure 13.b shows that the efficiency of search in the zigzag design is approximately uniform when the workspace size is scaled.
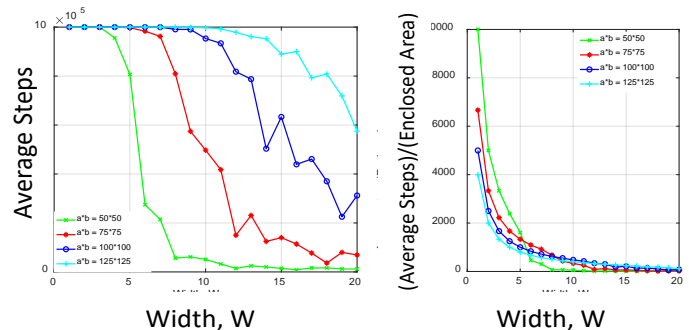

Figure 13: Zig-zag obstacle design: (a) The average success is simulated for several widths and workspaces sizes. (b) Effeciency of successful search.

7.6. Comparison Between Obstacle Designs

A simulation study is conducted to evaluate how effective the obstacle designs are, in terms of their ability to guide an AGV

from initial position $(1,1)$ into a final desired position located at $(a-1, b-1)$. Consider the test conducted on a workspace size of $(50{\times}50)$ unit square. The average success is simulated for passage, L-shaped and zigzag designs, and for various width. The results shown in Figure 14 shows that the passage design performed better than other types for all simulated widths. The L-shape obstacle design was the least effective, however, the decay profile is smooth, and it could be approximated by Cosine Annealing function.
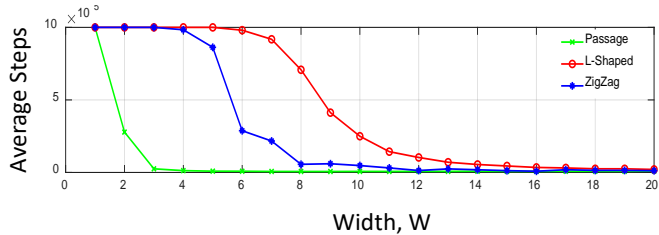
Figure 14: Comparison between passage, L-shaped and zigzag design.

8. Simulation: Randomly Walking AGV within Random Static Obstacles and in Confined Workspace

A case study is put forward to examine the relation between a free ranging AGV, and *m* number of random obstacles placed inside a workspace. Let the degree of obstacle aggregation be defined by scattered distance

$$D_s = \sum_{j=1}^{m} \left(x_j - \bar{x}\right)^2 + \left(y_j - \bar{y}\right)^2 \qquad (4)$$

where $(\bar{x}, \bar{y})$ is the mean value of the scattered obstacles. Simulations are carried out to calculate the average success for *m* number of various subsets of randomly placed obstacles. Each *m* subset of obstacle was also iterated, *q,* number of times.

A simulation example is conducted for an AGV roaming within a confined workspace area of $(50{\times}50)$. The AGV is assumed to be initially placed at $(1,1)$ unit, with target destination located at $(25,25)$ unit. Consider that *m* subsets belong to individual values in set $\{1,4, 9,16, 25, 36\}$. The number of iterations tested per subset is set to $q=50$. The average success is plotted against the scattered distance and linear accumulative distance (Equation 3) for every iteration, with results shown in Figure 15.

Let the *dispersion of randomly placed obstacles*, **δ**, be defined by the range of D_s for *q* iterations of a subset *m*. Also, let the *variation in the average success*, **ξ**, be defined by the range of the average successes obtained for *q* iterations of a subset *m*. The following observation can be depicted from Figure 15.a: (1) Although the average success computed for each iteration seems random, the mean-value calculated for all the average success iterations increases as *m* increases, i.e. has a negative impact, and (2) **δ** and **ξ** widen as *m* increases. The linear cumulative indicator, *D*, used in the simulation shown in Figure 15.b, references the obstacles relative to the initial and final positions, instead of the mean-value. The difference between *D* and D_s indicators is regarded to the dispersion of data points. The *D* value of each *m* obstacles correspond to distinct aggregation as compared to D_s.
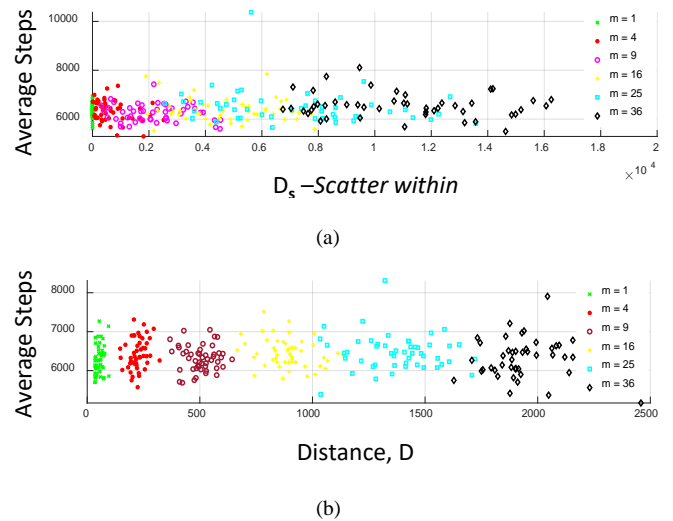

(a)

(b)

Figure 15: The effect of randomly placed static obstacles on the success of a free ranging AGV to find its destination .(a) compared with scattered distance. (b) compared with linear accumulative distance.
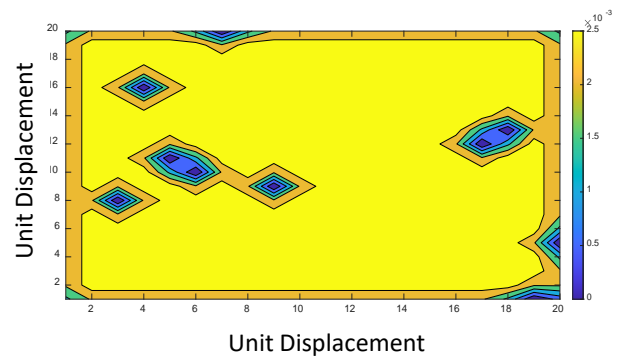

Figure 16: Probability of occurance for randomly placed static obstacles.

An example of long-term probability is simulated for an AGV performing random walk in a workspace $(20{\times}20)$ unit square, amidst 10 randomly generated static obstacles. Figure 16 shows long-term likelihood of the AGV to be present in the light region, "yellow-color", whereas the dark regions, "blue-color", corresponds to avoided region.

9. Simulation: Multiple Randomly Walking AGVs in Confined Workspace

This study mimics a real-world collaborative workspace where human-robot or robot-robot interactions co-exist in a confined workspace with randomly placed static obstacles. Mobile robots or AGVs, and human motion are assumed to undergo independent SRW.

The first simulation case study is based on AGV performing SRW motion from the initial point to the final point in the presence of another independent AGV that is in 5-state Markov Chain motion. The Markov Chain implies that the AGV can only jump from one state into one of the other 5-states located within the workspace. Let the workspace size be assigned to $(50{\times}50)$ unit square, with the initial and final desired position for one AGV be located at point $(1,1)$ unit, and $(49,49)$ unit, respectively. Algorithm (2) is implemented to compute the average number of successful steps and the average *collision probability* by using 300 random walks per test.
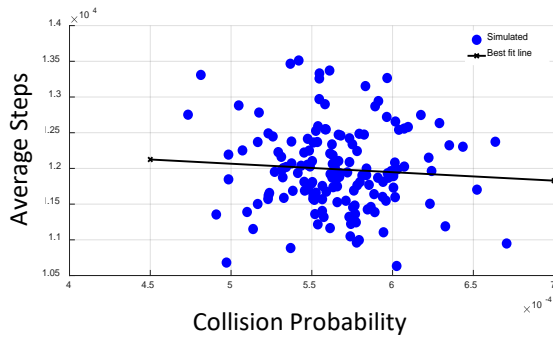
Figure 17: Simulation of collision probability of an AGV and a '5 state Markov Chain' AGV, in workspace free of obstacles.

The simulation calculates the average success against the collision probability for 150 random tests, with results shown in the scatter plot in Figure 17. The linear regression fit has a slight negative slope; indicating that an increase in the probability of the collision between the two AGVs has resulted in a decrease in the average success of the free ranging AGV. The mean values of the scatter plot data are approximately, $\sim(1.2 \times 10^4, 5.6 \times 10^{-4})$, which is uniquely characterizes the average number of successful steps and probability of collision for the given boundary conditions.

The second case study considers (m+1) independent AGVs in dynamic motion, where all are assumed to undergo simple random walk, i.e. they can travel from any point to another point in a step wise manner. One of these AGVs is tasked to navigate from same the previous initial to the final desired positions, while others are in continuous random motion. The obstacle avoidance is enforced between any combination of AGVs, while all are kept inside workspace. Conversely, this case study is equivalent to a robot swarming among a crowd of randomly moving humans.
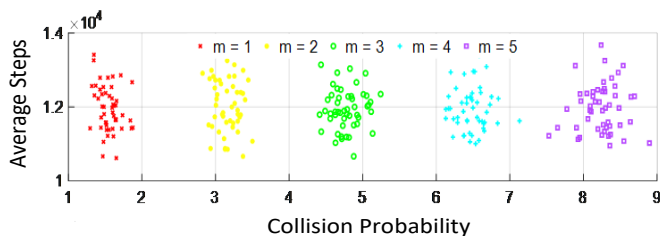

Figure 18: Collision Probability of an AGV performing SRW with 'm' number of SRW AGVs.

Simulations are carried out for several number of AGVs described in group set $m = \{1, 2, 3, 4, 5\}$, where each group is simulated 50 iterations. The scattered data shown in Figure 18 indicates that the mean values of the average steps evaluated for each group are approximately equal. I.e., the number of roaming AGVs, m, has little impact on the average success. However, the increase of number of AGVs installation results in increase in the probability of collision.

The effect of the diagonal distance measured between the initial and final positions, R, is examined within the context of this case study. Let the workspace be modified into (100,100). The average success is simulated for same set of groups and plot

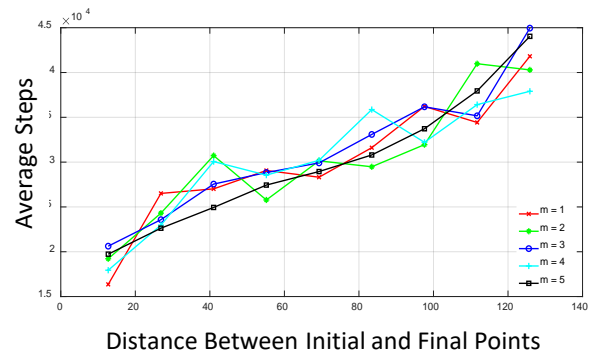against R. This relationship is linear for all m, as shown in Figure 19.

Figure 19: The effect of distance between initial and final points on the success of a free ranging AGV to finds its destination.

In addition, the linear plots confirm that in an industrial workspace, increasing random dynamic obstacles has no impact on the average number of steps required for an AGV to find its destination.

10. Conclusions

This paper examined the random walk locomotion in confined workspace. The goal of the research is to help design a safe shared workspace where humans and robots could co-exist, and further propose a navigation-free method that could help AGVs with malfunctioned navigation retrieve to a safe location at worse case scenarios. The research tested the hypothesis that placing obstacles in workspace could improve the ability of AGV to locate its destination. Several patterns were tested including L-shape, passage, and zigzag. In general, it can be concluded that as the area (characterized here by the width) of obstacle region increases, the AGV will take greater number of steps to reach its final position. However, there exist a threshold at which the increase of the width would have no effect. Passage Design' is the most efficient way to guide AGV to its final position. In the case of dynamic obstacles, if the obstacles are in Markov chain, the lost AGV will reach its final position in a minimum average number of steps. Finally, as the number of randomly placed obstacles increases, the average number of steps taken by the AGV to reach its final position increases.

11. Future Work

Future work will extend the 2D random walk to study 3D complex interaction of swarm of robots based on Brownian motion model.

Acknowledgement

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] R. Bostelman, "AGMs and forklifts gaining sight for safety," Material. Handling and Logistics Magazine., **68**(1), 34–37, 2013, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=912983.

[2] SICKUSAblog, "Additional solutions for collision avoidance with indoor AGVs," SICKUSAblog, https://sickusablog.com/additional-solutions-collision-avoidance-indoor-agvs/, Jul. 2019.

[3] R. Koster, T. Le-Anh, J. van der Meer, "Testing and classifying vehicle dispatching rules in three real-world settings," Journal of Operations. Management, **22**(4), 369–386, 2004, https://doi.org/10.1016/j.jom.2004.05.006.

[4] A. P. Vancea, I. Orha, "A survey in the design and control of automated guided vehicle systems," Carpathian Journal of Electronic and Computer Engineering, **12**(2), 41–49, 2020, doi: 10.2478/cjece-2019-0016.

[5] E. Cardarelli, L. Sabattini, V. Digani, C. Secchi, C. Fantuzzi, "Interacting with a multi AGV system," in IEEE Intelligent Computer Communication and Processing International Conference, 263–267, 2015, doi: 10.1109/ICCP.2015.7312641.

[6] A. Prusak, J. Bernshausen, H. Roth, J. Warburg, C. Hille, H.-H. Gtting, T. Neugebauer, "Applications of automated guided vehicle (AGV) and industry robots with PMD-camera," 2007, 299-303.

[7] C. Liu, M. Tomizuka, Designing the robot behavior for safe human–robot interactions, Springer Internatinoal Publishing, 2017.

[8] L. Sabattini., V. Digani, C. Secchi, G. Cotena, D. Ronzoni, M. Foppoli, F. Oleari. "Technological roadmap to boost the introduction of AGVs in industrial applications," In 2013 IEEE 9th International Conference on Intelligent Computer Communication and Processing 203–208, 2013, doi: 10.1109/ICCP.2013.6646109.

[9] B. S. P. Reddy & C. S. P. Rao, "A hybrid multi-objective GA for simultaneous scheduling of machines and AGVs in FMS," Internatinal Journal Advanced Manufacturing Technology., **31**, 602–613, 2006, doi: https://doi.org/10.1007/s00170-005-0223-6.

[10] M. V. S. Kumar, R. Janardhana, and C. S. P. Rao, "Simultaneous scheduling of machines and vehicles in an FMS environment with alternative routing," Internatinal Journal Advanced Manufacturing Technology, **53**(1-4), 339–351, 2011, doi: 10.1007/s00170-010-2820-2.

[11] J. Jerald, P. Asokan, G. Prabaharan, and R. Saravanan, "Scheduling optimisation of flexible manufacturing systems using particle swarm optimisation algorithm," Internatinal Journal Advanced Manufacturing Technology, **25**(9–10), 964–971, 2005, doi: 10.1007/s00170-003-1933-2.

[12] N. Krishnamurthy, R. Batta, M. Karwan, "Developing conflict-free routes for automated guided vehicles," Operations Research, **41**(6), 1077–1090, 2018, https://doi.org/10.1287/opre.41.6.1077.

[13] D. Zeitler, M. Werner, G. Koff, "Dynamic object avoidance with automated guided vehicle", Washington, DC, US Patent *20050021195*, 2005.

[14] D. Emanuel, R. Kunzig, R. Taylor "Method and apparatus for collision avoidance" Washington, DC, US Patent *8346468B2*, 2013.

[15] T. Park, "AGV control system and method," Washington, DC, US Patent *7305287B2*, 2007.

[16] J. Kondo, "Travel control device for unmanned conveyence vehicle,", Washington, DC, US Patent *8406949B2*, 2013.

[17] S. Lee, W. Jeong, "Apparatus and method for estimating a position and an orientation of a mobile robot" Washington, DC, US Patent, *20100070125A1* 2010.

[18] P. Lasota, T. Fong, J. A. Shah, "A Survey of methods for safe human-robot interaction," Foundation and Trends in Robotics., **5**(3), 261–349, 2017, doi: 10.1561/2300000052.

[19] G. Michalos, S. Makris, P. Tsarouchi, T. Guasch, D. Kontovrakis, G. Chryssolouris, "Design considerations for safe human-robot collaborative workplaces," Procedia CIRP, **37**, 248–253, 2015, doi: 10.1016/j.procir.2015.08.014.

[20] R. Inam, K. Raizer, A. Hata, R. Souza, E. Forsman, E. Cao, S. Wang, "Risk Assessment for human-robot collaboration in an automated warehouse scenario," in 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), 743–751, 2018, doi: 10.1109/ETFA.2018.8502466.

[21] E. Magrini, F. Ferraguti, A. Ronga, F. Pini, A. De Luca, F. Leali, "Human-robot coexistence and interaction in open industrial cells," Robotics Computer-Integrated Manufacturing, **61**, 101846, 2020, doi: 10.1016/j.rcim.2019.101846.

[22] I. Draganjac, D. Miklic, Z. Kovacic, G. Vasiljevic, S. Bogdan, "Decentralized control of multi-AGV systems in autonomous warehousing Applications," Transactions on Automation Science Engineering, **13**(4), 433–1447, 2016, doi: 10.1109/TASE.2016.2603781.

[23] R. Miller, R. Subrin, Automated Guided Vehicles and Automated Manufacturing. Society of Manufacturing Engineers, 1987.

[24] N. Möller, S. Ove Hansson, J. Holmberg, C. Rollenhagen, Handbook of Safety Principles, Wiley, 2018.

[25] T. Helmer, Development of a methodology for the evaluation of active safety using the example of preventive pedestrian protection, Springer Internatinoal Publishing, 2015.

[26] MMH Staff, "AGVs enhance automotive powertrain production," Modern Material Handling, Oct. 2018.

[27] AGV Technology, "Understanding AGV safety systems," AGV Network, https://www.agvnetwork.com/automated-guided-vehicles-safety-systems, Jun. 2020.

[28] C Frese, A Fetzner, C. Frey. "Multi-sensor obstacle tracking for safe human-robot interaction", in International Symposium on Robotics, 784-791, 2014, permalink http://publica.fraunhofer.de/documents/N-298607.html.

[29] J. H. Chen and K. T. Song, "Collision-free motion planning for human-robot collaborative safety under cartesian constraint," in 2018 IEEE International Conference on Robotics and Automation (ICRA), 4348-4354, 2018, doi: 10.1109/ICRA.2018.8460185.

[30] P. Wenger, P. Chedmail, "Ability of a robot to travel through its free work space in an environment with obstacles," International Journal of Robotics Research, **10**(3), 214–227, 1991, doi: 10.1177/027836499101000303.

[31] N. Vahrenkamp, H. Arnst, M. Wachter, D. Schiebener, P. Sotiropoulos, M. Kowalik, & T. Asfour. "Workspace analysis for planning human-robot interaction tasks," in 2016 IEEE-RAS 16th International Conference on Humanoid Robots, 1298–1303, 2016, doi: 10.1109/HUMANOIDS.2016.7803437.

[32] F. Rezazadegan, J. Geng, M. Ghirardi, G. Menga, G. Camuncoli, "Risked-based design for the physical human-robot interaction (pHRI): an overview," Chemical Engineering Transactions, **43**, 1249–1254, 2015, doi: 10.3303/CET1543209.

[33] E. Kim, R. Kirschner, Y. Yamada, S. Okamoto, "Estimating probability of human hand intrusion for speed and separation monitoring using interference theory," Robotics and Computer-Integrated Manufacturing., **61,** 101819, 2020, doi: 10.1016/j.rcim.2019.101819.

[34] A. Lambert, D. Gruyer, G. Saint Pierre, "A fast Monte Carlo algorithm for collision probability estimation," in 2008 10th International Conference on Control, Automation, Robotics and Vision, 406-411, 2008, doi: 10.1109/ICARCV.2008.4795553.

[35] J. van den Berg, P. Abbee, K. Goldberg, "LQG-MP: Optimized path planning for robots with motion uncertainty and imperfect state information," International Journal of Robotics Research, **30**(7), 895–913, 2011, doi: 10.1177/0278364911406562.

[36] S. Patil, J. van den Berg, R. Alterovitz, "Estimating probability of collision for safe motion planning under Gaussian motion and sensing uncertainty," in IEEE International Conference on Robotics and Automation, 3238-3244, 2012, doi: 10.1109/ICRA.2012.6224727.

[37] F. Flacco, T. Kroeger, A. De Luca, O. Khatib, "A depth dpace approach to human-robot collision avoidance," in 2012 IEEE International Conference on Robotics and Automation (IRCA 2012), 338 - 345, 2012, http://toc.proceedings.com/15154webtoc.pdf.

[38] J. H. Liang, C. H. Lee, "Efficient collision-free path-planning of multiple mobile robots system using efficient artificial bee colony algorithm," Advances in Engineering Software, **79**, 47–56, 2015, doi: 10.1016/j.advengsoft.2014.09.006.

[39] N. Pedrocchi, F. Vicentini, M. Malosio, L. M. Tosatti, "Safe human-robot cooperation in an industrial environment," International Journal of Advanced Robotic Systems, **10(1)**, 2013, doi: 10.5772/53939.

[40] M. Mayyas, S. P. Vadlamudi, M. A. Syed, "Fenceless obstacle avoidance method for efficient and safe human--robot collaboration in a shared work space," International Journal of Advanced Robotic Systems, **17**(5), 2020, doi:10.1177/1729881420959018.

[41] D. Claes, K. Tuyls, "Multi robot collision avoidance in a shared workspace," Autonomous Robots, **42**(8), 1749–1770, 2018, doi: 10.1007/s10514-018-9726-5.

[42] C. Park, J. Park, D. Manocha, "Fast and bounded probabilistic collision detection for High-DOF trajectory planning in dynamic environments," Transactions on Automation Science and Engineering, **15**(3), 980-991, 2018, doi: 10.1109/TASE.2018.2801279.

[43] J. H. Cho, D. S. Pae, M. T. Lim, T. K. Kang, "A Real-Time Obstacle Avoidance Method for Autonomous Vehicles Using an Obstacle-Dependent Gaussian Potential Field," Journal of Advanced Transportation, **2018**, 2018, 5041401, doi: 10.1155/2018/5041401.

[44] S. Carpin, G. Pillonetto, "Motion planning using adaptive random walks," IEEE Transactions on Robots., **21**(1), 129–136, 2005, doi: 10.1109/TRO.2004.833790.

[45] L. Palmieri, A. Rudenko, and K. O. Arras, "A fast random walk approach to find diverse paths for robot navigation," IEEE Robotics and Automation, **2**(1), 269–276, 2017, doi: 10.1109/LRA.2016.2602240.

[46] B. Pang, Y. Song, C. Zhang, H. Wang, R. Yang, "A swarm robotic exploration strategy based on an improved random walk method," Journal of Robotics, **2019**, 2019, doi: 10.1155/2019/6914212.

[47] M. Biskup, Random walks. 1827.

[48] A. Singh, "Reinforcement learning : Markov-decision process (Part 1)," 2019, https://towardsdatascience.com/introduction-to-reinforcement-learning-markov-decision-process-44c533ebf8da.

[49] E. Behrends, Introduction to Markov chains, Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, 2000.

[50] Ryan O'Donnell, Great Ideas in theoretical computer science: Random walks and Markov chains [Video] You Tube, July 15, 2017 https://www.youtube.com/watch?v=86N-awvgi1w, 1:19:31.

[51] N. Etemadi, "An elementary proof of the strong law of large numbers," Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, **55**, 119–122, 1981, doi: 10.1007/BF01013465.

Neural Network for 2D Range Scanner Navigation System

Giuseppe Spampinato[*], Arcangelo Ranieri Bruna, Ivana Guarneri, Davide Giacalone

STMicroelectronics, System, Research and Application, Catania, 95100, Italy

A B S T R A C T

Navigation of a moving object (drone, vehicle, robot, and so on) and related localization in unknown scenes is nowadays a challenging subject to be addressed. Typically, different source devices, such as image sensor, Inertial Measurement Unit (IMU), Time of Flight (TOF), or a combination of them can be used to reach this goal. Recently, due to increasing accuracy and decreasing cost, the usage of 2D laser range scanners has growth in this subject. Inside a complete navigation scheme, using a 2D laser range scanner, the proposed paper considers alternative ways to estimate the core localization step with the usage of deep learning. We propose a simple but accurate neural network, using less than one hundred thousand overall parameters and reaching good precision performance in terms of Mean Absolute Error (MAE): one centimeter in translation and one degree in rotation. Moreover, the inference time of the neural network is quite fast, processing eight thousand scan pairs per second on Titan X (Pascal) GPU produced by Nvidia. For these reasons, the system is suitable for real-time processing and it is an interesting complement and/or integration for traditional localization methods.

1. Introduction

This paper is an extension of the work originally presented in ICARA [1]. Further investigations to increase performance of the approach are done changing deep learning parameters, moreover problem simplification and data augmentation have been tested.

The field of the proposed paper is navigation system and related localization, which is still considered a challenging task [2]. The core function of these kind of systems is for sure the localization itself. Main goal of this vital function is the correct estimation of the step-by-step position of the moving robot in unknown scenes. To perform this task different data sensors can be used. Moreover, to reach better estimation usually previous step data are stored inside an updating map.

About localization approaches proposed in literature, two main localization groups can be recognized: vision-based and laser-based.

Vision-based localization techniques just use images to achieve their goal. Usually, features inside previous and current image are calculated and matched to retrieve the global displacement. In literature, different vision-based techniques have been proposed: effective prioritized matching [3], ORB-SLAM [4], monocular semi-direct visual odometry (SVO) [5], camera pose voting [6], localization based on probabilistic feature map [7],

etc. More sophisticated solutions, like multi-resolution image pyramid methods, have been proposed to reach more robust feature matching [8]. These approaches are usually robust, but they do not have useful distance information, i.e., it is difficult to map the estimated trajectory (in pixel) in the real world (in cm).

On the other hand, laser-based localization techniques just use laser scans to achieve their goal. A features matching approach, as in vision-based localization, is not simple to be implemented. This is mainly due to the poor information of laser scans compared to images. In fact, in the case of image details (e.g., corners), we can note only weak variations in range measurements and then a lack of distinctive features to be analyzed [9].

Usually, Bayesian filtering are largely used in laser-based techniques to consider the robot position as a problem of probability distribution estimation based on grid maps [10,11]. Apart Bayesian filtering, in literature other laser-based proposed techniques are: iterative closest point (ICP) [12] and related variants [13], which minimize the matching error between two point-clouds estimating the related transformation, perimeter based polar scan matching (PB PSM), Lidar odometry and mapping (LOAM) [15], which achieves real time processing by running in parallel two different algorithms, and so on. Even if these approaches usually achieve precise localization, since distance information is available, they can fail in scene changing conditions. In fact, when an object is moving in the scene, due to

the occlusions, we can have lack of information in the moving object area and then the estimated localization can be wrong.

Recently, in vision-based localization, to extract and match image features, deep learning approaches have been successfully used. These promising approaches allow to estimate camera position. A lot of deep learning approaches have been proposed in literature: PoseNet [16], which for pose regression task uses the convolutional neural networks (CNNs), Deepvo [17], which for the same task uses recurrent neural networks (RNNs), undeepVO [18], which estimates the monocular camera pose using a deep learning unsupervised method, and so on. Unfortunately, at moment the deep learning-based methods do not achieve the same pose estimation accuracy of classical vision-based localization approaches.

Inspired by vision-based localization approaches based on deep learning algorithms, few attempts of deep learning methods have also been suggested for laser-based localization: in [19] authors estimate odometry processing 3D laser scanner data with a series of CNNs, in [20] authors trained for giving steering commands a navigation model target-oriented, in [21] authors performed loop closure and matching of consecutive scans making use of a CNN network, in [22] authors improved the odometry estimation considering also temporal features using a RNN, able to model sequential long-term dependencies, and so on.

Regarding the deep learning laser-based localization, as indicated in the case of vision-based localization, unfortunately they still do not achieve the accuracy of the pose estimation compared to classical laser-based localization. For this reason, in literature some authors propose the integration of deep learning approaches with the classical ones: in [23] the authors make use of Inertial Measurement Unit (IMU) in combination with CNNs for 3D laser scanners for assisted odometry, in [24] the authors use the result of vision-based localization approaches based on CNN as starting seed for Monte Carlo localization algorithm, to speed-up algorithm convergence, also increasing robustness and precision, and so on.

It is easy to understand that the field of deep learning localization, in particular about laser-based approaches, has not yet been intensively explored and, at moment, it is still considered a challenging process. In fact, just a small number of papers discuss about this subject [22]. Our choise is to go further in this investigation, to obtain a simple deep learning laser-based localization, using only data taken by 2D laser scanners. Moreover, we tried to reduce as far as possible the number of parameters used by the proposed network, to deal with the low-cost resources constraint.

Our contribution to the research in the field of navigation system, using deep learning approaches, with only 2D laser scan input is firstly the exploration of state of art algorithms. Another contribution is to indicate a methodology to generate the ground truth for the neural network without using real sensors but simulating them with existing powerful navigation tools.

Novelties of the proposed system are in both neural network dataset generation and training. In particular, in data generation we indicate a methodology to choose properly the angle resolution trying to reduce the collisions per frame (to avoid loss of important data) and to maximize array density (to avoid working with sparse data). In this way, the neural network was more able to solve the regression problem.

In the training phase, the novelty is the demonstration with real tests that in regression problems the choice of input/output values scale is vital to let the neural network working. In fact, after several experiments, we obtained the correct scale measures for distances and angles.

At last, the great contribution was to find, after lots of experiments with various neural network hyper parameters, a really light network to solve the localization problem with good performances in terms of mean absolute error between estimated positions and ground truth.

The proposed research is composed by the following Sections: Section 2, where the proposed deep learning laser-based localization is described; Section 3, where the experimental results obtained are deeply described; Section 4, where final considerations are remarked.

2. Proposed Approach

A typical navigation system is described in Figure 1. A starting moving object position (x,y,α) is considered, where (x,y) are the horizontal and vertical position in the cartesian axis and α is the orientation angle. Usually, at the beginning the position is assumed to be (0,0,0). Every time laser scan data is available (θ, d), where θ is the angle and d is the distance from object in front of the laser beam, the localization step will calculate the new position (x',y',α'). The system will then decide next movement. Depending on how it is programmed the moving object (for example the robot), the system can decide to continue moving (in the case an obstacle is not found) or to stop motors (when an obstacle is found). The correct command are then send to the motor control (which interact with the IMU) to update the movement. Positions and movements are updated each time.

Inside the navigation system, the proposed deep learning approach is applied on the core localization step. From this point, this article will focus only on the localization step and all the research will be focused on this particular block of the navigation system.

To reach this objective we used a wheeled robot equipped with the laser scanner A2 RPLidar on the top. This rotating laser scanner has twelve meters as maximum range, view at 360 degrees, running up to fifteen Hz. Thanks to the robot, we acquired a custom dataset in various environments (apartment, laboratory and office).

The ground truth generation schema used is shown in Figure 2. At the beginning, the dataset acquisition is needed to record the input Lidar dataset. Each scan is composed by multiple couples (d, θ), where d is the distance from the object and θ is the related angle. Once the dataset was obtained, we needed to generate the ground truth position (x, y, α) for each sample taken. Since we do not have the real position of the robot for each scan contained in the acquired dataset, we needed a simulation environment to obtain a ground truth.

For this purpose, we make usage of the MATLAB Navigation Tool. The generated ground truth was tested using a simple Mat2Map program to display the path of obtained positions (x, y, α) for each sample taken and the map generated by Lidar scans. In this way, we also checked the robustness of Navigation Tool. Even if it is a very slow method, we tested it in different conditions and we conclude to be very precise, so it was used as reference. It is based on Google Cartographer [25], which builds multiple

submaps and try to align upcoming scans with previous nearby submaps, generating the constrains on a graph.

Once we generated the custom dataset and related ground truth, we perform our experiments using the TensorFlow framework with Keras wrapper in a Python environment. In this configuration, to obtain the best compromise between quality and complexity, we tried different data binarization and augmentation with various neural network configurations.
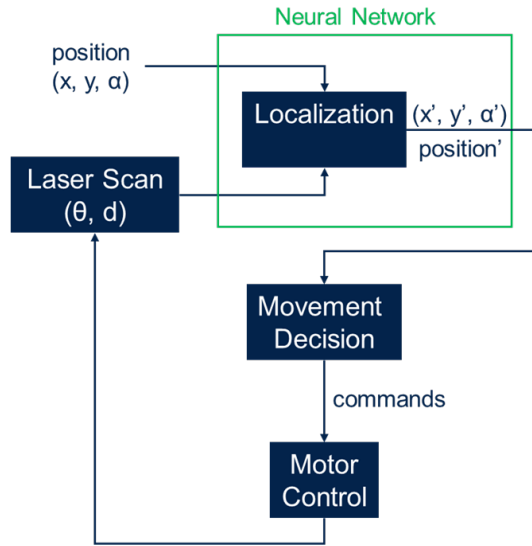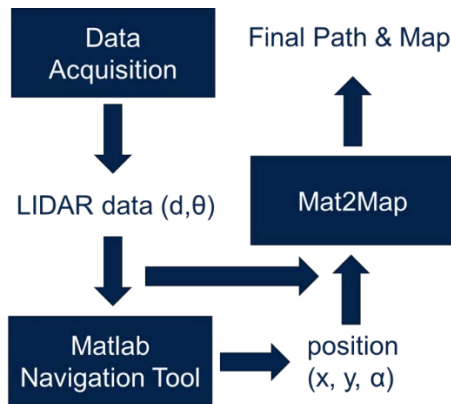

Figure 1: Navigation system.

Figure 2: Ground truth generation schema.

2.1. Dataset Generation

The main challenge in deep learning approaches is the acquisition of large amounts of data, to allow the neural network to work fine in any real scenario. Our generated dataset consists in about 51,000 samples, which we tested to be enough in our experiments for training the proposed neural network. Each sample is composed by two subsequent scans acquired by Lidar. Each scan is expressed by multiple couples (d, θ), that is the distance and the angle from the nearest found object.

Data acquisition cannot be used as it is. We instead need to encode acquired data into a panoramic like depth image, so we need a sort of binarization of each scan before to be paired with the

following one. Data binarization is inspired by a previous work [21] to encode laser scans into a 1D vector. To obtain a 1D vector, in each scan, all distances are binned into angle bins, according to the chosen angle resolution. In this way, we stored all depth values inside a 1D vector, where all the possible depths are represented (from 0° to 360°). As soon as two subsequent scans are binarized, we can couple them to be used as neural network input.

Laser range scanners usually give (scan by scan) distances from the identified nearest object for constant angles, so binarization is simple, because we have a fixed number of possible angles to be considered into the 1D vector. Instead, in the chosen laser range scanner A2 RPLidar in each scan the angles can vary, from 0° to 360°. For this reason, it is not possible to fix we the angle resolution as in previous works, e.g., in [22] the authors use 0.10° and in [21] the authors use 0.25°, so a preliminary investigation to find optimal angle resolution is needed. In this research, we tried at the same time to maximize array density and to minimize collisions per frame. Array density is for each scan the number of non-zero value bins, while collisions per frame is the total number of data ranges which are in the same bin.

Table 1 shows the impact of the chosen β (angle resolution) on N (total number of bins) and then on mean collisions per frame and mean array density. Since A2 RPLidar have got a 360 degrees view, the laser data is separated into β degree bins, for a total amount of N=360°/β° bins. Of course, increasing β (and then decreasing the total number of N bins) the array density increases and of course collisions per frame will become bigger. In our experiments, we tried different angle resolutions β to make at the end the proper decision about which configuration to use in the proposed neural network.

As indicated in Table 1, particularly at higher angle resolutions β, collision is an important aspect to solve to guarantee the neural network to work property. In [22] authors chose to take the mean of all distances are in the same bin, probably because in their experiments the collision occurred rarely and distances at the same degree was similar. In our experiments, we consider two main aspects: laser range scanner is more precise for lower distances and average of two different distances at the same degree can introduce false objects distances. For these reasons, we choose to take the minimum distance (instead of average distance) for distances falling in the same degree.

Table 1: Dataset Binarization

β	N	Collisions per frame	Array Density
0.10	3600	0.05	9%
0.25	1440	0.13	22%
0.50	720	0.48	46%
1.0	360	2.54	89%

2.2. Neural Network

The network we propose in this work, from a consecutive pair of two scans (s_{t-1}, s_t) done by the chosen Lidar, obtains the robot displacement between them, trying the estimation of their relative pose transformation:

$$T = [\Delta x, \Delta y, \Delta \alpha] \qquad (1)$$

where Δx and Δy are respectively the horizontal and vertical translations and $\Delta\alpha$ is the rotation angle between the two scans (s_{t-1}, s_t). We can only estimate the displacement of the robot in two dimensions, since we choose to use just a two dimension sensor.

The final objective of the neural network is to learn the unknown function g(), which, at time t, maps (s_{t-1}, s_t) to the pose transformation T :

$$T_t = g(s_{t-1}, s_t) \qquad (2)$$

In the training step the unknown function g() is learned. Moreover, thanks to the accumulation of the estimated local poses from the starting of the process up to time t, we obtain the robot global position at time t. The chosen loss functions are: mean absolute error (MAE) and mean square error (MSE), which are commonly used in deep learning regression problems.

The strategy we tried to implement here is to fit our regression problem with standard deep learning 2D image matching problems. The difference is that, in our case, instead of 2D images, we have obtained (thanks to dataset binarization) 1D panoramic depth images. In this way, as in the case of images, we can use consecutive CNN to extract spatial features obtained by the sensor in the tested conditions. After that, additional fully connected layers (dense layers) allow to the neural network to understand patterns within extracted spatial features to provide the matching and then the current robot position estimation in the unknown environment.
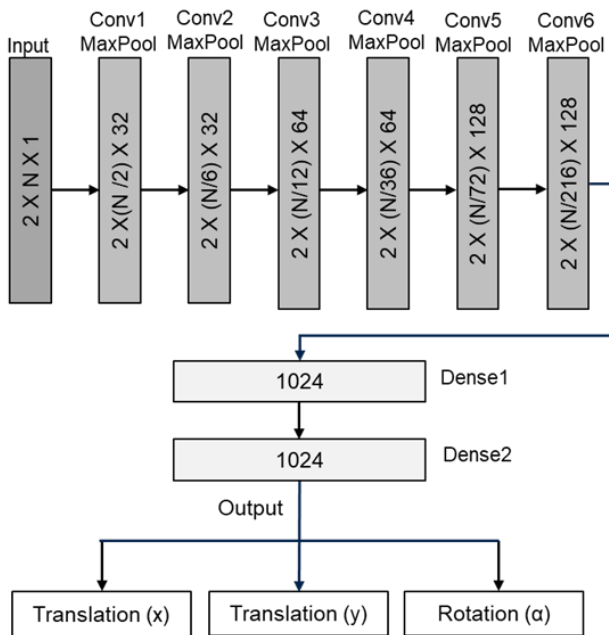

Figure 3: Proposed deep learning approach.

Table 2: Parameters of Neural Network

N	NN Model	NN Parameters
3600	CNN+LSTM	21,078,563
3600	CNN+Dense	5,343,779
1440	CNN+LSTM	15,835,683
1440	CNN+Dense	2,722,339

720	CNN+LSTM	14,262,819
720	CNN+Dense	1,935,907
360	CNN+LSTM	13,214,243
360	CNN+Dense	1,411,619

The proposed neural network is shown in Figure 3. The suggested neural network has been inspired by the one indicated in [22], but some differences should be evidenced. Firstly, we make lots of trials to obtain a low complexity neural network, without to lose in quality performance. Moreover, we used max pool layers (instead of average pool layers) reducing complexity and extracting only most important features spatially distributed. Better results were obtained using max pooling, since as aforementioned (see Table 1), the binarized laser scans tends to be sparse, when the angle resolutions β tends to increase. For the same reason, we also eliminate the stride parameter of the various neural network convolutional filters, which can eliminate useful information for sparse binarized laser scans. To maintain the same dimension, instead of using stride parameter in the convolutional filters, we used other max pooling before applying these filters.

As aforementioned, we experimented different neural network configurations, in particular varying:

- β (angle resolution) and then N (number of bins), to define the correct dimension of input data;
- last two network layers, trying both dense layers [21] and long short term memory (LSTM) layers [22].

The calculation of the NN parameters is shown in Table 2. This number is impacted by N (number of bins) and by the neural network model used. As indicated, when N is increased and when LSTM layers are used, the total number of NN parameters and then the time to be executed will increase too. It is to note that the changes made compared to [22], that is max pool in replacement of strides and average pool, do not impact the overall NN parameters.

2.3. Training

For the sake of clearness, now we indicate in detail the input and ground thru output of the proposed neural network. Input for the network is composed by composed by a set of couples of consecutive Lidar scans (θ, d). These scans are preprocessed allowing to binarize them into sets of 2 X N matrixes, as indicated in Section 2.1. In these matrixes, N depends on β (angle resolution), like indicated in Table 1.

About the reference (ground thru) output of the proposed NN, it is composed by a set of vectors $T = [\Delta x, \Delta y, \Delta\alpha]$, which are the ground thru positions of the moving object, obtained by applying Navigation Tool (MATLAB) to the input Lidar scans (θ, d).

The various experiments were performed in a Python environment with the use of TensorFlow framework and Keras wrapper. Moreover, a workstation was used for training execution, that is a Xeon ES-2630 (Intel) octacore machine with 62 GB of RAM and a Titan X (Pascal) GPU produced by Nvidia. The chosen GPU has got 12 GB of RAM and it is equipped by 3584 CUDA cores, to allow lots of parallelization in training step for faster execution.

The details for training step are the following: 0.0001 function cost minimization learning rate, 500 epochs for training the neural network, 32 batch size and Adam training optimizer used. We tried other kinds of training optimizers, but we did not notice any significant difference in regression performances.

3. Experimental Results

Lots of tests have been executed with different neural network configurations, as indicated in Table 2, and with different indoor environments. As expected, it is important to note that classical Convolutional Neural Networks (CNNs) work better in the case of dense input datasets. For this reason, we used max pooling instead of average pooling in final tests. For the same reason, even if, at the beginning, we tried in our experiments all the different neural network configurations, final research was focused on β (angle resolution) set to one degree and then N (number of bins) set to 360. This choice also allows us to reduce the total neural network parameters and the overall complexity.

Figures in this Section are representative of a particular testing to underline how (depending on scaling applied) the neural network tends to converge (generalizing the regression problem) or not and to underline how the final suggested neural network fits our needs (lightness and precision). In particular, in the X axis the evolution of the network in various epochs (trials) is represented and in the Y axis the loss in precision is represented (first trials in mean square error, after we used mean absolute error). When train and validation curves are similar with low loss the neural network works properly, while when they are different a problem occurs. In this Section we try to explain a particular reason (scale) of this problem.

Table 3 shows the results obtained using input distances and output positions expressed in millimeters. Results are really bad: the proposed network seems to make a sufficient regression work for training set, but for validation and test set it does not reach good performance at all. In general, as expected, max pooling strategy reaches better performance than average pooling and reducing the angle resolution and then the N dimension of the input binarized scans we obtain better loss values (MSE).

To better understand the evolution of this first experiment done on the proposed neural network epoch by epoch, the first 150 epochs are displayed in Figure 4. This graph is referred to the case N = 1440 with max pooling (train loss = 2.48 MSE; validation loss = 216 MSE; test loss = 233 MSE), but similar considerations can be done on the other different configurations. It is easy to note that while the curve for train decreases, the curve for validation is flat, so the proposed network, in this case, is not able to solve the overexposed regression problem and to generalize it.

Table 3: Test Results (MSE) with Input Dataset (Millimeters)

N	Model	Train Loss	Validation Loss	Test Loss
3600	AvePool	2.78	250	258
3600	MaxPool	2.75	248	250
1440	AvePool	2.30	228	236
1440	MaxPool	2.48	216	233

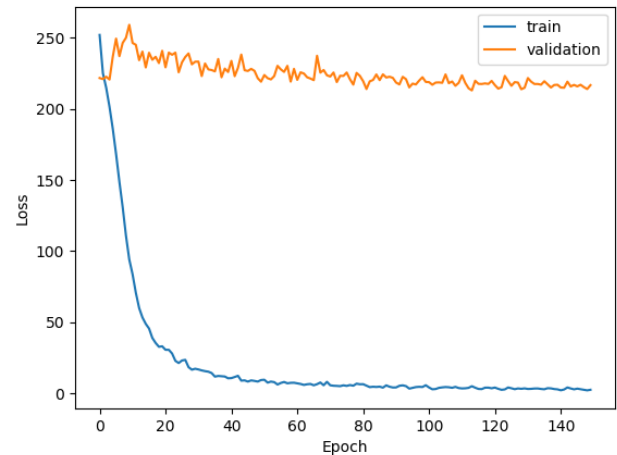| 720	AvePool	2.36	221	240
720	MaxPool	3.27	220	238
360	AvePool	7.54	220	235
360	MaxPool	7.02	218	230

Figure 4. Results obtained with input dataset (millimeter).

In the used approach, the main issue is the scale discrepancy in the input variables (x and y expressed in millimeters and α expressed in degree), which often increases the difficulty to correctly model the neural network to solve the regression problem. A common trick used in these cases is to pre-process the input variables before they are fed to the neural network [26]. In the same way, also the outputs of the network (d expressed in millimeters and θ expressed in degree) should be processed to obtain the correct output values.

A commonly used pre-processing step is just a simple linear scaling of network variables [26], so we just changed the distance measure passing from millimeter to centimeter and the related results are shown in Table 4. As indicated, better results are obtained compared to the first tentative with input distances and output positions expressed in millimeters. Even if results are better, we must again to note that they are not good enough: again, the proposed network seems to make a sufficient regression work for training set, but for validation and test set it does not reach similar good performance. Moreover, as expected, max pooling strategy reaches better performance than average pooling and reducing the angle resolution and then the N dimension of the input binarized scans we obtain better loss values (MSE).

Table 4: Test Results (MSE) with Input Dataset (Centimeters)

N	Model	Train Loss	Validation Loss	Test Loss
3600	AvePool	0.09	4.12	3.90
3600	MaxPool	0.03	3.98	3.83
1440	AvePool	0.28	3.97	3.77

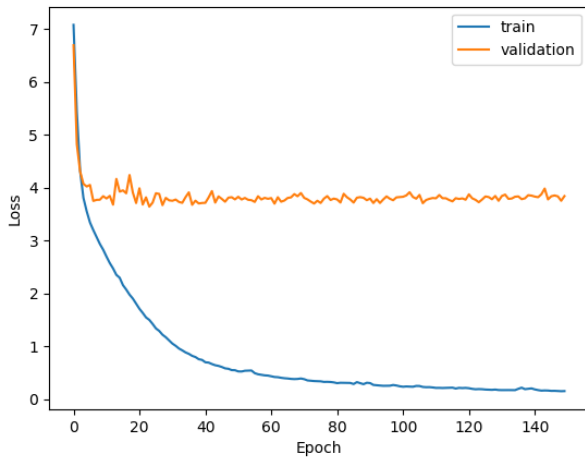| 1440	MaxPool	0.11	3.87	3.68
720	AvePool	0.18	3.86	3.76
720	MaxPool	0.13	3.84	3.67
360	AvePool	0.28	3.75	3.70
360	MaxPool	0.19	3.67	3.64

Figure 5. Results obtained with input dataset (centimeter).

To better understand the evolution of the second experiment done on the proposed neural network epoch by epoch, the first 150 epochs are displayed in Figure 5. To allow a visual comparison with Figure 4, also this graph is referred to the case N = 1440 with max pooling (train loss = 0.13 MSE; validation loss = 3.84 MSE; test loss = 3.683 MSE). Of course, similar considerations can be done on the other different configurations. It is important to put into evidence that MSE is not a linear measure, but quadratic. This is the reason why the loss is drastically reduced in comparison with the previous experiment. Moreover, the train curve correctly decreases, as in previous case, while in the current test the validation curve is not completely flat, but it starts to go down. As first experiment, in the current test the proposed network cannot model the neural network to generalize and correctly solve the regression problem, but improvements noticed give us an important hint to work with: to obtain the best regression results, firstly we must find the correct rescaling to apply to input and output data measures (distance and angle).

At this point, we make some rescaling experiments. To maintain similar scale in both input dataset and output variables, we tried to scale translation data by one thousand (in this way we use meters, instead of millimeters as measure) and rotation data by one hundred. This rescaling configuration gives us best results. Furthermore, we decide to make use of Mean Absolute Error (MAE), instead of Mean Squared Error (MSE). In this way, we obtained similar loss curves, but with results simpler to understand and comment, because MAE is a linear measure, while MSE is quadratic. Figure 6 shows that good results are finally reached (train loss: 0.011 MAE; validation loss: 0.011 MAE; test loss: 0.010 MAE), using the new scaling factors to be applied to input and output data with a very simple configuration (only 1,411,619 network parameters). Like other experiments, the train curve correctly decreases, but this time the validation curve also

decreases with similar slope. This indicates that finally we reached our main goal: the neural network can now correctly generalize and solve the proposed regression problem.
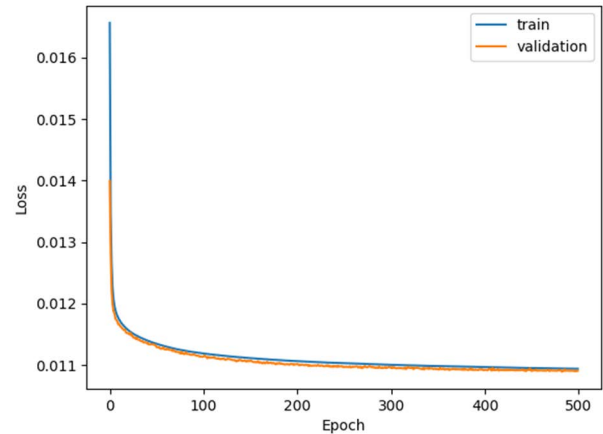

Figure 6: Results obtained with input dataset (meter) with 1,411,619 parameters

In our research, we tried to further reduce the parameters in the proposed neural network for lighter solutions, to be implemented in microcontrollers with low resources and in particular with memory (RAM and FLASH). To reach this goal, we tried to reduce the elements in the last two layers (dense layers). After several trials, we realized that the results are still good also deleting last two layers, obtaining a very low neural network parameters (96,547). Figure 7 shows that in this experiment, even if at the beginning loss values are higher than previous test because the neural network is simpler, at the end of the epochs, train and validation curves are very similar and this network reaches similar results. In fact, Table 5 shows that numerical results between the prosed full neural network with 1,411,619 parameters and the proposed reduced neural network with 96,547 reaches similar performances in terms of MAE regression loss.
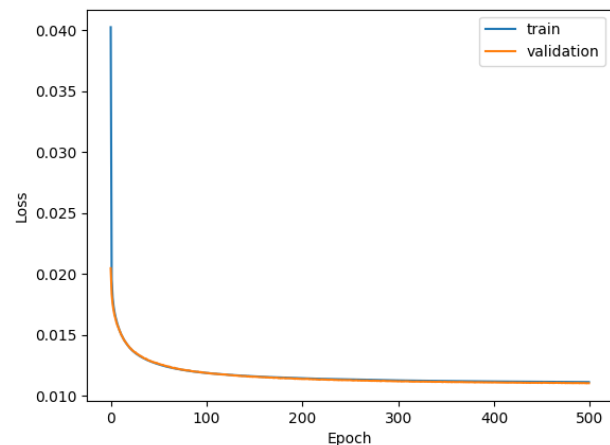

Figure 7: Results obtained with input dataset (meter) with 96,547 parameters

Table 5: Test Results (MAE) with Input Dataset (Meters) (N = 360)

P	Model	Train Loss	Validation Loss	Test Loss
1.4M	MaxPool	0.011	0.011	0.010
96K	MaxPool	0.011	0.011	0.010

Table 6: Test Results (MAE) For (x,y) Only (N=360)

P	Model	Train Loss	Validation Loss	Test Loss
1.4M	MaxPool	0.009	0.009	0.010
1.4M	MaxPool Augmentation	0.009	0.009	0.010
96K	MaxPool	0.010	0.010	0.010
96K	MaxPool Augmentation	0.010	0.009	0.009

Further investigations to increase performance of the proposed neural network have also been carried out. In particular, different normalization techniques (scaling) have been tested in our testing environment, like MinMaxScaler() and StandardScaler(), but no substantial improvement has been obtained. The same behavior (no improvement) has been also noticed changing the function loss (Euclidean distance) and extending the neural network using LSTM layers in replacement of Dense layers.

At this point, we tried to reduce the problem limiting the output of the neural network to only spatial components, i.e., to a vector $T' = [\Delta x, \Delta y]$. Moreover, in this simplified problem version, we also tried an intensive data augmentation, obtained with different strategies. In particular, we take input scans in reverse order, just odds and even scans and finally odds and even scans in reverse order. In this way, we obtained about 204,000 samples, that is about four times the original dataset.

Table 6 summarizes the results obtained by the proposed neural network in the case of simplified problem. Comparing the results with the ones in Table 5, we can note a negligible improvement in train and validation loss, but not in the test loss, so results are almost the same. Even with the intensive augmentation, we obtain a slight but not significant improvement for the simpler neural network with 96,547 parameters.

4. Conclusions

The problem of estimating the moving robot localization, with the only usage of data coming from 2D laser scanner, has been addressed by this paper using a simple deep learning approach. For the dataset used, in the final version composed by 204,000 samples with only 96,547 parameters, the proposed neural network achieved good precision performance in terms of Mean Absolute Error (MAE): one centimeter in translation and one degree in rotation. We also tried to increase performance of the proposed network using different strategies and parameters, but no significant improvements have been obtained.

Even if the encouraging results presented here are almost comparable with classical localization estimation approaches, at moment the approaches based on deep learning could be used in replacement of other state-of-the-art algorithms, since the latter ones are more flexible and can potentially reach better performances. Anyway, the overexposed approach remains a good proof of concept and it can be used for future explorations.

Furthermore, the proposed neural network can be used as interesting complement or can be integrate in classic localization methods, since it works in real-time, taking less than 130 µs to elaborate each estimation on Titan X (Pascal) GPU produced by Nvidia.

Conflict of Interest

None of the authors have any kind of conflict of interest related to the publication of the proposed research.

References

[1] G. Spampinato, A. Bruna, I. Guarneri, D. Giacalone, "Deep Learning Localization with 2D Range Scanner," International Conference on Automation, Robotics and Applications (ICARA), 206-210, 2021, DOI: 10.1109/ICARA51699.2021.9376424.

[2] G. Spampinato, A. Bruna, D. Giacalone, G. Messina, "Low Cost Point to Point Navigation System," International Conference on Automation, Robotics and Applications (ICARA), 195-199, 2021, DOI: 10.1109/ICARA51699.2021.9376545.

[3] T. Sattler, B. Leibe, L. Kobbelt, "Efficient & Effective Prioritized Matching for Large-Scale Image-Based Localization," IEEE Transaction on Pattern Analysis and Machine Intelligence, 1744-1756, 2017, DOI: 10.1109/TPAMI.2016.2611662.

[4] R. Mur-Artal, J. M. M. Montiel, J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," IEEE Transactions on Robotics, 31(5), 1147-1163, 2015, DOI: 10.1109/TRO.2015.2463671.

[5] C. Forster, M. Pizzoli, D. Scaramuzza. "SVO: Fast semi-direct monocular visual odometry," IEEE International Conference on Robotics and Automation (ICRA), 15-22, 2014, DOI: 10.1109/ICRA.2014.6906584.

[6] B. Zeisl, T. Sattler, M. Pollefeys, "Camera Pose Voting for Large-Scale Image-Based Localization," Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2704-2712, 2015, DOI: 10.1109/ICCV.2015.310.

[7] H. Kim, D. Lee, T. Oh, H. Myung, "A Probabilistic Feature Map-Based Localization System Using a Monocular Camera," Sensors, **15**(9), 21636-21659, 2015, DOI: 10.3390/s150921636.

[8] E. Olson, "M3rsm: Many-to-many multi-resolution scan matching," IEEE International Conference on Robotics and Automation (ICRA), 5815-5821, 2015, DOI: 10.1109/ICRA.2015.7140013.

[9] G. D. Tipaldi, K. O. Arras, "Flirt-interest regions for 2d range data," IEEE International Conference on Robotics and Automation (ICRA), 3619-3622, 2010, DOI: 10.1109/ROBOT.2010.5509864.

[10] S.I. Roumeliotis, G. A. Bekey, W. Burgard, S. Thrun, "Bayesian estimation and Kalman filtering: A unified framework for mobile robot localization," Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2985-2992, 2000, DOI: 10.1109/ROBOT.2000.846481.

[11] S. Park, K. S. Roh, "Coarse-to-Fine Localization for a Mobile Robot Based on Place Learning With a 2-D Range Scan," IEEE Transactions on Robotics, 528-544, 2016, DOI: 10.1109/TRO.2016.2544301.

[12] P. Besl, H.D. McKay, "Method for registration of 3-D shapes," Sensor Fusion IV: Control Paradigms and Data Structures, International Society for Optics and Photonics, 239-256, 1992, DOI: 10.1109/34.121791.

[13] F. Pomerleau, F. Colas, R. Siegwart, S. Magnenat, "Comparing icp variants on real-world data sets," Autonomous Robots, **34**(3), 133–148, 2013, DOI: 10.1007/s10514-013-9327-2.

[14] C. Friedman, I. Chopra, O. Rand, "Perimeter-based polar scan matching (PB-PSM) for 2D laser odometry," Journal of Intelligent and Robotic Systems: Theory and Applications, **80**(2), 231–254, 2015, DOI: 10.1007/s10846-014-0158-y.

[15] J. Zhang, S. Singh, "LOAM: Lidar Odometry and Mapping in Real-time," Robotics: Science and Systems, 109-111, 2014, DOI: 10.15607/RSS.2014.X.007.

[16] A. Kendall, M. Grimes, R. Cipolla. "Posenet: A convolutional network for real-time 6-dof camera relocalization," Proceedings of the IEEE International Conference on Computer Vision, 2938-2946, 2015, DOI: 10.1109/ICCV.2015.336.

[17] S. Wang, R. Clark, H. Wen, N. Trigoni, "Deepvo: Towards end-to-end visual odometry with deep recurrent convolutional neural networks," IEEE International Conference on Robotics and Automation (ICRA), 2043-2050, 2017, DOI: 10.1109/ICRA.2017.7989236.

[18] R. Li, S. Wang, Z. Long, D. Gu, "Undeepvo: Monocular visual odometry through unsupervised deep learning," IEEE International Conference on

Robotics and Automation (ICRA), 7286-7291, 2018, DOI: 10.1109/ICRA.2018.8461251.

[19] H. M. Cho, H. Jo, S. Lee, E. Kim, "Odometry Estimation via CNN using Sparse LiDAR Data," International Conference on Ubiquitous Robots (UR), 124-127, 2019, DOI: 10.1109/URAI.2019.8768571.

[20] M. Pfeiffer, M. Schaeuble, J. Nieto, R. Siegwart, C. Cadena, "From perception to decision: A data-driven approach to end-toend motion planning for autonomous ground robots," IEEE International Conference on Robotics and Automation (ICRA), 1527-1533, 2017, DOI: 10.1109/ICRA.2017.7989182.

[21] J. Li, H. Zhan, B. M. Chen, I. Reid, G. H. Lee, "Deep learning for 2D scan matching and loop closure," International Conference on Intelligent Robots and Systems (IROS), 763-768, 2017, DOI: 10.1109/IROS.2017.8202236.

[22] M. Valente, C. Joly, A. de La Fortelle, "An LSTM Network for Real-Time Odometry Estimation," IEEE Intelligent Vehicles Symposium (IV), 1434-1440, 2019, DOI: 10.1109/IVS.2019.8814133.

[23] M. Velas, M. Spanel, M. Hradis, A. Herout, "CNN for IMU assisted odometry estimation using velodyne LiDAR," IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), 71-77, 2018, DOI: 10.1109/ICARSC.2018.8374163.

[24] S. Xu, W. Chou, H. Dong, "A Robust Indoor Localization System Integrating Visual Localization Aided by CNN-Based Image Retrieval with Monte Carlo Localization," Sensors, **19**(2), 249, 2019, DOI: 10.3390/s19020249.

[25] W. Hess, D. Kohler, H. Rapp, D. Andor, "Real-time loop closure in 2d lidar slam," IEEE International Conference on Robotics and Automation (ICRA), 1271-1278, 2016, DOI: 10.1109/ICRA.2016.7487258.

[26] C. M. Bishop, "Neural Networks for Pattern Recognition," Oxford University Press, 296-298, 1996, ISBN:978-0198538646.

ASTES

Designing a Model of Consciousness Based on the Findings of Jungian Psychology

Toshiki Watanabe[*,1], Hiroyuki Kameda[2]

[1]*Graduate School of Bionics, Computer and Media Scienses, Tokyo University of Technology, Tokyo, 192-0982, Japan*

[2]*School of Computer Sciense, Tokyo University of Technology, Tokyo, 192-0982, Japan*

ARTICLE INFO	ABSTRACT
	As artificial intelligence (AI) develops, it is expected that humans and AI will become more closely related than now. At the same time, however, the more closely humans and AI are related to each other, the more clearly they will face a moral dilemma, i.e., artificial intelligence will face a moral dilemma. To solve the moral dilemma problem, AI should understand and take into account human values and ethics. From this point of view. we designed a consciousness model based on Jungian post-psychological notes. As a result, we found that in order to implement the model of consciousness on a computer, it is necessary to design it with a structure similar to that of human beings, referring to human structures in various fields.

1. Introduction

This paper is an extension of work originally presented in SCIS-ISIS 2020 [1]

1.1. The social value of artificial intelligence technology

Artificial intelligence (AI) is currently being used in a variety of fields, including medicine, weather, and robotics. It has become an indispensable part of our lives, and its use is still being actively researched.

It is expected to evolve further in the future and will be applied and used in even wider fields than at present. According to an announcement by Japan's Ministry of Internal Affairs and Communications(https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r01/html/nd113230.html), the use of AI is expected to lead to more efficient operations and the development of new products and business models. It is also expected to have the effect of creating innovation by freeing humans from cumbersome tasks and enabling them to focus on creative tasks that only humans can perform. Several research organizations have analyzed the effect of AI on productivity. For example, according to an analysis by Accenture, the labor productivity,which is defined as (results produced or amount of value-added) / (quantity of labor), e.g, number of workers or workers ∗ working hours, in Japan is expected to increase by 34% in 2035 compared to the baseline. The value of AI will become even higher in the future.

1.2. Current Status and Trends of Artificial Intelligence

In recent years, AI has become so prevalent in our society that we can ask so-called voice assistants such as Alexa (Amazon), Siri (Apple), and Google Assistant to tell us today's weather, or help input a destination into car navigation or show us the route on a map. In addition, there is a lot of research that focuses on interaction with humans, such as mentioned in [2] and[3], where robots and androids communicate with humans to heal and support them. In the future, humans and AI will be more closely related to each other.

As humans and AI become more closely involved, some of the decisions that have traditionally been made by humans will be delegated to them. For example, in the area of automated driving, Waymo currently commercializes Level 4 automated driving in the U.S., and Japan has also released a roadmap for fully automated driving by 2030 (https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20200715/2020_roadmap.pdf), while Honda LEGEND (https://www.honda.co.jp/LEGEND/), which is currently Level 3 automated driving, is currently being sold to consumers.

However, with the development of these technologies, AI is facing the problem of moral dilemma. A thought experiment of Massachusetts Institute of Technology's Moral Machine (https://www.moralmachine.net/) exposes problems like the trolley problem to gather human perspectives on how self-driving cars should behave. We believe that this kind of problem is an important issue that must be considered in the future of AI.

[*]Corresponding Author: Toshiki Watanabe, g212003578@edu.teu.ac.jp

1.3. Position of this study

In this study, we hypothesized that if humans were to delegate decision-making to AI in the future, they would be able to make better and more flexible decisions by understanding and taking into account human values, ethics, and local culture and customs than now. We hypothesized that if the conscious and unconscious minds that control self-control could be realized on a computer, it would be able to understand and take into account human values, ethics, and local culture and customs. Self-control is considered to be important for this, and we thought that the relationship between the conscious and unconscious mind, which controls behavior and judgment, makes this possible. Therefore, we decided to model the conscious and unconscious as a decision-making system for AI and reproduce it on a computer for verification.

Therefore, to realize the conscious and unconscious on a computer, we designed a model of consciousness based on the findings of Jungian psychology.

2. Related works

2.1. The treatment of consciousness in the field of psychology

Freud is one of the first persons to discover the unconscious. He likened the human mind to the tip of an iceberg and divides conscious into three layers: the consciousness, the pre-conscious which can be made conscious, and the unconscious which cannot be made conscious, He also stating that repressed desires are pushed into the unconscious.

In his work, the author described the structure of the mind as consisting of the *es*, which is a mass of instincts, the *superego*, which is responsible for the protection of ethics, and the *ego*, which regulates them and guides conclusions to action [4].

The authors did not break down the human being structurally, as his study was centered on the theory of the purpose, which states that the individual mind cannot be divided and acts according to its present purpose [5].

The authors divided human conscious into the *consciousness*, the *personal unconscious*, and the *collective unconscious*.

Consciousness is that in the sense of being aware, and so on., while unconsciousness refers to other things.

"Many of the difficult perceptual computations, such as invariant face recognition or speaker-invariant speech recognition. that ware recently addressed by AI correspond to nonconscious computations in the human brain. [...] Many of these C0 computations have now been captured by AI, particularly using feedforward convolutional neural networks (CNNs)."

The personal unconscious is the consciousness that has been suppressed by experiences and can be viewed as things that we are no longer conscious of, but which we can become aware of if we want to.

The collective unconscious is the unconscious that is common to all human beings and cannot be conscious.[6]

In Jungian psychology, the unconscious accounts for about 90% of the total, with the collective unconscious accounting for a large percentage of the total.

He was also the first person to use the term *complex*, as it is commonly referred to use in an even broader sense than its general scope.

In a recent study on the role of the conscious and unconscious mind the authors in [7],

"(1) Recent findings in social psychology research indicate that many human behaviors, including self-control, depend on automatic processes that do not require consciousness (2) We examine the relationship between conscious and unconscious processes in self-control and propose effective self-control in which the two processes work together. The purpose of this study is to examine the relationship between conscious and unconscious processes in self-control and to propose an effective self-control system in which the two processes work together. The role of conscious processes seems to be to form new coalitions through abstract mental editing. […]One of the important functions of conscious processes was also to make unconscious self-control more effective. Thus, it is only by assuming the collaboration of the two that the essential functions of unconscious and conscious processes can be seen."

Through experiments, he came closer to elucidating the roles and functions of the conscious and unconscious.

2.2. Treatment of consciousness in the medical field

In the field of medicine, consciousness is often treated as a level of conscious or arousal. According to the website(https://bsd.neuroinf.jp/wiki/%E6%84%8F%E8%AD%98) ,it says

"The level of consciousness or arousal is the state of awareness. The level of consciousness is the highest when you are awake and clear-headed, lower when you are sleepy and fuzzy-headed, or when you are asleep and not dreaming, or when you are under deep anesthesia."

Typical scales for the level of consciousness include the Grady Coma Scale. In [8] the authors, the authors say

"The grade I patient is only slightly confused. The grade II patient requires a light pain stimulus (such as a sharp pin tapped lightly over the chest wall) for appropriate arousal, or may be combative or belligerent. The grade III patient is comatose but will ward off deeply painful stimuli such as sternal pressure or nipple twist with an appropriate response. The grade IV patient reacts inappropriately with either decorticate or decerebrate posturing to such deeply painful stimuli, and the grade V patient remains flaccid when similarly stimulated."

Consciousness in neuroscience and neurology deals with both psychological and medical consciousness. In [9] the authors discussed how to make computers conscious from the perspective of neuroscience. They argue that

He also claims that In other words, much of what AI does today, such as face and voice recognition, corresponds to unconscious computations in the human brain. They also claim

In other words, he claimed that if the Unconscious processing (C0) can be equipped with Global availability (C1) and Self-monitoring (C2), the machine will behave as if it has conscious.

2.3. Related studies in the field of engineering

In related research in the field of engineering, models of the mind and emotion have been actively studied.

" We contend that a machine endowed with C1 and C2 would behave as if it were conscious—for instance,"

In [10], the author builds a model that is aware of the whole picture of the mind and is hierarchical at five levels, from raw data to complex concepts, and can integrate low-level signal processing with high-level symbolic processing, which AI is good at. They built a model called the "Aesopworld" and implemented it in the Aesopworld project.

In [11], the author built a model of the mind based on the passive consciousness hypothesis, dividing it into two major parts: the conscious system and the unconscious system, assuming the passivity and illusion of consciousness. Thereby he proposed a method of conceiving the mind that has clear boundaries and is feasible, unlike conventional cognitive models.

In [2], the robot used cold reading and other techniques in communication to read the minds of customers and use them in sales. In [3] the authors adopt a model of personality to "heal" people.

These can be said to be equipped with unique models to communicate with humans.

In [12] the authors analyzed the trends in incorporating emotions into AI and what people need to be aware of when building models. People would like to create a foothold for clarifying the nature of emotions, which have been considered ambiguous, by incorporating knowledge about emotions from different fields such as psychology, neurology, and cognitive science, and other emotion research using a constructive approach. The reason why people can't reach the essence of emotions is probably that we can't abolish naive psychology and that the emotion research in various fields is conducted in different ways. He pointed out that it is important to abolish naive psychology to build a model of emotions and the mind.

2.4. Position of this study

In this research, we focus on being able to reproduce human consciousness on a computer. This is because if we can reproduce human consciousness on a computer, we will be able to systematically explain what people think and how they behave. And we believe that this will be useful not only in the field of AI but also in various other fields such as psychology and brain science.

3. Consciousness model in this study

In this study, we modeled consciousness in psychology, while including ethics, culture, and the human thinking and judgment,

and aimed to construct a model of consciousness as a decision-making system, so that AI with the model can understand and take into account their values, ethics, and local culture and customs.

Psychology remains at a conceptual level and is difficult to implement directly in a computer, but it most clearly represents the role of consciousness and human behavior. On the other hand, research in the field of engineering has concrete models and roles to be implemented in computers, but they do not represent human behavior. In addition, the role of consciousness and human behavior was clear in the field of brain science, although the level of conscious in the medical field is different in meaning from the consciousness dealt with in this study. However, we still could not find a model that could be immediately implemented in a computer.

Compared to previous studies, my research is based on psychology, which allows me to design a model that eliminates naive psychology as much as possible, and by dividing the functions of consciousness, I can create a more detailed and clear model for designing programs.

The psychology we refer to when modeling is Jungian psychology, which divides the flow of thought into the consciousness, personal unconscious, and the collective unconscious, and explains the role of them in particular detail.

In [9] the authors found that it is effective to divide the functions of consciousness, so our model in this study is also designed by dividing the functions to some extent to clarify the roles of each process.

3.1. Definition of our Consciousness Model

We define the consciousness model as a decision-making function that selects action based on ethics, culture, and values accumulated through personal experience.

3.2. Requirements for an Consciousness model

The model of consciousness will be set to meet the following requirements.

- To fulfill the roles of consciousness, personal unconscious, and collective unconscious in Jungian psychology.

- Input and output are based on the human body.

- The model has memory retention and storage functions.

- The model can be reproduced as a program on a computer.

3.3. Pre-design of Consciousness Model

Consciousness in Jungian psychology is regarded as a system, and the functions of "consciousness," "personal unconscious," and "collective unconscious" are assumed to be interrelated and to make decisions by processing some data.

The procedure for constructing a model of consciousness is as follows: First, we clarify the functions of conscious in Jungian psychology. Next, we consider the connections among the functions of consciousness, and the roles these connections play in each other. We define the functions of consciousness as shown in Table1.

Table 1: Functions of Consciousness

Main components	Functions
Consciousness	Able to retrieve information from the unconscious that is related to the recognized information.
	Able to elicit needs and desires
	Able to repress needs and desires
Personal Unconscious	Complex (a desire or want that has been strongly suppressed by experience, education, and so on.)
	Suppressed needs and desires (currently unconscious needs and desires)
Collective unconscious	Human instincts (maternal instincts, sexual desires, and other archetypes shared by all human beings)

Assuming that the consciousness, personal unconscious, and collective unconscious have functions as shown in Table 1, there should be the interaction of data transfer and processing amang main components.

Therefore, as a result of considering the data transfer and processing between the functions of consciousness as a decision-making system, Figure 1 was obtained.
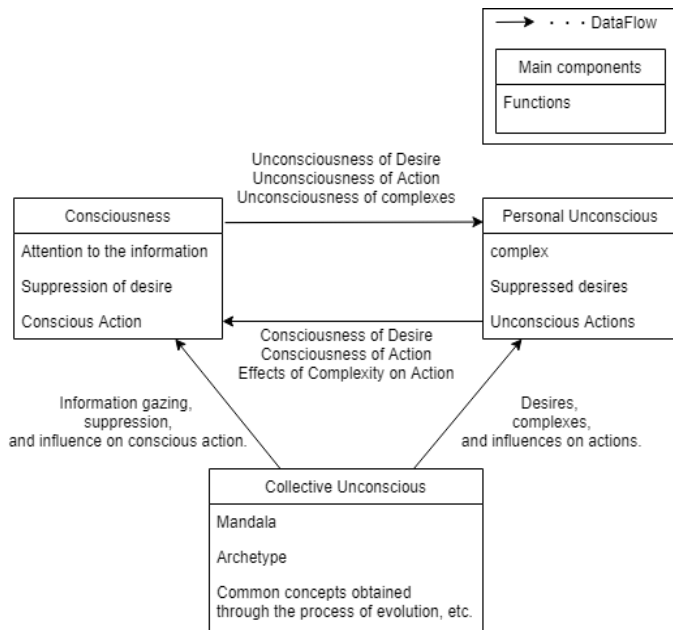
Figure 1: Conceptual diagram of consciousness

Figure 1 shows the connections between consciousness, the personal unconscious, and the collective unconscious, but this is too abstract to be reproduced as a program on a computer.

There must be connections between processes in the consciousness mind, connections between processes in the unconscious mind, and connections between small processes between the conscious mind and the unconscious mind.

Therefore, we further extended Figure 1 to consider the connections between each process and created Figure 2.

3.4. Overview of the Consciousness Model

By extending the conceptual diagram of consciousness in Figure 1, we obtained the following Figure 2. Using this figure as a model of consciousness, we can consider the operation of each process.
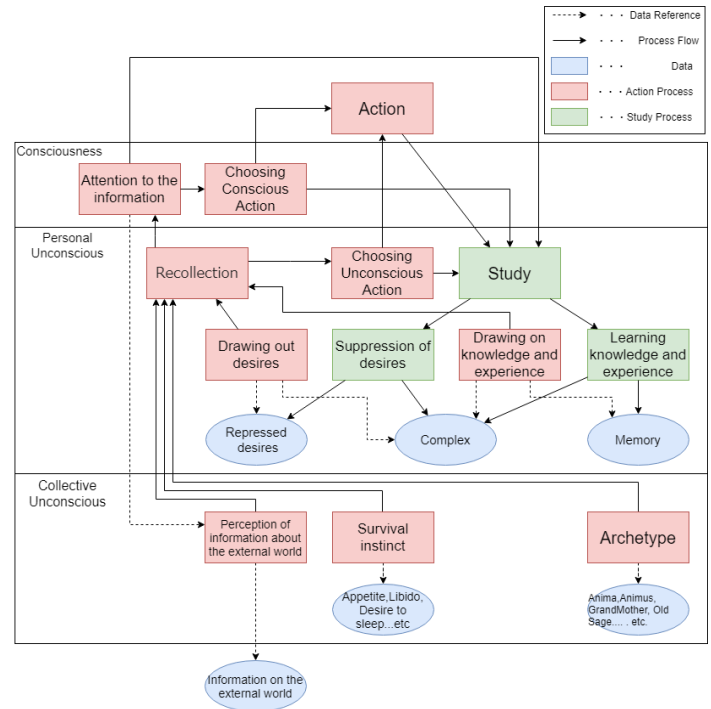

Figure 2: Consciousness Model

The process shown in Figure 2 can be divided into three major parts: "data," "Action process," and "Study process". For data, these include values obtained from sensors and episodes. The arrows between data and process represent data referrence.

The action process is the process of taking action based on the data, and the study process is the process of retaining and storing the data as experiences and desires in the individual's memory after taking action.

For example, when a sensor detects hunger, the survival instinct perceives hunger and recalls associates information about food and how to eat it. Then, priming memory and other memories are sent to unconscious behavioral processes to select actions. Episodic memory and other memories are sent to information gazing to link the perception of external information with the information in episodic memory. After that, the conscious action selection process selects the action, and the action process selects the final action from the conscious and unconscious actions.

The focused information and selected behaviors are sent to the learning process, where the information is divided into the desire suppression process and knowledge-experience learning, and is prepared in the form of each memory before being stored.

The followings are the explanations of the components in Figure 2.

- Action

The actual action is taken by the system or the signal that directs it. The process of acting is selected by conscious and unconscious action selection.

- Information on the external world

Information that is not internal to the system, such as the five senses in humans, or sensors in machines. And it is unprocessed one, such as what you see with your camera or eyes, or what you feel.

- Attention to the information

To pay attention to objects and phenomena that are highly related to the objects recognized through the recognition of information about the external world, and to perform a deeper recall of these objects and phenomena.

- Choosing Conscious Action

To select the higher rewarding action required by the information you have been paying attention to.

- Choosing Unconscious Action.

To select highly learned behaviors picked up by recall.

- Recollection

The process of retrieving retained information in the right place and at the right time. Memory recall is the process of recalling a specific memory from among the memories stored in the brain. It induces memory destabilization, refixation, extinction learning, and memory association. In terms of the movement of the process, it refers to the action of searching for knowledge and experience to achieve desires.

- Drawing out desires

To search for and elicit relevant desires.

- Drawing on knowledge and experience

To search for and elicit relevant knowledge.

- Study

The act of filling in the blanks of knowledge. It is close to the meaning of the word in psychology and pedagogy and is distinguished from learning. It refers to the overall act of learning, such as extracting, categorizing, and storing features from the information obtained at that time as a result of an action, and adjusting each parameter. Knowledge is all that is gained through experience and includes all information stored in the unconscious. The rewards received from actions and related information are parameterized and recorded into knowledge, experiences, wants desires, and complexes.

- Suppression of desires

To repress what could not be done or failed in the drawn-out desire-lust.

- Learning knowledge and experience

To record the results and rewards of actions taken, and to store a series of actions as episodic memory.

- Repressed desires

A set of data repressed by desire.

- Memory

It is classified into sensory memory, short-term memory, and long-term memory, and it records information about what happens to us and what we feel, both temporarily and long-term. Sensory memory has the shortest storage time and is present in each sense. Only the information to which we pay attention becomes short-term memory. Short-term memory lasts only a few tens of seconds and is limited in the amount of information that can be stored at one time. Long-term memory is a part of short-term memory that is retained and becomes long-term memory through memory fixation. In this study, data is stored by learning, and organized by study.

- Perception of information about the external world

To recognize the type of object and distance from the external world information.

- Survival instinct

It elicits life-sustaining desire parameters such as sexual desire, appetite, and desire to sleep. The survival instinct is considered to be part of the collective unconscious.

It is the minimum means of survival stored as data.

- Archetype

To draws information about repression related to external world information and desires from stored data, such as anima and animus. Instincts and concepts cultivated in the group are individually shared data, not necessarily shared things in common.

Of the collective unconscious, the history and common sense cultivated by humankind fall into this category.

- Complex

In this study, we believe that complexes can be treated as a cohesive group of closely related memories and impulses, and as a collection of related memories. Complexes are said to be complexes of impulses, memories, and ideas that are intertwined by some emotion. By using these complexes, we will be able to handle memories that are highly relevant to recall. The parameters stored in the memory are treated as data that is associated with a distribution or set.

- Appetite, Libido, Desire to sleep...and so on.

Internal information such as a sensation of human internal organs, mental state, drowsiness, and so on., and remaining battery power and continuous operation time of the robot.

- Anima, Animus, Great Mother, Old Wise Man...and so on.

Conceptual things that humanity has in common, such as a sense of fear.

4. Discussion and further works

4.1. Discussion

The assumed environment and data are built on the premise that the input is received from the sensors installed in the computer, and the action is derived from the recall from this model.

In the future, when we implement this consciousness model, we need to consider various methods for each process.

Therefore, to clarify the data flow of this model and to make it easier to implement, we created a DFD(Data Flow Diagram) that is supposed to be installed in a robot or a simulator, and as a result, we obtained the following Figure 3.
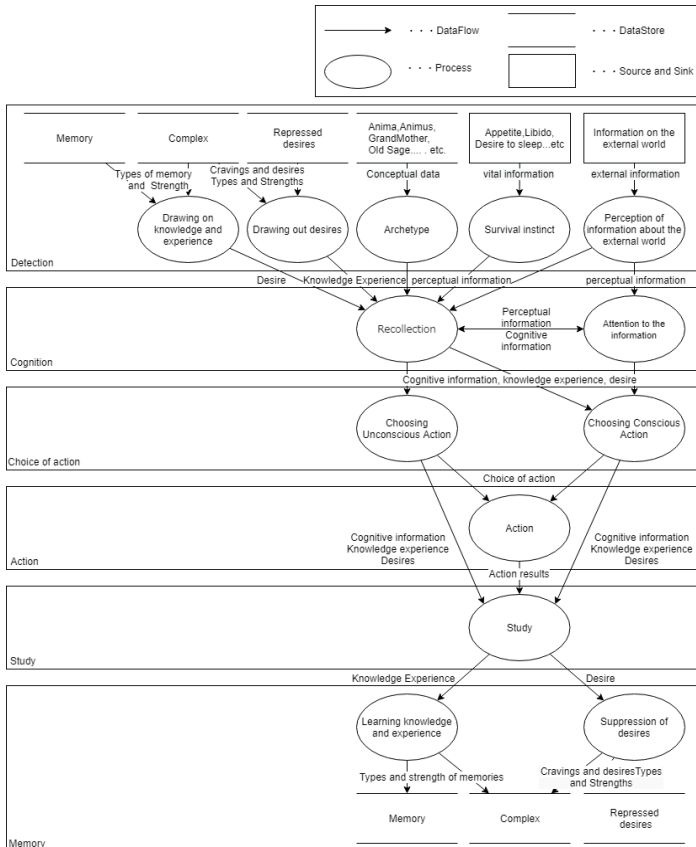

Figure 3: Diagram of consciousness DFD

Figure 3 shows the DFD version of the consciousness model in Figure 2. This DFD indicates that data will be processed from the top and passed to the next process. The arrows represent the direction in which the data is passed, and the names between the arrows represent the data.

Figure 3 shows that the flow of data is roughly in the following order: detection, data retrieval, action selection, action, study, and memory. This is roughly in line with the order of the process of detection-cognition-judgment-action, which is called the basic human thought process (https://ipsj-is.jp/isdic/1070/). We believe that this model can reproduce the basic thought process of human beings and still meet the requirements of the consciousness model defined in this study.

To implement the model on the agent in Unity for model validation, we will create a class diagram from Figure 2 for object-oriented implementation. Based on the obtained diagram, we plan to code and implement the model with the help of other researchers.

4.2. Further works

It will be necessary to implement the consciousness model in computers and robots to prove whether the model can function as

a decision-making program and make decisions based on an understanding of human ethics and values.

When we implement the consciousness model in the future, we need to be especially careful about the methods we use for each process, because some methods may cause unintended behavior. For example, the data structure of memory should be based on the structure of memory used in psychology and brain science, and the implementation should be based on human evidence.

In the learning process, it is also necessary to consider specific implementation methods on computers, such as the learning method and how to process the data to be handled.

5. Conclusion

In this study, we have developed a model of consciousness as a decision-making system based on Jungian psychology for implementation on computers and robots.

In the future, various issues remain to be solved in implementing this model, but we believe that solving them one by one will not only solve the ethical problems of AI but also lead to a better understanding of human thinking and brain mechanisms.

References

[1] T. Watanabe, H. Kameda, "Proposal of a computer-reproducible model of consciousness," in 2020 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (SCIS-ISIS), IEEE: 1–4, 2020, doi:10.1109/SCISISIS50064.2020.9322763.

[2] M. Watanabe, K. Ogawa, H. Ishiguro, "Minami-chan: Application and Verification of Androids to Society Through Selling," **57**(4), 1251–1261, 2016.

[3] T. Harada, S. Yamamoto, K. Yui, H. Kanda, Y. Chiba, H. Kameda, "Healing Robot PDDIN – Design and Implementation of prototype system -," The 65th National Convention Lecture and Paper Series, **2003**(1), 71–72.

[4] H. Myouki, Introduction to Freud, ChikumaShinsho, 2000.

[5] I. Kishimi, Introduction to Adlerian Psychology: For Better Relationships, bestsellers, 1999.

[6] H. Kawai, Introduction to Jungian Psychology, baifukan, 1967.

[7] M. Oikawa, H. Oikawa, "Consciousness and the unconscious in self-regulation: The effects of conscious compilation on goal priming," The Japanese Journal of Psychology, **81**(5), 485–491, 2010, doi:10.4992/jjpsy.81.485.

[8] H.K. Walker, W.D. Hall, J.W. Hurst, Clinical Methods, 3rd edition, Butterworth, 1990.

[9] S. Dehaene, H. Lau, S. Kouider, "What is consciousness, and could machines have it?," Science, **358**(6362), 486–492, 2017, doi:10.1126/science.aan8871.

[10] N. Okada, "A Computational Model of the Mind," IEICE ESS Fundamentals Review, **1**(4), 30–45, 2008, doi:10.1587/essfr.1.4_30.

[11] T. Maeno, "How to Make a Conscious Robot-Fundamental Idea based on Passive Consciousness Model-," Journal of the Robotics Society of Japan, **23**(1), 51–62, 2005, doi:10.7210/jrsj.23.51.

[12] H. Chie, H. Takato, N. Takayuki, "OS-18 Emotion and AI," Artificial Intelligence, **34**(6), 881–887, 2019, doi:10.11517/jjsai.34.6_881.

Impact Assessment of the Renewable Energy Sources Implementation in Bulgarian Single-Family Houses on the Greenhouse Gas by HOMER Pro Software

Dilyana Gospodinova[*], Peter Dineff

Department of Electrical Apparatus, Technical University of Sofia, Sofia, 1000, Bulgaria

A R T I C L E I N F O	A B S T R A C T
	It is widely known that cities now house more than half of the world's population. Within this framework, this study presents the possibilities for real-world application of renewable energy sources (RES) in urban areas, as well as their contribution to the urban deployment of the new energy paradigm. A comparison is made between hybrid power systems, operating on Bulgarian territory and conventional (traditional) power systems. Proposed hybrid system is designed and actually implemented to power a single-family house and consists of wind turbine (WG) photovoltaics (PV), lithium-ion batteries for energy storage and suitable converter. The HOMER Pro software was used to model and explore the long-term continuous implementation of a hybrid power system and greenhouse (GHG) gas emissions investigation. The article discusses the amount of carbon dioxide (CO_2) and nitrogen oxides (NO_X), that can be reduced by using a hybrid power system (solar and wind) in conjunction with a battery storage system (BSS - lithium-ion batteries) in single-family houses. Renewable energy sources combined with energy storage, according to this report, result in a 50% reduction in dangerous carbon dioxide and nitrogen oxide emissions. The proposed system is optimized based on the lower cost of energy (COE) and proper dispatch strategies (load following and cycle charging), resulting in greenhouse gas emissions distributed by regional Bulgarian cities.

1. Introduction

This paper is a continuation of work that was originally presented at the 2020 12th Electrical Engineering Faculty Conference (BulEF), [1]. The main task of this extended article is to expand knowledge and demonstrate a modern ability to share computing and modern technologies to assess important parameters such as the generation of harmful emissions. Kumar discuss [2] the opportunity to assess the generated harmful emissions is of particular importance and is a societal challenge for all of us. The rapid development of renewable technologies, as well as energy storage technologies, as well as their ever-increasing disposal, could to some extent provide a solution to the existing problem – environmental pollution generated at electricity produced for domestic use [3].

Many literary sources define the different types of green energy; some say there are five, while others say there are six or even seven [4–9]. Electricity production using fossil fuels has historically provided adequate costs but emits GHG.
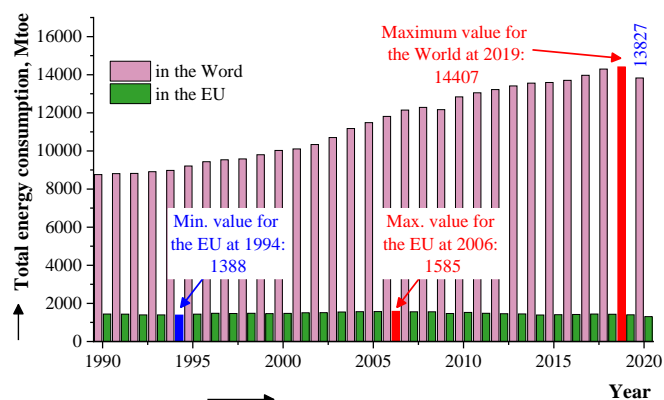

Figure 1: Total energy consumption over 1990 – 2020 in the Word and the EU.

According to the European Environment Agency, the European Union ranks third in the world in terms of greenhouse gas emissions, trailing only China and the United States [10]. As reported by Enerdata, contrary to expectations, in the face of a global pandemic, global energy consumption growth slowed by

362

4% in 2020, compared to an average of 2% per year from 2000 to 2018 and a 0.8% slowdown in 2019, Figure 1, [11].

Except for China, which is the world's largest energy consumer, most countries have reduced their energy consumption.

The difference in total power consumption between Europe and the rest of the world is between 6 and 10.6 times, and the trend is upward Figure 1. The data show that the global trend of increasing energy consumption continues. The year 2020 is an exception for both the world and Europe, due to the global pandemic caused by COVID-19 and its spread. Globally, energy consumption is steadily increasing, whereas in Europe, consumption is decreasing, most likely as a result of the European Union's policy to increase efficiency. Lockdown measures and limited economic activity had a significant influence on energy consumption in 2020. It dropped by about 7% in the European Union. Energy consumption fell across Africa and the Middle East as well, [11]. Under the Paris Agreement, the European Union (EU) has pledged to reduce GHG emissions by at least 40% by 2030 compared to 1990 (2016). The European Union is taking it a step further, aiming for nearly zero emissions by 2050 under the European Green Pact, [12]. The European Union is dedicated to achieving lofty goals in the fight against hazardous climate change. The European Commission announced a strategy for long term and achieving a climate friendly economy by 2050 in November 2018. By the second half of this decade, the EU hopes to be carbon neutral, [13]. The European Union has passed regulations in a variety of areas to attain the goal of reducing GHG emissions. The emissions trading scheme has undergone changes, for example, which oversees the progressive reduction of emissions from energy and industry while requiring polluters to pay a penalty. Various industries, such as agriculture and construction, have established national emission reduction targets. Because energy is at the heart of the emissions problem, the EU is working toward a strategy of cleaner sources and more efficient use. In 2018, Parliament passed laws requiring renewable energy sources (sun, wind, and water) to account for 32% of overall consumption and allowing households on their own to generate their own "green" energy, [14]. Furthermore, the EU has implemented rules on buildings and household appliances to increase energy efficiency by 32.5% by 2030, [15]. The "Clean Air For Europe programme (CAFE)" 6 of 2013 maintained the goal of achieving full compliance with current air quality regulations throughout the EU as soon as possible, with 2020 and 2030 targets specified.

The policy initiatives themselves are: the *first basis* consists of ground-level ozone, particulate matter, and nitrogen oxides ambient air quality standards, hazardous heavy metals, and a variety of other pollutants set out in the directives on ambient air quality; the *second basis* according to [16] is the Directive on national emission ceilings for the most major transboundary air pollutants, which establishes national emission reduction targets.

It is widely known that more than half of the world's population now lives in cities. Human and monetary activities both require energy. Fossil resources are decreasing in tandem with rising energy demand. In terms of global development, energy supply and security are major issues. The hunt for new energy sources based on RES has been prompted by the scarcity of energy and the demand for sustainable systems. In comparison to traditional fuel-powered generators, numerous renewable energy (RE) solutions are now well developed, cost-effective, and reliable, [17].

The true potential of solar energy is found in the numerous applications it offers at the municipal level, as it can be used to generate both heat and power in public and residential buildings and facilities. Solar energy has grown to be one of the most popular RES in Europe and Bulgaria as a result of its low cost, Figure 2. Solar energy is becoming a more common source of energy and a viable alternative to overcome energy shortages. According to [18], the main benefits of solar energy are accessibility, lack of pollution i.e. obtaining "green" energy, ease of operation, availability, and tranquillity. Furthermore in [19], the author point out, solar panels are known for their low maintenance, ease of installation and disassembly, and lack of moving components. A solar panel, as a RES technology, is simple to install and safe in urban environments. PV panels can be distributed and put in any sun- location exposure to the sun [20].
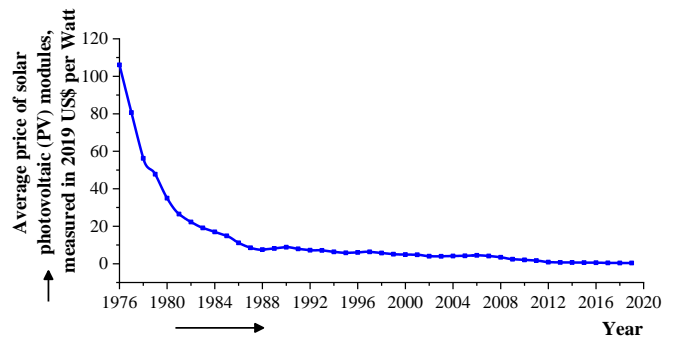

Figure 2: The average price of solar photovoltaic modules around the world, measured in US dollars per watt in 2019, [21].

PV systems and generators powered by the wind (WT) can be combined to generate power in a hybrid system. This system produces no greenhouse emissions. A system like this has the advantages of being practicable, dependable, simple, almost no maintenance and low cost. The economics of a solar-wind hybrid system is always rising due to constant reductions in component pricing. Because the energy obtained from the sun and wind fluctuates over time, a battery storage device must be included in the hybrid system, [21].

This extended article presents also new comparative data for Bulgaria, the EU and the World in connection with the use of RES and their share in the overall energy mix, Figure 3.
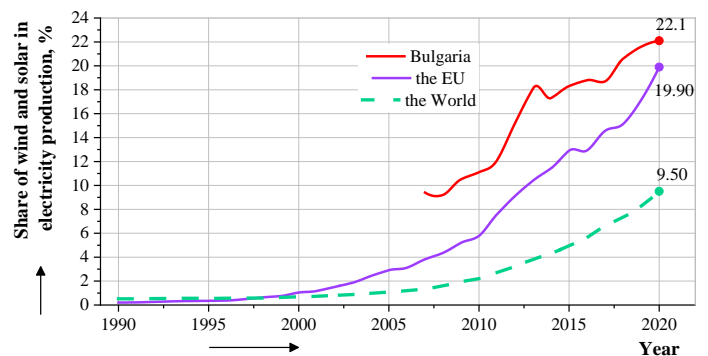

Figure 3: The percentage share of wind and solar in electricity production over 1990 – 2020 in the World, European Union and Bulgaria.

Increasing 0the amount of renewable energy in Bulgaria provides further economic, scientific, and political benefits: decreases the state's reliance on fossil fuels. Bulgaria first began promoting renewable energy in 2007, well after the formation and

execution of the institutional and legislative framework in the other old EU member states. In two periods (2007÷2012 and 2012÷2016), the state saw a significant increase in its percentage of renewable energy. Bulgaria now has approximately 1700 MW of installed renewable energy capacity (mostly wind and photovoltaic). Bulgaria is one of 11 EU countries that has already met its 2020 renewable energy goal. Energy from renewable sources accounted for 20.5% of gross final energy consumption in 2018, according to the Republic of Bulgaria National Statistical Institute [22-24].

A hybrid power station that combines PV and WT is promising future energy generation technology. Furthermore, Prasad proposes PV and wind power have an effect of synergy, with PV and wind power dominating during daytime and nocturnal hours, respectively [25]. Several studies have suggested that battery energy storage systems (BESS) be used to provide a variety of services for performance in order to deal with the variability of PV/wind energy generation [26-28]. The operation of a hybrid PV-WT system requires a battery storage system. The installation of a battery storage system enhances dependability, efficiency, and load smoothing. The hybrid power system is balanced thanks to the battery storage system [18,29].

The current computer-based study provides new valuable systematic price information connected to the so important chemical source of electricity - lithium-ion batteries as shown in Figure 4.
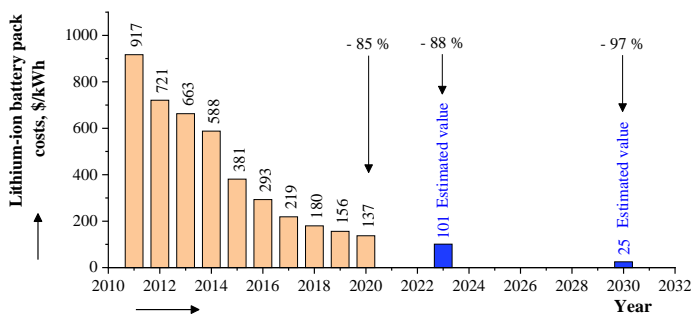

Figure 4: Costs of lithium-ion battery packs worldwide between 2011 and 2030 [30].

Energy storage is critical for integrating renewable energy sources into the power system. Furthermore, the combination of rooftop solar panels and battery storage is becoming increasingly popular in the hose-hold market. The cost of manufacturing lithium-ion batteries is steadily decreasing [30, 31]. These facts support the argument made in this article for the use of battery energy storage devices.

The complicated architecture of the power grid makes calculating the actual amount of emission reductions from energy efficiency and renewable energy measures difficult. Because air hazard pollutants like ozone (O_3), carbon monoxide (CO), sulfur dioxide (SO_2), and nitrogen oxides (NO_X) can be harmful to human health, it is critical to understand how to limit their emissions. Increased energy efficiency is one of the most well-known technologies. The usage of renewable energy sources is the second option.

The quantity of harmful emissions that can be decreased by employing a hybrid power system (WT and PV) in conjunction with a BSS is estimated in this article.

Numerous experiments have been conducted to study how to reduce costs by dispatching a battery without using forecasting techniques. There are also numerous battery optimization studies that use only one forecasting technique to achieve the least costs or greatest revenues. Although the complexity of the models was minimized in these investigations, the impact of different forecasting approaches on the optimization results could not be observed. As a result, the forecasting findings from several forecasting methodologies are integrated into the optimization model in this study [32-36].

The impact of a hybrid energy system with integrated energy storage technology on a single-family house in terms of harmful CO_2 and NO_X emissions has been investigated using an appropriate software product. The available literature does not present results on the distribution of harmful gas emissions after the implementation of RES on Bulgarian territory. The preliminary quantitative assessment yields result indicating a reduction in harmful CO_2 and NO_X emissions.

This article explains the contributions of the study in contrast to the article [1]. This paper's major contributions include the following: 1) The presentation of a dispatch strategy proposed by HOEMR Pro. 2) The demonstration of applying HOMER Pro software product to efficiently solve a hybrid system complex optimization problem, from a green gas emissions point of view, a single-family house power system that includes different renewable energy sources and battery storage system. 3) The use of a Load following (LF) dispatch strategy resulted in significantly lower hybrid power system operation costs. 4) The impact of the RES implementation and energy storage on the amount of generated emissions of CO_2 and NO_X, as well as the economic effect on the territory of Bulgaria, which is a member of the European Union, was assessed.

2. Design specification for the proposed hybrid system

This article takes a computer-assisted method. The simulation compares greenhouse gas emissions from hybrid energy supply systems in several Bulgarian regions. The proposed system configuration is simulated, optimized, and analysed using HOMER. The simulation assesses the system's technical feasibility and performs system optimization based on several system configurations to identify which one is the most suitable.

Forecasting PV and/or wind power generation is critical for developing a battery optimization plan for battery energy management. To meet the load, the HOMER software provides alternative dispatch algorithms to manage the operation of the generator and storage bank.
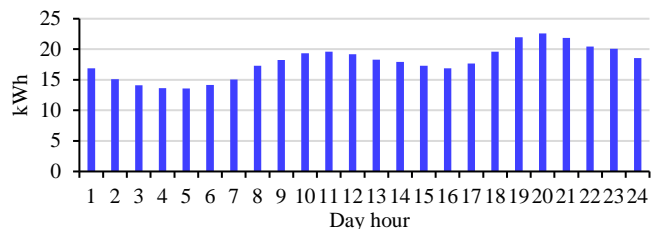

Figure 5: Daily profile for a Sofia-area (Bulgaria) urban household.

The shown standardized load profile is standard and pertains to residential customers in the Sofia metropolitan area, as recognized by Bulgaria's Energy and Water Regulatory Commission

(EWRC). The distribution company in the Sofia area of Bulgaria offered the developed load profile, Figure 5.

The HOMER Pro software is capable of simulating tree dispatch strategies - cycle charging, load following, and combined [37]. Which is ideal relies upon numerous variables, including the spans of the generators and battery bank, the fuel cost, the operation and maintenance (O&M) cost of the generators, etc.

When used in a load-following strategy (LF), a generator only produces enough power to meet the demand. LF works best in systems that use a lot of renewable energy because the renewable energy output occasionally exceeds the load. A generator is used in the cycle charging (CC) approach, and excess power is used to charge the battery bank. In systems with little or no renewable energy, cycle charging is preferred [37].

One of the objectives of this article is to compare alternative working strategies (LF- and CC-) in terms of greenhouse gas emissions in order to determine the best dispatch technique. It has been chosen to model both dispatch strategies, HOMER simulates each system separately.

Figure 4 depicts a schematic diagram of the microgrid designed by HOMER. A grid-connected hybrid system generates power from the sun and wind, which is then converted into grid-compliant electricity.
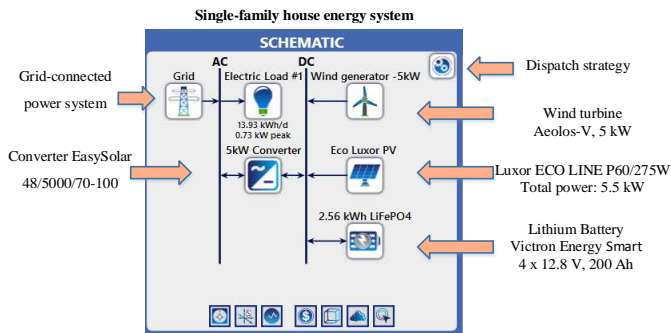

Figure 6: A schematic diagram of the hybrid power system that was designed to power a single-family home in Bulgaria.

2.1. Hybrid system components

The solar panels (Luxor ECO Line P60/275W) chosen had an initial and replacement cost of 3050 EUR, as well as an annual operation and maintenance cost of 20 EUR. It has a service life of 25 years. Without a tracking system, the selected panel's external current is DC, and the ground reflection coefficient is assumed to be 20%. The average efficiency coefficient was assumed to be 90% in this study. The coefficient of performance was used to represent the performance degradation caused by factors such as snow, shadows, high temperature, dust, wiring loss, and so on. It is usually 90%, but in hot weather it drops to 70%.

Four (4) identical series-connected 12V, 200Ah Li-ion Victron Energy batteries with an initial and replacement cost of 9600 EUR and an annual operating and maintenance cost of 50 EUR were used. The HOMER software accepts that the battery's highlights remain constant throughout its useful life and are not influenced by external factors such as temperature.

A converter is required for any system that includes both AC and DC components. For the simulation, a 5 kW converter with an initial and replacement cost of 4000 EUR is used. For grid compatibility in Bulgaria, the output voltage is set to 230V at

50Hz. The inverter has a 25-year lifespan and an efficiency of 95% as an inverter and 95% as a rectifier at a relative capacity of 100%.

A 5kW direct current wind generator was used. It had an initial and replacement cost of 800 EUR, as well as an annual operating and maintenance cost of 50 EUR. It has a 25-year life expectancy and was installed at a height of 10 meters above ground.

Reference [38] provides a detailed description of the peak electric load and standardized load profile.

2.2. HOMER's Greenhouse Gas Emission Assessment Capabilities

In [39], the author stated that there is a plethora of energy modelling software for renewable energy systems simulation available on the web. However, only five of them are appropriate for simulation procedures, and their environmental impact approach is appropriate.

The software needed to input technical and economic data, as well as energy component specs and resource data for renewable and non-renewable resources. Once all of the data has been entered into the software, HOMER calculates the environmental impact based on direct CO_2, CO, NO_X, and SO_2 emissions.

HOMER calculates emissions for each pollutant by multiplying the emissions factor (in g/kWh) by the annual fuel usage (in kWh) [37].

3. Discussion of simulation results

A 25-year prediction was used in the simulation. HOMER gave simulation results for a hybrid power system. All of the results reported here are for a hybrid system in Sofia, Bulgaria - 42°39.4'N, 23°21.2'E.

HOMER ran 114 simulation alternatives for the proposed hybrid power system, Figure 6, and all of them were feasible. This microgrid uses 27 kWh per day and has a max power output of 5 kW. The electrical load was served by the generation sources in the proposed system. This optimization has a net present cost of 20 800 EUR, while the cost of generating RE is 0.16 EUR per kilowatt-hour.

PV penetration of 136% for energy supply increases net grid purchase, whereas wind penetration of 11.4% decreases it. A renewable portion of 72.6% appears to be an appropriate load sharing between the grid, wind generator and PV panels.
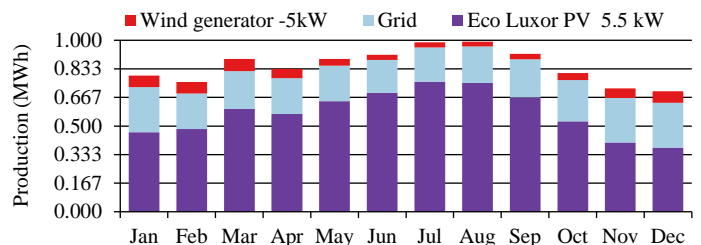

Figure 7: Average monthly electric production of a grid-connected hybrid power system in Sofia, Bulgaria.

Figure 7 indicates that the hybrid power system generates a total of 7516 kWh/year to sustain the load, lowering the grid purchase to 2697 kWh/year. It is clear that the solar component of the energy system has the biggest share, followed by power from the conventional grid. Despite a clear trend of increasing

renewable energy generation from wind turbine during the winter months of the year, the share of this type of source remains low.

In [1], it clearly shows the economics and operational statistics for the solar system, wind generator, and converter separately. At 136% penetration, the solar system's reported value is 5.50 kW. The Aeolos Wind Energy Ltd. wind turbine installation, rated at 5.00 kW, produces 581 kWh/yr of electricity.
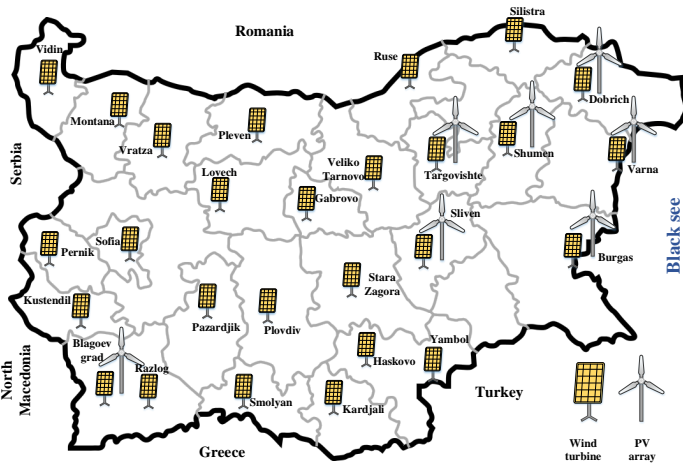

Figure 8: Preferred RES implementation by Bulgarian geographical area in accordance with [32], as determined by HOMER software.

HOMER calculates and displays the total amount of each pollutant generated yearly by the power system in kilograms per year of the six pollutants listed below: carbon dioxide (CO_2), carbon monoxide (CO), unburned hydrocarbons (UHC), particulate matter (PM), sulfur dioxide (SO_2), and nitrogen oxides (NO_x). Pollutants are produced through the use of fuel and biomass in generators, boilers, and reformers, as well as by the use of grid electricity [29]. The HOMER Pro software application allows for the presentation of Bulgaria's current carbon dioxide (CO_2) emissions situation. HOMER adds 8.93 EUR/ton CO_2 emissions to the total annual cost of the hybrid power system. Grid emissions are lowered as a result of power sales to the grid. According to our previous article [28], the recommended distribution of used RES corresponds to that shown in Figure 8.

The proposed distribution map was based on the demand for the cheapest per-kWh energy. The HOMER software package allows to estimate the cost of energy for each proposed system, and this parameter, as the most crucial, was tracked throughout the simulation process. The presented hybrid power supply system was multiplied and applied to each of Bulgaria's provinces. One of the goals of this research is to develop a roadmap for the use of renewable energy sources in single-family houses across the country, resulting in significant reductions in GHG emissions and the cost of energy, [40].

Figure 9 shows the results and comparison of electricity costs with renewable energy sources in a single-family home, as well as the distribution by major Bulgarian district cities. Two power systems were compared: one powered by the local electricity distribution company, and the other combining conventional electricity and RES (PV and WT) with a battery storage system.

In general, the cost of energy produced by a hybrid power system is higher than that produced by the conventional power system.
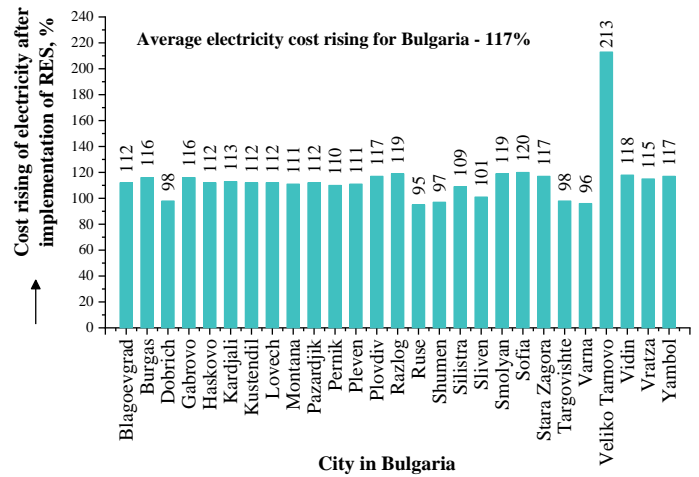

Figure 9: Costs of energy (COE) comparison between hybrid power system and conventional grid-connected system based on preferred energy sources shown in Figure 8.

When powering a single-family home, the proportion of renewable energy in the overall energy mix is an important factor in the modelling and analysis of the results. Figure 10 shows a clear trend toward a high percentage of energy generated from renewable sources in the total energy mix. The share of renewable energy accounts for between 67.7% and 79.0% of total energy consumption. This percentage stays high enough to power Bulgaria's entire territory with renewable energy.
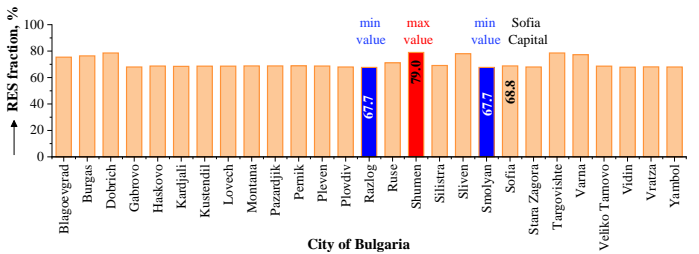

Figure 10: Distribution of RE fractions (%) in Bulgaria's major cities using a hybrid system, calculated by HOMER Pro.

Harmful CO_2 emissions for a single-family house with a hybrid power supply system are between 1.318 and 1.905 kg per year. Figure 11 shows that the baseline level of CO_2 emissions when using solely a conventional electrical system is 3.213 kg/year.

When a hybrid energy system is combined with battery storage, the percentage change with a negative sign goes from 41% to 59%, indicating a considerable reduction in harmful CO_2 emissions. Several locations of the country, including Shumen, Varna, Targovishte, Sliven, and Dobrich, have the potential to reduce CO_2 emissions by twofold.

Nitrogen oxides (NO_X) are a group of seven chemical molecules that pollute the air. The results for nitrogen oxides are also presented in that article in a different manner. The situation with nitrogen oxides, or NO_X, is identical to CO_2 emissions. Nitrogen oxides emissions from a single-family hose powered by the traditional energy system are 6.81 kg per year. In a single-family house, using a hybrid power system reduces hazardous nitrogen oxides by 41% to 57%, depending on the location where the home is located.
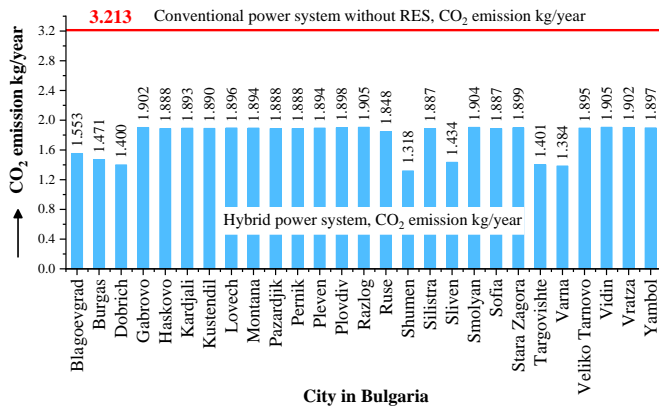
Figure 11: Comparison of CO_2 (kg/year) emissions by Bulgarian regions for conventional grid-connected and hybrid power systems.

4. Conclusions

The renewable energy fractions in the proposed hybrid power system for a single-family dwelling in Bulgaria are high, reaching 79 percent in some regions of the country.

Despite the limited percentage of wind energy, there are some areas of the country where a greater number of wind turbines would result in a further reduction in electricity prices.
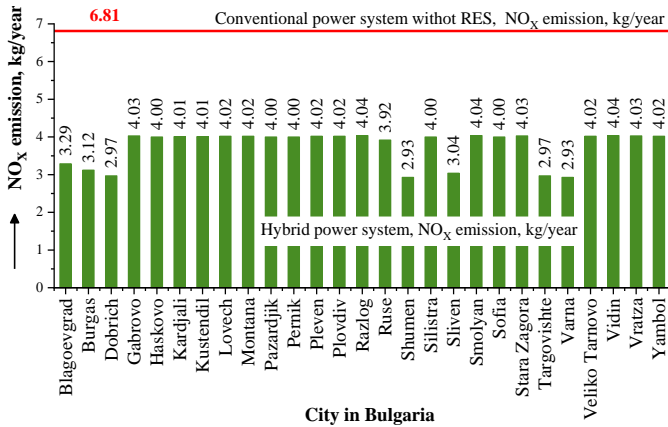

Figure 12: Comparison of NOx (kg/year) emissions by Bulgarian regions for conventional grid-connected and hybrid power systems.

The HOMER Pro software has been successfully used to anticipate greenhouse gas emissions from grid-connected single-family houses in Bulgaria.

When a hybrid power system is used in conjunction with proper energy source and storage management, harmful CO_2 emissions can be reduced by up to 59% in some parts of the country. Similarly, NO_X generation reductions in some sections of the country might reach 57%, which is an excessively high percentage.

To summarize, the software package Homer Pro may be used effectively for quickly assess harmful greenhouse gas emissions, both in Bulgaria and throughout the European Union's territory, which is particularly important for transboundary air pollution.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] D. Gospodinova, P. Dineff, K. Milanov, "Greenhouse Gas Emissions Assessment after Renewable Energy Sources Implementation in Bulgarian Grid-Connected Single-Family Houses by HOMER Pro Software," 2020 12th Electrical Engineering Faculty Conference, BulEF 2020, 2020, doi:10.1109/BULEF51036.2020.9326082.

[2] M. Kumar, Social, Economic, and Environmental Impacts of Renewable Energy Resources, IntechOpen, 2020, doi:10.5772/INTECHOPEN.89494.

[3] K.P. Tsagarakis, A. Mavragani, A. Jurelionis, I. Prodan, T. Andrian, D. Bajare, A. Korjakins, S. Magelinskaite-Legkauskiene, V. Razvan, L. Stasiuliene, "Clean vs. Green: Redefining renewable energy. Evidence from Latvia, Lithuania, and Romania," Renewable Energy, **121**, 412–419, 2018, doi:10.1016/J.RENENE.2018.01.020.

[4] N.K.M.A. Alrikabi, "Renewable Energy Types," Journal of Clean Energy Technologies, 61–64, 2014, doi:10.7763/JOCET.2014.V2.92.

[5] IRENA, Overview of renewable energy, Overview of Renewable Energy, 27, Sep. 2021.

[6] U.S.E.I.A. (EIA), Renewable energy explained - types and usage, 2021.

[7] J. Kantenbacher, R. Shirley, Renewable Energy: Scaling Deployment in the United States and in Developing Economies, Butterworth-Heinemann: 89–109, 2018, doi:10.1016/B978-0-12-813964-6.00005-7.

[8] I. Yahyaoui, Renewable Energies and Irrigation, Elsevier: 1–14, 2017, doi:10.1016/B978-0-12-812039-2.00001-6.

[9] N. El Bassam, "Energy basics," Distributed Renewable Energies for Off-Grid Communities, 113–122, 2021, doi:10.1016/B978-0-12-821605-7.00005-2.

[10] European Environment Agency, Air pollution, 2020.

[11] Enerdata, World Energy Consumption Statistics | Enerdata, Global Energy Consumption Fell in 2020 (-4%), Due to Lockdown Measures and Transport Restrictions, 2021.

[12] United Nations, Paris agreement, 2015.

[13] European Commission, 2050 long-term strategy, 2018.

[14] European Commission, Fourth Biennial Report from the European Union Required under the United Nations Framework Convention on Climate Change, Brussels.

[15] European Commission, Energy label and ecodesign, About the Energy Label and Ecodesign, 2020.

[16] European Commission, A Europe that protects: Clean air for all.

[17] Y. Yongchang, Y. Chuanan, "Implementation of a MPPT Controller Based on AVR Mega16 for Photovoltaic Systems," Energy Procedia, **17**, 241–248, 2012, doi:10.1016/j.egypro.2012.02.090.

[18] A. Mohanty, M. Viswavandya, D.K. Mishra, P.K. Ray, S. Pragyan, "Modelling & Simulation of a PV Based Micro Grid for Enhanced Stability," Energy Procedia, **109**, 94–101, 2017, doi:10.1016/j.egypro.2017.03.060.

[19] K. Adam, V. Hoolohan, J. Gooding, T. Knowland, C.S.E. Bale, A.S. Tomlin, "Methodologies for city-scale assessment of renewable energy generation potential to inform strategic energy infrastructure investment," Cities, **54**, 45–56, 2016, doi:10.1016/J.CITIES.2015.10.015.

[20] A. Gasparatos, C.N.H. Doll, M. Esteban, A. Ahmed, T.A. Olang, "Renewable energy and biodiversity: Implications for transitioning to a Green Economy," Renewable and Sustainable Energy Reviews, **70**, 161–184, 2017, doi:10.1016/J.RSER.2016.08.030.

[21] H. Ritchie, M. Roser, Energy, 2020.

[22] Statista, Electricity: share of renewable energy 2007-2018 | Statista, Share of Energy from Renewable Sources in Electricity Generation in Bulgaria from 2007 to 2018, 2021.

[23] European Commission, Europe 2020 targets: statistics and indicators for Bulgaria | European Commission, Europe 2020 Targets: Statistics and Indicators for Bulgaria, 2020.

[24] National Statistical Institute, Share of renewable energy in gross final energy consumption | National statistical institute, 2021.

[25] A.A. Prasad, R.A. Taylor, M. Kay, "Assessment of solar and wind resource synergy in Australia," Applied Energy, **190**, 354–367, 2017, doi:10.1016/J.APENERGY.2016.12.135.

[26] H. Zhao, Q. Wu, S. Hu, H. Xu, C.N. Rasmussen, "Review of energy storage system for wind power integration support," Applied Energy, **137**, 545–553, 2015, doi:10.1016/J.APENERGY.2014.04.103.

[27] A. Berrada, K. Loudiyi, "Operation, sizing, and economic evaluation of storage for solar and wind power plants," Renewable and Sustainable Energy Reviews, **59**, 1117–1129, 2016, doi:10.1016/J.RSER.2016.01.048.

[28] O.M. Toledo, D. Oliveira Filho, A.S.A.C. Diniz, "Distributed photovoltaic generation and energy storage systems: A review," Renewable and Sustainable Energy Reviews, **14**(1), 506–511, 2010, doi:10.1016/J.RSER.2009.08.007.

[29] F. Baghdadi, K. Mohammedi, S. Diaf, O. Behar, "Feasibility study and energy conversion analysis of stand-alone hybrid renewable energy system," Energy Conversion and Management, 2015, doi:10.1016/j.enconman.2015.07.051.

[30] Statista, Worldwide - lithium ion battery pack costs | Statista, Lithium-Ion Battery Pack Costs Worldwide between 2011 and 2030, 2021.

[31] M.S. Ziegler, J.E. Trancik, "Re-examining rates of lithium-ion battery technology improvement and cost decline," Energy & Environmental Science, **14**(4), 1635–1651, 2021, doi:10.1039/D0EE02681F.

[32] I. Miranda, N. Silva, H. Leite, "A Holistic Approach to the Integration of Battery Energy Storage Systems in Island Electric Grids with High Wind Penetration," IEEE Transactions on Sustainable Energy, **7**(2), 775–785, 2016, doi:10.1109/TSTE.2015.2497003.

[33] S. Mazzola, C. Vergara, M. Astolfi, V. Li, I. Perez-Arriaga, E. Macchi, "Assessing the value of forecast-based dispatch in the operation of off-grid rural microgrids," Renewable Energy, **108**, 116–125, 2017, doi:10.1016/J.RENENE.2017.02.040.

[34] C. Zhang, Y. Xu, Z.Y. Dong, J. Ma, "Robust Operation of Microgrids via Two-Stage Coordinated Energy Storage and Direct Load Control," IEEE

Transactions on Power Systems, **32**(4), 2858–2868, 2017, doi:10.1109/TPWRS.2016.2627583.

[35] H.H. Abdeltawab, Y.A.R.I. Mohamed, "Market-oriented energy management of a hybrid wind-battery energy storage system via model predictive control with constraint optimizer," IEEE Transactions on Industrial Electronics, **62**(11), 6658–6670, 2015, doi:10.1109/TIE.2015.2435694.

[36] M.R. Sandgani, S. Sirouspour, "Coordinated Optimal Dispatch of Energy Storage in a Network of Grid-Connected Microgrids," IEEE Transactions on Sustainable Energy, **8**(3), 1166–1176, 2017, doi:10.1109/TSTE.2017.2664666.

[37] H.P. Software, HOMER Pro, 2020.

[38] D. Gospodinova, K. Milanov, M. Minchev, P. Dineff, "Techno-Economic Feasibility Analysis of Nearly-Zero Hybrid Energy System for the City of Sofia in Bulgaria," in 2019 11th Electrical Engineering Faculty Conference, BulEF 2019, 2019, doi:10.1109/BulEF48056.2019.9030697.

[39] H. Sharma, É. Monnier, G. Mandil, P. Zwolinski, S. Colasson, "Comparison of environmental assessment methodology in hybrid energy system simulation software," Procedia CIRP, **80**, 221–227, 2019, doi:10.1016/j.procir.2019.01.007.

[40] D. Gospodinova, K. Milanov, P. Dineff, "Route map for renewable energy sources implementation for household in Bulgaria," in IOP Conference Series: Materials Science and Engineering, IOP Publishing: 012016, 2020, doi:10.1088/1757-899X/878/1/012016.

Problems of Increasing the Intelligence of Algorithms for Optimal Distribution of the Current Load on the Combined Heat and Power Plant and Ways to Solve Them

Arakelyan Edik[1,*], Kosoy Anatoliy[1], Andryushin Alexander[1], Mezin Sergey[1], Yagupova Yulia[1], Leonov Maxim[1], Pashchenko Fedor[2]

[1]*ACS TP dept., National Research University "MPEI", Moscow 111250, Russia*

[2]*V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow 117997, Russia*

A B S T R A C T

The problem of optimal distribution of the current load of a combined heat and power plant with a complex composition of equipment is considered. A multi-criteria approach is proposed for parts of graphs with a constant load in time, considering the factors of economy, reliability and ecology. For sections of the graph with a time-variable load, we consider the formulation of a dynamic optimization problem. In order to increase the intelligence of solving the tasks set, we propose a variant of building an intelligent system that, in addition to a number of intelligent functions, aims to: switch to the principle of advanced (predictive) control when performing the dispatching schedule of loads; select the number of criteria under consideration; evaluate the values of unmeasured or poorly measured parameters; identify the current technical condition of equipment, as well as recommend either a solution available in the database, or perform new calculations.

1. Introduction

This article is a continuation of the work originally presented at the international conference "Managing the Development of Large-scale Systems" (MLSD-20) [1]. In this report, the problems and increasing the intelligence of algorithms for optimal load distribution at a thermal power plant were considered in the following conditions:

- the optimal distribution of the time-constant load of the station with a given composition of generating equipment was considered, i.e. the problem of static optimization;
- a single-criteria task was solved according to the economic criterion – the maximum profit of the station from the sale of electricity in various sectors of the electricity market and power and heat to the thermal consumer, or minimizing the total fuel consumption at the station.

The technical capabilities of automation complexes are several generations ahead of the software implementation of applied algorithms for optimal control and control of the technological process of electricity production. It is more rational, in the conditions of high prices for the work of talented programmers and technologists, not to try to solve this problem with the old catching-

up methods, but to approach this problem cardinally, entrusting the solution of the complex problem of the efficiency of electricity generation at thermal power plants to artificial intelligence. The intermediate stage of the transition to the management of the entire plant by AI is the transition to the management of individual levels, whether it is the level of the station or the level of the power unit. [2]-[4].

Determination of the optimal operating mode for power plants and power systems is an urgent problem in the power industry. Over time, the solution of this optimization problem is being improved and at present, in connection with the transition of energy to market relations, the optimization of power plant operation modes seeks not only to reduce the cost of electricity generated, but also to increase reliability and reduce emissions of harmful fuel combustion products into the environment [4]-[7].

The relevance of considering the reliability factor is due to the fact that the task of choosing the optimal operating modes of thermal power plants (TPP) equipment includes two interrelated tasks: choosing the composition of generating equipment and the distribution of heat and electric load with a known composition of generating equipment. Currently, in accordance with the Regulations of the electricity and power market, the selection of the composition of generating equipment and its coordination with

*Corresponding Author: Arakelyan Edik, Edik_arakelyan@inbox.ru

the system operator (SO) of the power system is carried out 2-3 days before the operational day, and the station performs load distribution per day when preparing an application for entering the electricity and capacity market "for the day ahead". Due to the fact that the agreed composition of the equipment during optimization calculations and execution of the dispatching schedule of loads cannot be changed, considering the reliability factor in solving the problem becomes obvious and mandatory. For TPPs located near urban buildings, the environmental factor is also important. At the same time, the task is not to ensure emissions within the permissible regulatory limits, but to minimize them as much as possible.

In connection with the above, in the development of the report at the conference, this article discusses the problems and ways to improve the intelligence of the algorithm for optimal distribution of thermal and electrical loads of a combined heat and power plant (CHPP) with a complex composition of equipment with a multi-criteria problem statement - in terms of efficiency, reliability and environmental friendliness.

At the same time, in addition to the sections of the load schedule with a constant electrical load in time, the sections with a variable load (the stages of unloading and loading) are also considered. This is due to the fact that the duration of the unloading/loading processes of the station is several hours, while the efficiency indicators of the power units vary within a fairly wide range, which leads to significant additional fuel losses if the optimal sequence of unloading/loading of the power units involved in the process of passing the load failure is not observed.

To assess the amount of computational procedures required, we will consider the stages of preparing and entering the power plant market for electricity and capacity in the market conditions of the power plant. In accordance with the Regulations of the electricity and capacity market, the following stages are considered for managing the operating modes of the CHPP equipment for the operational day:

- selection and approval of the composition of generating equipment for the operational day in order to form the optimal station request for each group of delivery points (GDP);
- drawing up a possible plan for electricity generation with projected volumes of heat output to form the optimal application of the station for participation in the competitive market "for the day ahead" for each GDP conducted by the administrator of the trading network (ATN);
- operational optimization when performing the dispatching load schedule for each GDP – optimal distribution of thermal and electrical loads between the switched-on units with a known composition of the operating equipment;
- optimization of the dispatching load schedule to generate an optimal operational price-accepting request for each GDP for the deviation of the electrical load from the dispatching schedule in the balancing market controlled by the SO.

2. Statement of a Multi-Criteria Static Problem of Optimal Load Distribution

Setting a multi-criteria task: for a given total electrical and thermal loads of a CHPP, it is necessary to determine such loads at which the total current costs for the production of electrical and

thermal energy will be minimal with the greatest reliability of the equipment and the least amount of harmful emissions into the environment [8].

The mathematical formulation of a static problem with three criteria is formed as follows: find

$$
\left.
\begin{aligned}
F_1(x) &= \sum_{j=1}^{n} f_{1j}(p_j, q_j) \rightarrow \min_{\mathbf{x}} \\
F_2(x) &= \sum_{j=1}^{n} f_{2j}(p_j, q_j) \rightarrow \min_{\mathbf{x}}, \\
F_3(x) &= \sum_{j=1}^{n} f_{3j}(p_j, q_j) \rightarrow \min_{\mathbf{x}}
\end{aligned}
\right\}
\quad (1)
$$

under the following conditions:

$$\sum_{j=1}^{n} p_j = P_\Sigma , \quad \sum_{j=1}^{n} q_j = Q_\Sigma \quad (2)$$

$$p_{j_{min}} < p_j < p_{j_{max}},$$

$$q_{j_{min}} < q_j < q_{j_{max}}, \quad (3)$$

$$p_j \neq q_j \leq S_{j_{max}}, j = 1, 2, \dots, n$$

where $F_1(x), F_2(x), F_3(x)$ - optimality criteria for efficiency, reliability and ecology; $f_{1j}(p_j, q_j)$, $f_{2j}(p_j, q_j)$, $f_{3j}(p_j, q_j)$ - energy characteristics of the j-th power unit, reflecting fuel consumption, average damage for unreliability of operation (due to emergency downtime and failures) and the amount of emissions of harmful substances into the environment; p_j, q_j - electrical and thermal power of the j-th unit; P_Σ, Q_Σ- total required electrical and thermal power set for the entire system; $p_{j\min}, p_{j\max}, q_{j\min}, q_{j\max}, S_{j\max}$ - permissible load limits; $\mathbf{x} = \{ x_1, \dots, x_n \}; \mathbf{p} = \{ p_1, \dots, p_n \}; \mathbf{q} = \{ q_1, \dots, q_n \}.$

The essence of optimization is to achieve a compromise solution for three criteria, with an optimization horizon of one to several days when forming an application for the selection of the composition of the included generating equipment for an operational day in 2-3 days, applications for the market for a day ahead, as well as operational applications for the balancing market. When implementing the dispatching schedule of loads, the number of considered criteria can be changed both in the direction of decrease and in the direction of increase

3. Statement of the Dynamic Problem of Optimal Load Distribution

The need to consider the formulation of a dynamic optimization problem is due to the following. As a rule, in the absence of control of the operating modes of power units in automatic mode, if it is necessary to involve several power units in changing the load of the station in order to eliminate the likelihood of non-fulfillment of the specified schedule, all power units begin the procedure for changing the load almost simultaneously, without taking into account their individual characteristics. This

often leads to significant deviations of the total load of the station from the required one, violation of the required time for the station to reach the specified load, especially when it is necessary to stop or start one or more power units. At the same time, in the absence of a clear schedule of load changes over time at each specific power unit, as a rule, there is an advanced or delayed load change, which leads to additional fuel losses, deviations in power generation from the required dispatch schedule, and other violations (for example, deviation of the permissible load change rates during unloading or loading).

To eliminate these disadvantages, it is necessary to organize such a sequence of unloading/loading of power units from a given initial state to the required final, also known state. It is assumed that for these states, with a known equipment composition, the problem of optimal distribution is solved on the basis of a multi-criteria approach, and for their docking with each other, the following conditions must be met:

- the change in the load of the power plant $\Delta N_{PP}(\tau_{u/l})$ over time is the sum of the change in the loads of each of the power units $\Delta N_i(\tau_{u/l})$ for the considered time $\tau_{u/l}$ of unloading/loading the station, taking into account the permissible deviation from the dispatch schedule α, i.e.

$$\Delta N_{PP}(\tau_{u/l}) = (1 \pm \alpha) \sum_{i=1}^{n} \Delta N_i(\tau_{u/l}), \qquad (4)$$

in this case, the condition below is met at each moment of time

$$N_{PP\tau} = \sum_{i=1}^{n} N_{i\tau} = \sum_{i=1}^{n} (N_{i0} + \Delta N_{i\tau}), \qquad (5)$$

where $N_{PP\tau}$ is the power of the station at the moment τ; $N_{i0}, \Delta N_{i\tau}$ - the power of the i-th block before the change in the power of the station and its increment in time.

- the rate of load change at each power unit is within the permissible unloading/loading speed

$$w_{iu/l}^{min} \le w_{up/l} \le w_{up/l}^{max}; \qquad (6)$$

- the capacity of each power unit can vary from the initial N_{i0} to the required final capacity N_{iF} in accordance with the dispatching schedule (it is assumed that the power unit capacity limits are taken into account when distributing the station load before and after the load change).

The features of solving this dynamic problem are as follows:

- the composition of the equipment before and after changing the load of the station is identical, and therefore it makes no sense to consider the criteria of reliability and environmental friendliness, so a single-criteria task is considered;

- as a criterion of two possible functions used in solving similar problems – the total fuel loss or the total fuel consumption during the load change at the station, it is advisable to use the function of fuel consumption during the load change;

- in contrast to the previously adopted methods in this paper, in addition to the fuel costs calculated from the consumption characteristics of each power unit, it is also necessary to take into account the additional fuel costs not taken into account by these characteristics in the unloading/loading processes associated with the unsteadiness of heat exchange processes in heat exchangers, unstable operation of regulators, etc. in transients in the equipment of the unit [9].

Taking into account the additional costs, the optimization criterion is presented in the form:

$$B_{PP}(N_{PPu/l}, \tau_{u/l}) = \sum_{i=1}^{n} \left[\int_{\tau_0}^{\tau_{u/Hl}} B_{i\tau} \delta\tau + \Delta B_{PPi}\left(\frac{\Delta N_{iu/l}}{N_i}, W_{iu/l} \right) \right] \quad (7)$$

where $B_{PP}(N_{PPu/l}, \tau_{u/l})$ - the total fuel consumption at the station during the load change; $N_{PPu/l}$ - the magnitude of the load change at the station; $\tau_{u/l}$ - the duration of the load change; B_i - the function of fuel consumption on the i-th block when the load changes; $\Delta B_{PPi}\left(\dfrac{\Delta N_{iu/l}}{N_i}, W_{iu/l} \right)$ - the function of additional costs in the unloading/loading processes, depending on the depth and speed of the load change on the i-th block.

The initial data of the task are: the initial and final power of the station, the duration of the load change $\tau_{u/l}$, hour (min.); the rate of load change at the station $W_{PP} = \dfrac{\Delta N_{PP}}{\tau_{u/l}}$; the energy characteristics of the units in the form of a dependence [8]; the functions of the dependence of additional fuel costs in the unloading/loading processes, the initial and final load distribution of the station before and after the load change between the generating equipment, the permissible unloading/loading speeds of the power units.

The desired (variable) values are: the moment of the beginning of unloading/loading of each power unit in relation to the beginning of the load change at the station; the dynamics of the load change of each power unit during the transition from the initial state of the station to the final state.

To solve this problem, it is proposed to use a dynamic programming algorithm (Bellman's method) [8], which is used to optimize the management of multi-stage processes or to optimize multi-link systems, when it is necessary to find the control action (or the state of each link of the system) at each stage, so that the general control criterion reaches the optimal value. At the same time, at each next stage, the result of optimization at the previous stage is taken into account. The peculiarity of the method consists in a sequential, step-by-step solution of the optimization problem, so that at each stage only a relative number of possible options are selected that provide the optimal values of the selected criterion.

4. Ways to Increase Intelligence in Solving the Optimal Distribution Problem Based on Multi-criteria Optimization

The analysis of the above statements of static and dynamic optimization shows the complexity of their solution by traditional

methods and it becomes obvious that their practical implementation will require the use of modern information and digital technologies. Below are some, far from exhaustive, ways to increase the intelligence of the automated control system in solving the tasks set.

- Transition to the principle of advanced (predictive) control when performing the dispatching schedule of loads. The essence of the proposal: first, based on the analysis of previous optimization calculations, to estimate the required delay time between the start of the optimization calculation and before the formation of the impact on the power units. There are two possible ways to implement it: 1) calculate it based on the average data in the chain – the task, the preparation of the source data, the solution, the solution check, the signal generation, the controller, the actuator; 2) create a database at the specified time and, if necessary, taking into account the dimension of the current optimization problem, find a suitable lead time;

- Before performing optimization calculations, determine the feasibility and necessity of selecting the number of criteria and, if there is more than one criterion, the priority of the criteria. The assessment can be carried out on the basis of expert assessments, based on the current load area on the daily schedule according to pre-accepted conditions (the significance of the criteria at different hours of the day, or by the amount of load). For example, when passing through the failures of load schedules, the environmental criterion can either be excluded from consideration, or can be taken into account as a restriction, while the significance of the remaining two criteria (efficiency and reliability) depends on the redundancy mode: if all units are in generation mode, then the efficiency criterion is more significant; if there are stopped units, then the reliability criterion. When passing the maximum load, the efficiency criterion plays a secondary role, because all units operate in an economical mode.

The development of the library module of the intelligent controller is inseparable from the problem of verification and validation of the results obtained. We can get an ideal imitator of a traditional PID controller, we can get an auto-tuning blog simulator based on linear regression models, or we can just get an intelligent controller with a defect in a hidden area of knowledge. The flaws in the training of the received intelligent controller may be hidden from us for many years of operation, and then, for reasons that we do not understand, the optimizer will go into this dark area of uncalculated or incorrectly calculated.

Developers are already presenting many options for really working intelligent controllers based on deep neural networks and machine learning principles, and technical capabilities allow you to scale and launch impressive neural network controllers in terms of complexity. But so far, there are problems of proving that these regulators find and follow the global extremum, which is an uncritical problem at this stage, and the problem of proving the stability and reliability of the operation of this intelligent regulator in the time and operational areas of determination.

So far, they are trying to hide the problem of this fundamental limitation behind the no less cumbersome problem of creating a digital twin of a power plant. It is assumed that we will have a mathematical model similar to a full-scale simulator of the station. And on this model, we will first, but in real time, check the

solutions of the intelligent controller before making adjustments to the controller of a real object. This is an interesting, but not entirely honest approach. Removing the issue of validation and verification of a developed and trained intelligent controller by justifying that everything is checked through a digital double generates, with due responsibility, questions about the validation and verification of a digital double. And if we take into account that the power plant as a control object changes over time – and here we get the problem of the impossibility of building a working model of a digital double without using artificial neural network technology. This again leads us to the problem of trust and the implication of the module's operation and proves the dead-end nature of this way of solving the problem.

At this stage, the solution will be a hierarchical distribution of the entire control object into functional groups, and then a distribution according to the importance and danger of violations of the technological process in each group. From this partition, we will get the area of permissible rights for the intelligent controller module. All these rights should be limited from above by the control of the human operator. [10, 11].

To implement the interaction described above, the software that runs the optimal load balancing algorithm must support a number of functions and features, which are presented below:

- Support multiple standard data transfer protocols: OPC UA, OPC DA, HART, Modbus, etc.
- Buffering, storing and analyzing information in a separate archive according to the selected parameters for a specified period of time. This information includes: key parameters of the unit operation, digitized consumption characteristics obtained online and via the regulatory and technical documentation.
- The ability to automatically obtain approved dispatch schedules that serve as similar data for the implementation of the calculation.
- Generation of reports and graphs based on the received data with an assessment of the degree of accuracy of the problem solution.
- Implementation of the authentication process, with login only for registered accounts. This is necessary to restrict access and prevent unauthorized changes to the logic of the program.
- Development taking into account cross-platform, in connection with the current trends in the industrial market.

The requirements described above are currently met by most modern PTCs, including domestic ones. This means that for the technical implementation of this algorithm and its integration into the automated process control system of the station will not require the purchase and adjustment of new equipment. It is enough to virtually isolate the unused power of the PTC, creating a virtual functional node for solving intellectual tasks. In this node, everything is also virtual, that is, without spending money on the purchase and laying of new cable connections, to bring existing signals from sensors and programs from the PTC. The output signals from the intelligent optimization node can be easily connected both directly with additional signals in the automated regulation systems routines of the block capacities, and in the form of recommendation information on the mnemonic circuits to the operator.

5. Problems of Integration of the Optimal Load Distribution Algorithm into the PTC Application Software and Possible Solutions

The analysis of the functional capabilities of modern automated control systems (ACS) of TPP built on the basis of PTC of both domestic and foreign production shows that the improvement of the quality of technological processes, optimization of their modes and parameters, based on new technical and software tools and achievements of the modern theory of automatic control of technological processes, occurs mainly at the level of equipment and power unit, and to a much lesser extent – at the level of power plants. It is obvious that the further development of the ACS through the intellectualization of power plant process control systems should be aimed at creating a single system for optimal control of technological and production processes in all operating modes of equipment, power units and the plant as a whole. The analysis of the current state of use in the design of ACS of domestic software and hardware complexes, their information and functional capabilities in terms of the possibility of integration into their application software of the proposed optimal distribution algorithm revealed the problems that arise in this case, both from the point of view of technical and information support. When considering the problems that arise, it is assumed that in the future, in addition to the optimal distribution algorithm, a number of other station-level control algorithms will be integrated into the application software of PTC, the implementation of which will significantly increase the overall level of intelligence of the automated control system as a whole.

The main requirements for the technical part, in which the intelligent algorithm will be directly implemented, are: sufficient server power, in which the software complexes will work; high speed of the server and the algorithms; the software must be able to transmit data according to the OPC UA specification, which is the newest and most advanced specification for information transmission (for the first time in the OPC family of technologies, the possibility of cross-platform compatibility was implemented). OPC UA also has the highest level of security in the OPC family of technologies. Its main advantage is the transmission of data over the Internet.

There are also basic requirements for systems that are imposed regardless of the software and hardware complex used, such as:

requirements for the protection of system information, including the protection of the system from the penetration of external software and technical influences that are made over the data transmission network;

- all changes to the system access settings must be logged in the system status log;

- access to the system should be divided into roles, for example: engineer, operator, etc.;

- the system should be able to restore databases after failures, failures and accidents.

- Some requirements for storing calculated data:

- Storing parameter values for a period specified by the user or the system;

- Full description of the database where parameter values are stored;

- Long-term storage of information;

- The user should be able to run algorithms and generate reports for any period (hours, minutes, seconds, etc.);

- The system should be able to observe changes in parameters in real time.

According to the prediction of parameters for calculating the power unit's technical and economic indicators (TEI) and carrying out optimal load distribution, the system should be able to select the past periods from the system's database, in the case when the equipment used coincides with the composition of the forecast period. Forecasting allows you to save a lot of time on entering the initial data. If it is impossible to predict the technological parameters of the system, it should be possible to manually enter the initial data by the operator.

The analysis showed that the above requirements are currently met by the majority of modern PTC, including domestic ones. This means that the technical implementation of the proposed algorithms and their integration into the automated process control system of the station will not require the purchase and adjustment of new equipment. It is enough to virtually isolate the unused power of the PTC, creating a virtual functional node for solving intellectual tasks. The output signals from the intelligent optimization node can be easily connected directly as additional signals to the ACS routines of the block capacities, or as recommendation information to the mnemonic circuits to the operator.

The technical difficulties of integrating the developed artificial intelligence modules into the existing software and hardware complexes at the stations do not require, with the proper level of qualification of the station personnel, a planned shutdown of the entire station for repair. Individual modules for adjusting the settings of existing regulators can be exported and activated on running equipment, since initially these modules only give recommendations. The subsequent switching of the output ports of these modules with the input ports of the controller settings can occur both in manual mode and in automatic mode, when the convergence of the expected and received results of the operation of the bundle of the intelligent controller and the digital double of the station is achieved.

And at the stage of a planned shutdown, it is recommended transfer some of the design functions of the automated process control system from the block level to the station level – those tasks that are associated with general station services (for example, calculation and analysis of the power units TEI, selection of the optimal pressure in the condenser), which will significantly relieve the information load of the block level controllers.

6. Conclusion

- The studied formulation of a multi-criteria static problem for sections of the load schedule with a constant and a single-criteria dynamic problem for sections with a variable load with an optimal distribution of the current load at a thermal power plant with a complex composition of equipment.

- To increase the intelligence of algorithms for optimal load distribution, a number of methods are proposed, including using machine learning and artificial intelligence methods. Approaches to the use of modern machine learning methods to the specifics of the tasks of power distribution of thermal power plants are proposed.

- Possible approaches to the methodology and practical approaches to the technical implementation of the integration of optimization problem algorithms into the PTC application software, their information and technical support, provided that all existing functions of the automated control system are performed simultaneously, are developed.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] E.K. Arakelyan et al., "Ways to increase the intelligence of algorithms for optimal load distribution at CHPP with complex equipment composition", Managing the development of large-scale systems (MLSD-2020), **2**, 904-911, 2020, doi: 10.1109 / MLSD49919.2020.9247651.

[2] E.K. Arakelyan et al., "Approaches to increasing the intelligence of automated control systems of large power plants by solving optimization problems of block and station levels", Managing the development of large-scale systems, **2**, 434-436, 2018, doi: 10.1109/MLSD.2018.8551935.

[3] Y. Tverskoy, "Features and problems of the current stage of development of the creation technology of automated control system of TPP", Thermal Engineering, **10**, 37-44, 2010.

[4] E.K. Arakelyan et al., "Problems of modern automated process control systems based on software and hardware complexes and a possible way to solve them", Vestnik MPEI, **6**, 15-20, 2014.

[5] R.E. Steuer, Multiple Criteria Optimization: Theory, Computations, and Application, John Wiley & Sons, 1986.

[6] Y. Sawaragi, H. Nakayama and T. Tanino. "Theory of Multiobjective Optimization", Mathematics in Science and Engineering, 1985.

[7] I. Gorlach and O. Wessel, «Optimal Level of Automation in the Automotive Industry» Engineering Letters, **16**, 141-149, 2008.

[8] V. Sabanin et al., "Mathematical and software support of the algorithm for correction of measured parameters for calculating technical and economic indicators at a thermal power plant", Vestnik MPEI, **1**, 21-27, 2003.

[9] E.K. Arakelyan et al., Optimization and optimal control, Moscow Publishing House of MPEI, 2008.

[10] E.K. Arakelyan et al., Improving the efficiency and maneuverability of equipment for thermal power plants, MPEI Publishing House, 1993.

[11] E.K. Arakelyan et al., "Evaluation of the optimal level of intelligence of the automated control system of high-power power units based on modern PTC" Managing the development of large-scale systems **5**, 574-576, 2019.

[12] I.N. Daimand et al. "Synthesis of intelligent automated control systems for complex technological processes", Automation in industry, **7**, 2013.