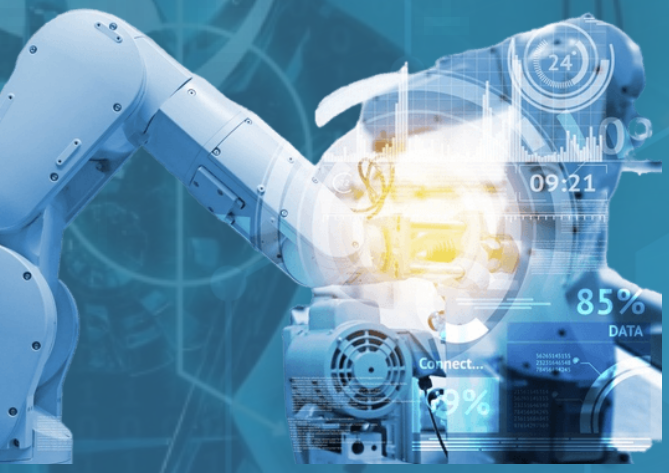
ASTES

Advances in Science, Technology & Engineering Systems Journal



VOLUME 6-ISSUE 4|JUL-AUG 2021

www.astesj.com
ISSN: 2415-6698

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 48 accepted papers in robotics and electronics domains.

Editor-in-chief
Prof. Passerini Kazmersk

ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

Volume 6 Issue 4 **July-August 2021**

CONTENTS

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 6, No. 4, 01-12 (2021)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Exploiting Domain-Aware Aspect Similarity for Multi-Source Cross-Domain Sentiment Classification

Kwun-Ping Lai*, Jackie Chun-Sing Ho, Wai Lam

Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong, 999077, China

ARTICLE INFO

Article history:
Received: 01 May, 2021
Accepted: 15 June, 2021
Online: 10 July, 2021

Keywords:
Domain-aware topic model
Topic-attention network
Adversarial training
Artificial neural networks

ABSTRACT

We propose a novel framework exploiting domain-aware aspect similarity for solving the multi-source cross-domain sentiment classification problem under the constraint of little labeled data. Existing works mainly focus on identifying the common sentiment features from all domains with weighting based on the coarse-grained domain similarity. We argue that it might not provide an accurate similarity measure due to the negative effect of domain-specific aspects. In addition, existing models usually involve training sub-models using a small portion of the labeled data which might not be appropriate under the constraint of little labeled data. To tickle the above limitations, we propose a domain-aware topic model to exploit the fine-grained domain-aware aspect similarity. We utilize the novel domain-aware linear layer to control the exposure of various domains to latent aspect topics. The model discovers latent aspect topics and also captures the proportion of latent aspect topics of the input. Next, we utilize the proposed topic-attention network for training aspect models capturing the transferable sentiment knowledge regarding particular aspect topics. The framework finally makes predictions according to the aspect proportion of the testing data for adjusting the contribution of various aspect models. Experimental results show that our proposed framework achieves the state-of-the-art performance under the constraint of little labeled data. The framework has 71% classification accuracy when there are only 40 labeled data. The performance increases to around 82% with 200 labeled data. This proves the effectiveness of the fine-grained domain-aware aspect similarity measure.

1 Introduction

Online shopping becomes more and more popular during the pandemic. Product reviews serve as an important information source for product sellers to understand customers, and for potential buyers to make decisions. Automatically analyzing product reviews therefore attracts people's attention. Sentiment classification is one of the important tasks. Given sufficient annotation resources, supervised learning method could generate promising result for sentiment classification. However, it would be very expensive or even impractical to obtain sufficient amount of labeled data for unpopular domains. Large pre-trained model, such as the Bidirectional Encoder Representations from Transformers model (BERT) [1], could be an universal way to solve many kinds of problems without exploiting the structure of the problem. In [2], the author apply large pre-trained model to handle this problem task, which has sufficient labeled data only in source domain but has no labeled data in target domain, with fine tuning on source domain and predicting on target domain. In [3], the author train the large pre-trained model using various sentiment related tasks and show that the model could directly apply to the target domain even without the fine-tuning stage. However, these large pre-trained models do not consider the structure of the problem and they have certain hardware requirement that might not be suitable in some situations. We focus on smaller models, which have a few layers, in this work in order to handle the constraint of little labeled data[1]. Besides using the gigantic pre-trained model, domain adaptation (or cross-domain) [4, 5] attempts to solve this problem by utilizing the knowledge from the source domain(s) with abundant annotation resources and transfers the knowledge to the target domain. This requires the model to learn transferable sentiment knowledge by eliminating the domain discrepancy problem. Domain adversarial training [6, 7] is an ef-

*Corresponding Author: Kwun-Ping Lai, Email: kplai@se.cuhk.edu.hk

[1]To give a brief comparison of our proposed framework and the large pre-trained model, we present the performance of the standard BERT-Large model in the experiment section. We ignore other variants of the large pre-trained models as they are not the major focus of this work.

fective method to capture common sentiment features which are useful in the target domain. Various works using domain adversarial training [8]–[11] achieve good performance for single-source cross-domain sentiment classification. It could be also applied to the large pre-trained model to further boost the performance [12]. Moreover, it is quite typical that multiple source domains are available, the model might be exposed to a wider variety of sentiment information and the amount of annotation requirement for every single domain would be smaller. A simple approach is to combine the data from multiple sources and form a new combined source domain. Existing models tackling single-source cross-domain sentiment classification mentioned above could be directly applied to this new problem setting after merging all source domains. However, the method of combining multiple sources does not guarantee a better performance than using only the best individual source domain [13, 14]. Recent works measure the global domain similarity [15]–[17], i.e. domain similarity between the whole source and target domain, or instance-based domain similarity [18]–[21], i.e. domain similarity between the whole source domain and every single test data point. We observe that these approaches are coarse-grained and ignore the fine-grained aspect relationship buried in every single domain. Domain-specific aspects from the source domain might have negative effect in measuring the similarity between the source domain and the target domain, or the single data point. For instance, we would like to predict the sentiment polarity of some reviews from the Kitchen domain and we have available data from the Book, and the DVD domain. Intuitively, the global domain similarity might not have much difference as both of them are not similar to the target. However, reviews related to the cookbook aspect from the Book domain, or reviews talking about cookery show from the DVD domain might contribute more to the prediction of Kitchen domain. Discovering domain-aware latent aspects and measuring the aspect similarity could be a possible way to address the problem. Based on this idea, we introduce the domain-aware aspect similarity measure based on various discovered domain-shared latent aspect topics using the proposed domain-aware topic model. The negative effect of domain-specific aspects could be reduced.

Existing models measuring domain similarity have another drawback. They usually train a set of expert models with each using a single source domain paired with the target domain. Then, the domain similarity is measured to decide the weighting of each expert model. Another way is to select a subset of data from all source domains which are similar to the target data. We argue that these approaches are not suitable under the constraint of little labeled data as each single sub-model is trained using a small portion of the limited labeled data which might obtain a heavily biased observation. The performance under limited amount of labeled data is underexplored for most of existing methods as they require considerable amount of labeled data for training. In [22], the author study the problem setting applying the constraint. However, they assume equal contribution for every source domain. We study the situation under the constraint of little labeled data and at the same time handling the contribution of source domains using fine-grained domain-aware aspect similarity.

To address the negative effect of domain-specific aspects during the domain similarity measure, and also the limitation of the constraint of little labeled data, we propose a novel framework exploiting domain-aware aspect similarity for measuring the contribution of each aspect model representing the captured knowledge of particular aspects. It is capable of working under the constraint of little labeled data. Specifically, the framework consists of the domain-aware topic model for discovering latent aspect topics and inferring the aspect proportion utilizing a novel aspect topic control mechanism, and the topic-attention network for training multiple aspect models capturing the transferable sentiment knowledge regarding particular aspects. The framework makes predictions using the measured aspect proportion of the testing data, which is a more fine-grained measure than the domain similarity, to decide the contribution of various aspect models. Experimental results show that the proposed domain-aware aspect similarity measure leads to a better performance.

1.1 Contributions

The contributions of this work are as follows:

- We propose a novel framework exploiting the domain-aware aspect similarity to measure the contribution of various aspect models for predicting the sentiment polarity. The proposed domain-aware aspect similarity is a fine-grained measure which is designed to address the negative effect of domain-specific aspects existing in the coarse-grained domain similarity measure.
- We present a novel domain-aware topic model which is capable of discovering domain-specific and domain-shared aspect topics, together with the aspect distribution of the data in an unsupervised way. It is achieved by utilizing the proposed domain-aware linear layer controlling the exposure of different domains to latent aspect topics.
- Experimental results show that our proposed framework achieves the state-of-the-art performance for the multi-source cross-domain sentiment classification under the constraint of little labeled data.

1.2 Organization

The rest of this paper is organized as follows. We present related works regarding cross-domain sentiment classification in Section 2. We describe the problem setting and our proposed framework in Section 3. We conduct extensive experiments and present results in Section 4. Finally, we talk about limitations and furture works in Section 5, and summarize our work in Section 6.

2 Related Works

Sentiment analysis [23]–[25] is the computational study of people's opinions, sentiments, emotions, appraisals, and attitudes towards entities [26]. In this work, we focus on textual sentiment data which is based on review of products, and the classification of the sentiment polarity of reviews. We first present the related works of single-source cross-domain sentiment classification. Next, we further extend to multiple-source case.

2.1 Single-Source Cross-Domain Sentiment Classification

Early works involve the manual selection of pivots based on predefined measures, such as frequency [27], mutual information [5, 28] and pointwise mutual information [29], which might have limited accuracy.

Recently, the rapid development of deep learning provides an alternative for solving the problem. Domain adversarial training is a promising technique for handling the domain adaptation. In [8], the author make use of memory networks to identify and visualize pivots. Besides pivots, [9] also consider non-pivot features by using the NP-Net network. In [10], the author combine external aspect information for predicting the sentiment.

Large pre-trained models attract people's attention since the BERT model [1] obtains the state-of-the-art performance across various machine learning tasks. Researchers also apply it on the sentiment classification task. Transformer-based models [2, 12, 3] utilize the amazing learning capability of the deep transformer structure to learn a better representation for text data during the pre-training stage and adapt themselves to downstream tasks (sentiment classification in our case) using fine tuning. However, we argue that the deep transformer structure has been encoded with semantic or syntactic knowledge during the pre-training process which makes the direct comparison against shallow models unfair. It also has certain hardware requirement which hinders its application in some situations.

Methods mentioned above focus on individual source only and they do not exploit the structure among domains. Although we can still directly apply these models to solve the problem by either training multiple sub-models and averaging predictions, or merging all source domains into a single domain, having a performance better than using only the single best source is not guaranteed. Therefore, exploring the structure or relationship among various domains is essential.

2.2 Multi-Source Cross-Domain Sentiment Classification

Early works assuming equal contribution for every source domain [30]–[32] could be a possible approach to handle the relationship between source domains and the target. Other solutions try to align features from various domains globally [33]–[22]. However, the source domain with higher degree of similarity to the target domain contributing more during the prediction process is a reasonable intuition. These methods fail to capture the domain relation. Recent works try to measure domain contribution in order to further improve the performance.

Researchers propose methods to measure the global domain similarity [15]–[17], i.e. the domain similarity between the whole source and target domain, or the instance-based domain similarity [18]–[21], i.e. the domain similarity between the whole source domain and every single test data point. In [15], the author measure the domain similarity using the proposed sentiment graph. In [17], the author employ a multi-armed bandit controller to handle the dynamic domain selection. In [18], the author compute the attention weight to decide the contribution of various already trained expert models. [20] also utilize the attention mechanism to assign importance weights. They incorporate a Granger-causal objective in their mixture of experts training. The total loss measuring distances of attention weights from desired attributions based on how much the inclusion of each expert reduces the prediction error. Maximum Cluster Difference is used in [19] as the metric to decide how much confidence to put in each source expert for a given example. In [21], the author utilize the output from the domain classifier to determine the weighting of a domain-specific extractor.

These methods measure the coarse-grained domain relation and ignore the fine-grained aspect relationship buried in every single domain. In addition, these methods do not consider the constraint of limited labeled data, which is the main focus of this work.

3 Model Descriptions

3.1 Problem Setting

The problem setting consists of the source domain group D_s and the target domain D_t. The source domain group has m domains $\{D_{s_k}\}|_{k=1}^{m}$ while there is only one target domain. For each source domain, we have two sets of data: i) the labeled data $L = \{x_i^l, y_i^l\}|_{i=1}^{n_L}$ and ii) the unlabeled data $U = \{x_j^u, d_j\}|_{j=1}^{n_U}$ where n_L and n_U are the number of data of labeled and unlabeled data respectively, and d_j is the augmented domain membership indicator. Note that y_i is the sentiment label for the whole review x_i and we do not have any fine-grained aspect-level information. The kth source domain can be written as $D_{s_k} = \left\{L_{s_k} = \{x_i^{l,s_k}, y_i^{s_k}\}|_{i=1}^{n_{L_{s_k}}}, U_s = \{x_j^{u,s_k}, d_j^{s_k}\}|_{j=1}^{n_{U_{s_k}}}\right\}$. The data of the target domain has similar structure except that we do not have the sentiment label, i.e. $D_t = \left\{\{x_i^t\}|_{i=1}^{n_t}, U_t = \{x_j^{u,t}, d_j^t\}|_{j=1}^{n_{U_t}}\right\}$ respectively. $n_{L_{s_k}}$ is the number of labeled data and they are the same for all k. We set all $d_*^{s_k}$ to k and all d_*^t to $m+1$. The objective of the multi-source cross-domain sentiment classification is to find out a best mapping function f so that given the training data $T = \{D_{s_1}, D_{s_2}, ..., D_{s_m}, D_t\}$, the aim is to predict the label of the target domain labeled data $\overline{\mathbf{y}}^{\mathbf{t}} = f(\mathbf{x}^{\mathbf{t}})$.

3.2 Overview of Our Framework

We describe our proposed framework exploiting domain-aware aspect similarity. Specifically, there are two components: i) the domain-aware topic model discovering domain-aware latent aspect topics, ii) the topic-attention network identifying sentiment topic capturing the transferable aspect-based sentiment knowledge. The first component captures both domain-specific and domain-shared latent aspect topics, and infers the aspect distribution of each review. It is an unsupervised model that utilizes only the unlabeled data. It is analogous to the standard topic model which discovers latent topics as well as topic distributions. However, the standard topic model is not capable of controlling discovered latent topics. Our proposed domain-aware topic model is capable of separating discovered latent topics into two groups: we name them as domain-specific aspect topics and domain-shared aspect topics. The topic control is achieved by using the domain-aware linear layer described in the latter subsection. Specifically, the model discover $n_{\texttt{spec}}$ domain-specific aspect topics for every domain, and $n_{\texttt{share}}$ domain-shared

aspect topics which are shared among all domains. Each review has a $n_{\text{spec}} + n_{\text{share}}$ dimensional aspect distribution with the first n_{spec} dimension corresponding to domain-specific aspect topics and the last n_{share} dimension corresponding to domain-share aspect topics. Discovered aspect topics and inferred aspect distributions have three important functions:

- By considering only domain-shared aspect topics, the negative effect of domain-specific aspect topics could be minimized for measuring the contribution during the inference process.
- The overall aspect distribution of the testing data reveals the importance of each discovered aspect topic following the assumption that the topic appearing more frequent is more important for the target domain.
- The aspect distribution of the unlabeled data could be used for picking reviews with a high coverage of a particular set of aspect topics.

Based on the domain-shared aspect distribution of the target domain, we divide discovered domain-shared aspect topics into groups with each group having unlabeled reviews from all domains with high aspect proportion forming the training dataset for the second component. Specifically, we divide domain-shared aspect topics into groups based on the overall aspect distribution of the target domain. We aim at separating aspect topics and train an expert model for each group of aspects. Each aspect model focuses on a particular set of aspects so as to boost the learning capability of that set of fine-grained aspect topics. Therefore, we need to construct the dataset carrying the information related to selected aspect topics. We select the unlabeled data from all domains with high aspect proportion of a particular set of aspect topics to form the aspect-based training dataset.

Each of the aspect-based training dataset guides the next component to focus on the corresponding aspect group and identify the related transferable sentiment knowledge. The obtained training dataset is jointly trained with the limited labeled data using the topic-attention network to generate an aspect model for each aspect-based training dataset. The topic-attention network is a compact model which is designed to work effectively under limited training data. The topic-attention network captures two topics simultaneously: i) the sentiment topic and ii) the domain topic. The sentiment topic captures the transferable sentiment knowledge which could be applied to the target domain. The domain topic serves as an auxiliary training task for constructing a strong domain classifier which helps the sentiment topic to identify domain-independent features by using domain adversarial training. These two topics are captured by the corresponding topical query built in the topic-attention layer. These topical queries are learnt automatically during the training process. The limited labeled data works with the sentiment classifier to control the knowledge discovery related to sentiment (sentiment topic captures sentiment knowledge while domain topic does not), while the unlabeled data works with the domain classifier to control the knowledge discovery related to domain. Finally, the framework makes predictions using various aspect models with contribution defined by the aspect distribution of the testing data. For example, if the testing data has a higher coverage regarding aspect group 1, then naturally the prediction made by the aspect model of group 1 should contribute more to the finally prediction as intuitively that aspect model would have more related sentiment knowledge to make judgement. We believe this fine-grained latent aspect similarity would provide a more accurate sentiment prediction than the traditional coarse-grained domain similarity due to the fact that we eliminate the negative effect of domain-specific aspects when measuring the similarity between the testing data and the expert models.

We first describe the architecture of the two components. Then, we describe the procedure of inferring the sentiment polarity of reviews of the target domain.

3.3 Domain-Aware Topic Model

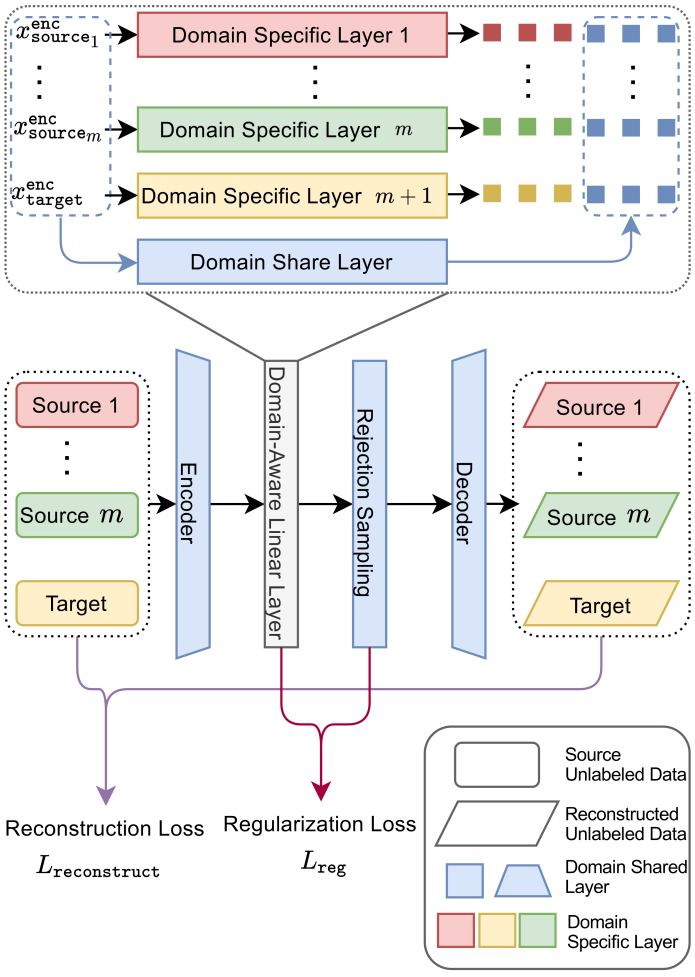


Figure 1: Diagram depicting the proposed domain-aware topic model. The middle part provides a high-level overview of the proposed domain-aware topic model. The model aims at inferring the dense representation of the unlabeled data from all domains in terms of aspect topic proportion. The model discovers domain-specific aspect topics and domain-shared aspect topics utilizing the domain-aware linear layer which is illustrated in the upper part of the figure. The model is trained by minimizing the reconstruction loss calculated by using the input data and the reconstructed data, and the regularization loss based on the inferred α and the predefined Dirichlet prior.

The domain-aware topic model follows the mechanism of the variational autoencoder framework (VAE) [35] which utilizes the encoder for inferring the latent variable (the Dirichlet prior α in our case representing the expected aspect distribution) and the de-

coder for reconstructing the input. Researchers try to apply the VAE network for achieving functionalities of standard topic model in a neural network way, such as inferring the topic proportion of the input and the word distribution of each topic. This provides some advantages such as reducing the difficulty of designing the inference process, leveraging the scalability of neural network, and the easiness of integrating with other neural networks [36]. However, the standard VAE using Gaussian distribution to model the latent variable might not be suitable for text data due to the sparseness of the text data. The Dirichlet distribution used in the topic model [37] has a problem of breaking the back-propagation. Calculating the gradient for the sampling process from the Dirichlet distribution is difficult. Researchers propose approximation methods [38, 39, 40, 41] in order to apply Dirichlet distribution to the neural topic model. We follow the rejection sampling method [42] in this work. Although discovered topics might carry extra information which might be helpful for identifying the hidden structure of the text data, it is not intuitive for applying this information to help the sentiment classification task. We introduce the domain-aware linear layer for controlling the formation of domain-specific and domain-shared aspect topics. To the best of our knowledge, we do not find any similar aspect topic control layer applied for multiple-source cross-domain sentiment classification in related works. The domain-aware linear layer identifies both domain-specific aspect topics and domain-shared aspect topics. We utilize domain-shared aspect topics only which could provide a more accurate measure for calculating the similarity. In addition, the inferred aspect topic proportion is used for constructing the aspect-based training dataset, and determining the level of contribution of each aspect model. Details of the architecture of the model are described below.

3.3.1 Encoder

The input of the encoder is the bag of words of the review. Specifically, we count the occurrence of each vocabulary in each review and we use a vector of dimension V to store the value. This serve as the input representing the review. The encoder is used to infer the Dirichlet prior of the aspect distribution of the input. The bag-of-words input is first transformed using a fully connected layer with RELU activation followed by a dropout layer.

$$\texttt{Layer}^{\texttt{enc}}(x) = \texttt{Dropout}\Big(\texttt{RELU}(W^{\texttt{enc}}x + b^{\texttt{enc}})\Big) \tag{1}$$

3.3.2 Domain-Aware Linear Layer

Next, the output is fed into the domain-aware linear layer for obtaining domain-specific and domain-shared features. The domain-aware linear layer has $m + 1$ sub-layers including m domain-specific sub-layers handling the feature extraction of the corresponding domain and 1 domain-shared sub-layer handling all domains as follows:

$$\texttt{Layer}^{\texttt{DL}}_{d_x}(x) = [W^{\texttt{DL}}_{d_x}x + b^{\texttt{DL}}_{d_x}; W^{\texttt{DL}}_{\texttt{shared}}x + b^{\texttt{DL}}_{\texttt{shared}}] \tag{2}$$

where d_x is the domain ID of the input x, and [;] represents the operation of vector concatenation. The output $x^{\texttt{DL}}$ is batch normalized and passed to the `SoftPlus` function to infer the Dirichlet prior α of the aspect distribution. To make sure each value in α is greater than zero, we set all values smaller than $\alpha_{\texttt{min}}$ to $\alpha_{\texttt{min}}$.

$$\alpha = \texttt{max}\Big(\texttt{SoftPlus}\Big(\texttt{BatchNorm}(x^{\texttt{DL}})\Big), \alpha_{\texttt{min}}\Big) \tag{3}$$

We use the rejection sampling method proposed in [42] to sample the aspect distribution z and at the same time it allows the gradient to back-propagate to α.

3.3.3 Decoder

The decoder layer is used for reconstructing the bag-of-word input. The sampled aspect distribution z is transformed by the domain-aware linear layer as follows:

$$\texttt{Layer}^{\texttt{dec}}(x) = [W^{\texttt{dec}}_{d_x}x; W^{\texttt{dec}}_{\texttt{shared}}x] \tag{4}$$

The output $x^{\texttt{dec}}$ is batch normalized and passed to the log-softmax function representing the log probability of generating the word.

$$y = \texttt{ln}\Big(\texttt{Softmax}\Big(\texttt{BatchNorm}(x^{\texttt{dec}})\Big)\Big) \tag{5}$$

3.3.4 Loss Function

The loss function includes the regularization loss and the reconstruction loss. The regularization loss measures the difference of the log probability of generating the aspect distribution z between two prior, α and $\overline{\alpha}$ as follows:

$$L_{\texttt{reg}} = lnP(z|\alpha) - lnP(z|\overline{\alpha}), \quad P(y|x) \sim Dir(x) \tag{6}$$

where α is inferred by the model and $\overline{\alpha}$ is the predefined Dirichlet prior. The reconstruction loss is the log probability of generating the bag-of-word input calculated as follows:

$$L_{\texttt{reconstruct}} = -\sum_{i=1}^{V} y_i x_i \tag{7}$$

where V is the vocabulary size, y_i is the log probability of the ith word generated by the model, and x_i is the count of the ith word in the input.

3.4 Topic-Attention Network

The topic-attention network aims at capturing the transferable sentiment knowledge from the limited labeled data of various source domains. To achieve this goal, the network is designed to capture two topics simultaneously: i) the sentiment topic, and ii) the domain topic. The sentiment topic identifies the transferable sentiment knowledge from the input data while the domain topic helps to train a strong domain classifier. We use the technique of domain adversarial training [6, 7, 43] to maintain the domain independence of the sentiment topic. However, instead of using the standard gradient reversal layer, we use the adversarial loss function [22] to achieve the same purpose with a more stable gradient and a faster convergence. The model has two training tasks: i) the sentiment task for identifying the sentiment knowledge, and ii) the auxiliary domain task for training a strong domain classifier. The adversarial loss function is applied to the domain classifier output of the sentiment topic and the sentiment classifier output of the domain topic to hold the indistinguishability property of these two topics. Details of the architecture of the model is described below.

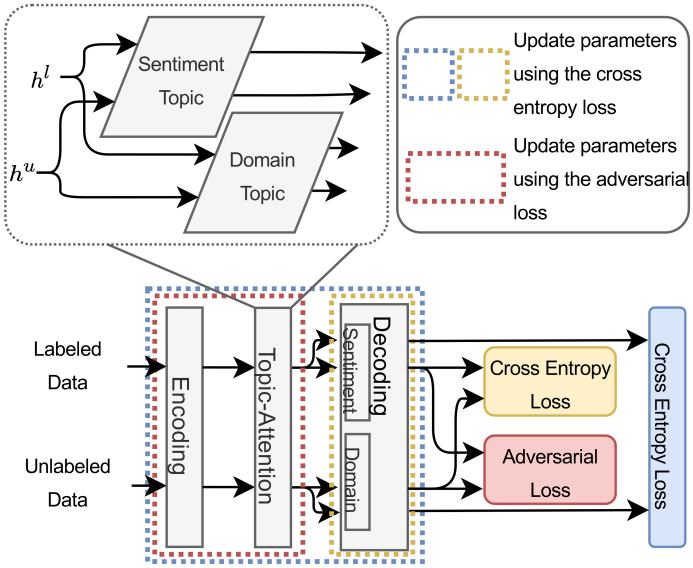


Figure 2: Diagram depicting the proposed topic-attention network. The bottom part provides a high-level overview of the proposed topic-attention network. The network captures two topics, i.e. the sentiment topic and the domain topic, from the review data and classifies the sentiment polarity and the domain membership. The adversarial loss maintains the indistinguishability of topics (domain indistinguishability of the sentiment topic and sentiment indistinguishability for the domain topic). Therefore, the sentiment knowledge captured by the sentiment topic could be transferred to the target domain. The colored dashed boxes show the scope of updating parameters for the corresponding loss.

3.4.1 Encoding Layer

Each word is mapped to the corresponding embedding vector and then transformed by a feed-forward layer with tanh activation for obtaining the feature vector h.

$$h = \texttt{tanh}\left(W^{\text{enc}}\,\texttt{Embedding}(x) + b^{\text{enc}}\right) \tag{8}$$

3.4.2 Topic-Attention Layer

The feature vector h_i of the ith word is re-weighted by the topical attention weight β_i^k calculated as follows:

$$\beta_i^k = \frac{m_i e^{q_k^\top h_i}}{\sum_{i'=1}^{k_w} m_{i'} e^{q_k^\top h_{i'}}} \times n_m \tag{9}$$

where k indicates the topic (either sentiment or domain topic), m_i is the word-level indicator indicating whether the ith position is a word or a padding, n_m is the number of non-padding words, and q_k is the topical query vector for topic k learnt by the model. Note that we have two topical query vectors representing two topics. The topical feature vector t_i^k of the topic k and the review i is obtained by summing feature vectors weighted by the corresponding topical attention weight β_*^k as follows:

$$t_i^k = \sum_{j=1}^{W_i} \beta_j^k h_j \tag{10}$$

where W_i is the number of words in review i. t_i^k represents extracted features of the review by topic k.

3.4.3 Decoding Layer

This layer consists of two decoders with each handling one training task, namely the sentiment decoder and the domain decoder for classifying the sentiment polarity and the domain membership respectively. Note that the review feature vector of labeled data is passed to the sentiment decoder while the unlabeled data of the aspect groups is passed to the domain decoder. Although we use the same t^k to represent the input feature vector in the following two equations, they are actually representing the review features captured from the labeled data, and unlabeled data respectively. Specifically, the review feature vector is linearly transformed and passed to the `Softmax` function for obtaining a valid probability distribution.

$$p^{\text{sen},k} = \texttt{Softmax}(W^{\text{sen}} t^k + b^{\text{sen}}) \tag{11}$$

$$p^{\text{dom},k} = \texttt{Softmax}(W^{\text{dom}} t^k + b^{\text{dom}}) \tag{12}$$

Note that there are four outputs generated by the decoding layer, including two outputs generated by the captured features of two topics passing to the sentiment decoder, and similarly the remaining two generated by the domain decoder. The two topics are sentiment and domain topic, i.e. $k = \{\texttt{sen}, \texttt{dom}\}$. Therefore, the four outputs are: $p^{\text{sen,sen}}$ and $p^{\text{sen,dom}}$ coming from the labeled data passing to the sentiment decoder (the first superscript) having specific features captured by the sentiment and domain topic (the second superscript) respectively, and $p^{\text{dom,sen}}$ and $p^{\text{dom,dom}}$ coming from the unlabeled data passing to the domain decoder having specific features captured by the corresponding topic.

3.4.4 Loss Function

We use the standard cross entropy loss to measure the classification performance:

$$L^{\text{sen},k} = -\frac{1}{n_L} \sum_{i=1}^{n_L} \ln p_{i,s_i}^{\text{sen},k} \tag{13}$$

$$L^{\text{dom},k} = -\frac{1}{n_U} \sum_{i=1}^{n_U} \ln p_{i,d_i}^{\text{dom},k} \tag{14}$$

where s_i and d_i are the class indicator specifying the sentiment polarity or the domain membership of the ith training data, and $p_{i,c}^*$ is the predicted probability regarding the cth class. Therefore, we have four cross entropy losses. The loss generated by the sentiment decoder from the sentiment topic and the loss generated by the domain decoder from the domain topic are used to update all parameters of the model using back-propagation. The remaining two are used to update the parameters of the decoding layer only. We introduce the adversarial loss function for doing adversarial training for both tasks as follows:

$$f_{adv}(p) = \sum_{i=1}^{c} (p_i - \frac{1}{c})^2 \tag{15}$$

where c is the number of classes and p_i is the predicted probability for the class i. Note that c for sentiment task is 2 while it is $m + 1$ for the domain task. We use the probability distributions generated by the sentiment decoder from the domain topic $p^{\texttt{sen,dom}}$, and by the domain decoder from the sentiment topic $p^{\texttt{dom,sen}}$, to calculate the adversarial losses, which are used to update the parameters of the encoding layer and the topic-attention layer.

3.5 Training Strategy

We first train the domain-aware topic model using the unlabeled data X^u from all domains. The model is then used for predicting the aspect proportion of the unlabeled data X^u and testing data X^t to obtain α^u and α^t. Note that the domain-aware topic model is an unsupervised model that does not utilize any labeled data from source domains nor target domain. The aspect score θ^t of the target domain is calculated using the mean value of the domain-shared aspect part of α^t over all testing data:

$$\theta^t = \frac{1}{n_t} \sum_{i=1}^{n_t} \alpha_i^t[-n_{\texttt{share}} :] \tag{16}$$

where $\alpha_i^t[-n_{\texttt{share}} :]$ represents the last $n_{\texttt{share}}$ dimensions of the vector α_i^t. Therefore, θ^t is a $n_{\texttt{share}}$ dimensional vector with each value representing the importance score of the corresponding aspect topic for the target domain. We divide the domain-shared aspect topics into k groups based on their importance score using θ_t in descending order. The set $\texttt{topic}_{g_{k'}}$ contains the topic indices of the k'th aspect group. For each group $g_{k'}$, we select top n unlabeled data from all domains based on the aspect topic score of the k'th group $\omega_{k'}^u$, which is the sum of the corresponding domain-shared aspect proportion of the k'th group for the uth review using its discovered aspect proportion:

$$\omega_{k'}^u = \sum_{i \in \texttt{topic}_{g_{k'}}} \alpha^u[i] \tag{17}$$

where $\alpha^u[i]$ represents the value in the ith dimension of α^u.

Next, we train k aspect models using the topic-attention network. For each aspect model, the limited labeled data X^l, Y^l is used for training the sentiment task while the group of selected unlabeled data $g_{k'}$ is used for training the auxiliary domain task. The last step is to utilize the obtained models for predicting the sentiment polarity of all testing data x^t. Let $AM_{k'}$ be the aspect model trained by using the dataset $\{X^l, Y^l, g_{k'}\}$, we denote the sentiment prediction of the sentiment topic generated by the model as $p_{k'}^{t'}$ for the target review $x^{t'}$ as follows:

$$p_{k'}^{t'} = AM_{k'}(x^{t'}) \tag{18}$$

Finally, we combine the sentiment predictions of the sentiment topic generated by all aspect models having each contributes according to the aspect proportion of the testing data to obtain the final prediction:

$$p^{t'} = \sum_{i=1}^{k} \omega_i^{t'} p_i^{t'} \tag{19}$$

where $\omega_i^{t'}$ is the contribution of the ith aspect model to the final prediction.

4 Experiment

4.1 Experiment Settings

We use the Amazon review dataset [5] for the evaluation of our proposed framework. The Amazon review dataset is a common benchmark for sentiment classification. We use 5 most common domains, namely Book, DVD, Electronics, Kitchen and Video. For each experiment cross, we reserve one domain as the target domain and use others as source domains. There are 5 combinations in total and we conduct experiments on these 5 crosses. For each domain, we follow the dataset setting in [9] collecting 6000 labeled data, with half positive and half negative polarity. We do further sampling to select a subset of the labeled data to fulfill the constraint of little labeled data. We first construct two lists with each having 3000 elements representing the index of the labeled data of positive and negative class respectively. We randomly shuffle the lists and pick first n indices. Next, we select the labeled data based on these indices. In order to have a comparable result for different size of labeled data, we fix the seed number of the random function so that the runs with different size of labeled data would obtain a same shuffle result. Therefore, the run with 20 labeled data contains the 10 labeled data from the run with 10 labeled data, and also another 10 new labeled data. Similarly, the run with 30 labeled data contains the 20 labeled data from the run with 20 labeled data. With this setting, we can directly estimate the effect of adding additional labeled data and compare the performance directly. We continue the process for other source domains. Finally, we construct 5 datasets having 10 to 50 labeled data for each target domain (there are 40 to 200 labeled data in total as there are 4 source domains). The unlabeled dataset includes all unlabeled data from all domains (including the target domain). All labeled data from the target domain is served as the testing data. We run every single run for 10 times and present the average accuracy with standard deviation in order to obtain a reliable result for model comparison.

4.2 Implementation Details

4.2.1 Domain-Aware Topic Model

The Dirichlet prior is set to 0.01. The minimum of inferred prior is set to 0.00001. We set the number of domain-specific and domain-shared topics to 20 and 40 respectively. We divide the domain-shared aspect topics into 5 groups. The domain-aware topic model is trained for 100 warm-up epochs, and stopped after 10 epochs of no improvement.

4.2.2 Topic-Attention Network

We use word2vec[2] embedding [44] to represent each word. We do not further train them to prevent overfitting. The batch size is

[2]It is a distributed representations of words in vector space. It helps various natural language processing task by putting similar words in a closer location.

[3]It is an optimization algorithm with adaptive learning rate. It considers the momentum of the gradient by using the moving average of the gradient. It also uses the moving average of the squared gradient to scale the learning rate of each individual parameter.

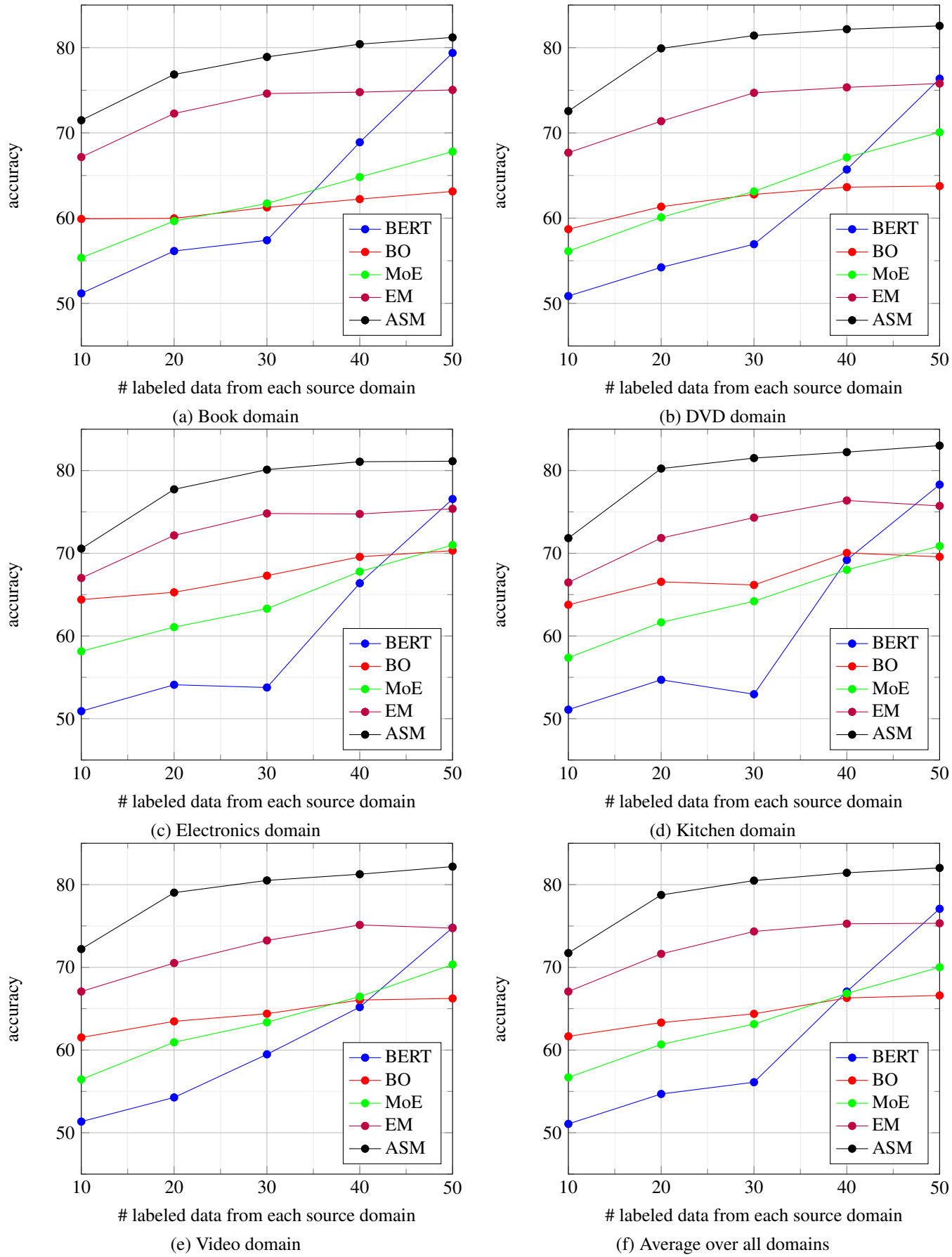


Figure 3: Figure showing the experimental results. Graph (a) to (e) shows the performance of Book, DVD, Electronics, Kitchen and Video domain as target domain respectively. Graph (f) shows the average accuracy of all domains.

set to the number of available labeled data. The topic-attention network is trained for 20 epochs. We use Adam[3] optimizer [45] for back-propagation for both models.

4.3 Evaluation Metric

We use accuracy to measure the evaluate the performance of various models. The target is a binary class. Therefore, correct cases involve the true positive (TP) and true negative (TN). Incorrect cases involve the false positive (FP) and the false negative (FN). The accuracy is calculated as follows:

$$\texttt{accuracy} = \frac{\texttt{TP} + \texttt{TN}}{\texttt{TP} + \texttt{TN} + \texttt{FP} + \texttt{FN}} \tag{20}$$

The average accuracy is calculated by taking the average of accuracy scores of multiple runs.

4.4 Main Results

Models used for performance comparison are as follows:

- BERT [1]: This is the **B**idirectional **E**ncoder **R**epresentations from **T**ransformers model, which is the popular pre-trained model designed to handle various text mining tasks. We use the BERT-Large model with fine tuning using the labeled data to obtain the prediction.
- BO [46]: This model employs **B**ayesian **o**ptimization for selecting data from source domains and transfer the learnt knowledge to conduct prediction on the target domain.
- MoE [19]: This is the **m**ixture **o**f **e**xpert model. It measures the similarity between single test data to every source domain for deciding the contribution of the expert models.
- EM [22]: This is the **e**nsemble **m**odel. It uses various base learners with different focuses on the training data to capture a diverse sentiment knowledge.
- ASM: This is the proposed framework exploiting the domain-aware **a**spect **s**imilarity **m**easure for obtaining a more accurate measure to adjust the contribution of various aspect models focusing on different aspect sentiment knowledge.

Results are presented in Figure 3 and Table 1. We use the classification accuracy as the metric to measure the performance. The proposed framework achieves the best average accuracy among all crosses. Its average performance is 71.73%, 78.75%, 80.49%, 81.43%, and 82.02% for 10, 20, 30, 40 and 50 labeled data cases respectively, or 40, 80, 120, 160 and 200 labeled data cases in total respectively.

4.5 Discussions

Our proposed framework performs substantially better than the comparison models. The proposed framework has an average of 4%, 7%, 6%, 6% and 6% absolute improvement over the second best result for 10, 20, 30, 40 and 50 labeled data cases respectively, or 40, 80, 120, 160 and 200 labeled data cases in total respectively. The variance of the proposed model is comparable to or better than the second best models. The result proves that our proposed framework is very effective for conducting multi-source cross-domain sentiment classification under the constraint of little labeled data. The model can capture transferable sentiment knowledge for predicting the sentiment polarity of the target reviews.

We also do comparative analysis to test the effectiveness of the proposed fine-grained domain-aware aspect similarity measure. It is based on the discovered aspect topics and also the aspect topic proportion for adjusting the contribution of various aspect models. We try to remove these two components to test the performance of the variants. The results are presented in Table 2. The first variant is *rand. select data + avg. pred.*, which means using the unlabeled data selected in a random way instead of using the aspect-based training dataset constructed by the domain-aware topic model, and combining the predictions of various aspect models by averaging them. In other words, the first variant removes both components. The second variant is *avg. pred.*. It keeps the first component (train the aspect models using the aspect-based training dataset) and only removes the second component. Therefore, it assumes equal contribution from various aspect models, just like the first variant. The last one is the proposed framework equipped with both components. Results show that the proposed fine-grained domain-aware aspect similarity measure improves the performance in general except the case having very few labeled data. We think the reason is that the aspect model could not locate the correct aspect sentiment knowledge from the limited data. Thus, the simply averaging the prediction of these biased aspect models would be better than relying on some models. Although the second variant (*avg. pred.*) has a better performance than the full framework in 10 labeled data case, the difference is very small (around 0.18%). Therefore, this comparative analysis could show that the proposed fine-grained domain-aware aspect similarity measure is effective for adjusting the contribution from different discovered aspects.

When comparing with the EM model [22] with similar network architecture but having an equal contribution for the source domains, the result shows that varying the contribution based on the domain-aware aspect similarity leads to a better performance.

We observe that our proposed framework has a small performance gain when giving more labeled training data, besides the case from 10 to 20. The EM model also has similar problem as mentioned in [22]. However, the BERT model [1] has an opposite behavior, which has a steady performance gain. We believe that the reason is due to the compact architecture of the topic-attention network which prevents overfitting the limited labeled data in order to have a better domain adaptation. Increasing the learning capability of the model and at the same time handling domain adaptation could be a future research direction.

5 Limitations and Future Works

The proposed framework involves two separate models handling their own jobs. These models do not share any learning parameters. Many works report that the single model handling various tasks would have a better generalization and thus leads to a better performance. One possible future work might consider integrating both

Table 1: Sentiment classification accuracy of different models

# Labeled Data	Model	Book	DVD	Electronics	Kitchen	Video	Average
10 (40)	BERT	51.17 ± 1.25	50.86 ± 1.06	50.91 ± 1.52	51.09 ± 2.54	51.35 ± 2.21	51.07
	BO	59.90 ± 1.88	58.70 ± 3.66	64.40 ± 2.30	63.77 ± 2.14	61.52 ± 3.28	61.66
	MoE	55.35 ± 3.65	56.12 ± 3.94	58.15 ± 4.87	57.37 ± 4.32	56.45 ± 3.82	56.69
	EM	67.16 ± 5.03	67.68 ± 4.55	67.01 ± 5.29	66.47 ± 5.40	67.08 ± 3.67	67.08
	ASM	**71.48 ± 4.70**	**72.56 ± 5.97**	**70.56 ± 4.44**	**71.83 ± 3.97**	**72.21 ± 4.27**	**71.73**
20 (80)	BERT	56.14 ± 6.14	54.22 ± 5.73	54.10 ± 4.70	54.70 ± 5.27	54.27 ± 5.60	54.69
	BO	59.97 ± 1.85	61.34 ± 2.65	65.28 ± 3.40	66.55 ± 2.54	63.47 ± 3.03	63.32
	MoE	59.65 ± 4.99	60.09 ± 5.66	61.07 ± 5.38	61.65 ± 5.09	60.94 ± 5.24	60.68
	EM	72.27 ± 2.67	71.37 ± 3.97	72.16 ± 2.74	71.84 ± 2.60	70.52 ± 1.38	71.63
	ASM	**76.85 ± 2.41**	**79.91 ± 1.85**	**77.73 ± 3.51**	**80.24 ± 1.58**	**79.03 ± 1.85**	**78.75**
30 (120)	BERT	57.40 ± 5.87	56.94 ± 6.88	53.77 ± 5.40	52.96 ± 2.05	59.48 ± 7.99	56.11
	BO	61.26 ± 2.03	62.78 ± 2.32	67.29 ± 2.89	66.17 ± 3.11	64.39 ± 2.51	64.38
	MoE	61.71 ± 5.47	63.13 ± 5.68	63.30 ± 6.43	64.20 ± 6.06	63.36 ± 5.92	63.14
	EM	74.61 ± 2.54	74.70 ± 1.35	74.81 ± 2.03	74.31 ± 1.10	73.25 ± 1.97	74.34
	ASM	**78.91 ± 2.13**	**81.42 ± 1.07**	**80.11 ± 1.58**	**81.51 ± 1.06**	**80.51 ± 1.75**	**80.49**
40 (160)	BERT	68.90 ± 8.55	65.70 ± 9.25	66.38 ± 8.45	69.19 ± 8.69	65.19 ± 9.77	67.07
	BO	62.23 ± 1.25	63.63 ± 2.45	69.57 ± 2.07	70.04 ± 2.22	66.05 ± 1.75	66.30
	MoE	64.82 ± 4.47	67.12 ± 5.22	67.78 ± 6.05	68.00 ± 5.89	66.46 ± 5.36	66.84
	EM	74.78 ± 1.06	75.34 ± 2.24	74.75 ± 1.89	76.38 ± 1.16	75.13 ± 1.77	75.27
	ASM	**80.41 ± 1.13**	**82.16 ± 0.86**	**81.07 ± 1.38**	**82.23 ± 1.02**	**81.26 ± 1.71**	**81.43**
50 (200)	BERT	79.38 ± 4.79	76.36 ± 8.25	76.56 ± 7.01	78.30 ± 8.08	74.79 ± 9.68	77.08
	BO	63.13 ± 2.63	63.76 ± 2.14	70.32 ± 1.93	69.57 ± 2.56	66.24 ± 1.67	66.60
	MoE	67.80 ± 2.24	70.06 ± 2.59	70.99 ± 2.59	70.87 ± 2.68	70.34 ± 2.76	70.01
	EM	75.04 ± 1.85	75.79 ± 1.96	75.38 ± 2.34	75.73 ± 2.27	74.74 ± 1.42	75.33
	ASM	**81.20 ± 0.86**	**82.56 ± 0.88**	**81.13 ± 1.30**	**83.02 ± 0.87**	**82.18 ± 1.01**	**82.02**

Table 2: Comparative analysis of the proposed framework.

# Labeled Data	Model	Avg. Accuracy
10 (40)	rand. select data + avg. pred.	70.98
	avg.	**71.91**
	ASM	71.73
20 (80)	rand. select data + avg. pred.	76.47
	avg.	78.18
	ASM	**78.75**
30 (120)	rand. select data + avg. pred.	78.03
	avg.	79.75
	ASM	**80.49**
40 (160)	rand. select data + avg. pred.	79.25
	avg.	80.72
	ASM	**81.43**
50 (200)	rand. select data + avg. pred.	79.63
	avg.	81.08
	ASM	**82.02**

models together forming an unified model to take the advantage of multi-task learning. This might further improve the performance for the sentiment classification task.

6 Conclusion

We study the task of multi-source cross-domain sentiment classification under the constraint of little labeled data. We propose a novel framework exploiting domain-aware aspect similarity to identify the contribution of discovered fine-grained aspect topics. This fine-grained similarity measure aims at addressing the negative effect of domain-specific aspects appearing in the existing coarse-grained domain similarity measure, and also the limitation caused by the constraint of little labeled data. Aspect topics are extracted by the proposed domain-aware topic model in an unsupervised way. The topic-attention network then learns the transferable sentiment knowledge based on the selected data related to discovered aspects. The framework finally makes predictions according to the aspect proportion of the testing data for adjusting the contribution of various aspect models. Extensive experiments show that our proposed framework achieves the state-of-the-art performance. The framework achieves a good performance, i.e. around 71%, even though there are only 40 labeled data. The performance reaches around 82% when there are 200 labeled data. This shows that our proposed fine-grained domain-aware aspect similarity measure is very effective under the constraint of little labeled data.

References

[1] 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, doi:10.18653/v1/N19-1423.

[2] B. Myagmar, J. Li, S. Kimura, "Cross-Domain Sentiment Classification With Bidirectional Contextualized Transformer Language Models," IEEE Access, 163219–163230, 2019, doi:10.1109/ACCESS.2019.2952360.

[3] J. Zhou, J. Tian, R. Wang, Y. Wu, W. Xiao, L. He, "SentiX: A Sentiment-Aware Pre-Trained Model for Cross-Domain Sentiment Analysis," in Proceedings of the 28th International Conference on Computational Linguistics, 568–579, 2020, doi:10.18653/V1/2020.COLING-MAIN.49.

[4] S. Ben-David, J. Blitzer, K. Crammer, F. Pereira, "Analysis of Representa-tions for Domain Adaptation," in Advances in Neural Information Processing Systems, **19**, 2007, doi:10.7551/mitpress/7503.003.0022.

[5] J. Blitzer, M. Dredze, F. Pereira, "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in Proceedings of the 45th annual meeting of the association of computational linguistics, 440–447, 2007.

[6] H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, "Domain-adversarial neural networks," NIPS 2014 Workshop on Transfer and Multi-task learning: Theory Meets Practice, 2014, doi:10.1007/978-3-319-58347-1 10.

[7] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, "Domain-adversarial training of neural networks," The Journal of Machine Learning Research, **17**(1), 2096–2030, 2016, doi:10.1007/978-3-319-58347-1 10.

[8] Z. Li, Y. Zhang, Y. Wei, Y. Wu, Q. Yang, "End-to-End Adversarial Memory Network for Cross-domain Sentiment Classification," in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, 2237–2243, 2017, doi:10.24963/IJCAI.2017/311.

[9] Z. Li, Y. Wei, Y. Zhang, Q. Yang, "Hierarchical attention transfer network for cross-domain sentiment classification," in Thirty-Second AAAI Conference on Artificial Intelligence, 5852–5859, 2018, doi:10.1609/AAAI.V33I01.33015773.

[10] K. Zhang, H. Zhang, Q. Liu, H. Zhao, H. Zhu, E. Chen, "Interactive Attention Transfer Network for Cross-Domain Sentiment Classification," in Proceedings of the AAAI Conference on Artificial Intelligence, 5773–5780, 2019, doi:10.1609/aaai.v33i01.33015773.

[11] Q. Xue, W. Zhang, H. Zha, "Improving domain-adapted sentiment classification by deep adversarial mutual learning," in Proceedings of the AAAI Conference on Artificial Intelligence, 9362–9369, 2020, doi:10.1609/AAAI.V34I05.6477.

[12] C. Du, H. Sun, J. Wang, Q. Qi, J. Liao, "Adversarial and Domain-Aware BERT for Cross-Domain Sentiment Analysis," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 4019–4028, 2020, doi:10.18653/v1/2020.acl-main.370.

[13] S. Zhao, B. Li, C. Reed, P. Xu, K. Keutzer, "Multi-source Domain Adaptation in the Deep Learning Era: A Systematic Survey," arXiv preprint arXiv:2002.12169, 2020.

[14] S. Zhao, Y. Xiao, J. Guo, X. Yue, J. Yang, R. Krishna, P. Xu, K. Keutzer, "Curriculum CycleGAN for Textual Sentiment Domain Adaptation with Multiple Sources," in The Web Conference (WWW), 2021, doi:10.1145/3442381.3449981.

[15] F. Wu, Y. Huang, "Sentiment Domain Adaptation with Multiple Sources," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (1: Long Papers), 301–310, 2016, doi:10.18653/v1/P16-1029.

[16] M. Yu, X. Guo, J. Yi, S. Chang, S. Potdar, Y. Cheng, G. Tesauro, H. Wang, B. Zhou, "Diverse Few-Shot Text Classification with Multiple Metrics," in Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, **1** (Long Papers), 1206–1215, 2018, doi:10.18653/v1/N18-1109.

[17] H. Guo, R. Pasunuru, M. Bansal, "Multi-Source Domain Adaptation for Text Classification via DistanceNet-Bandits," in The Thirty-Fourth AAAI Conference on Artificial Intelligence, 7830–7838, 2020, doi:10.1609/AAAI.V34I05.6288.

[18] Y.-B. Kim, K. Stratos, D. Kim, "Domain Attention with an Ensemble of Experts," in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (1: Long Papers), 643–653, 2017, doi:10.18653/v1/P17-1060.

[19] J. Guo, D. Shah, R. Barzilay, "Multi-Source Domain Adaptation with Mixture of Experts," in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 4694–4703, 2018, doi:10.18653/v1/D18-1498.

[20] M. Yang, Y. Shen, X. Chen, C. Li, "Multi-Source Domain Adaptation for Sentiment Classification with Granger Causal Inference," in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 1913–1916, 2020, doi:10.1145/3397271.3401314.

[21] Y. Dai, J. Liu, X. Ren, Z. Xu, "Adversarial Training Based Multi-Source Unsupervised Domain Adaptation for Sentiment Analysis," in The Thirty-Fourth AAAI Conference on Artificial Intelligence, 7618–7625, 2020, doi:10.1609/AAAI.V34I05.6262.

[22] K. Lai, J. C. Ho, W. Lam, "Ensemble Model for Multi-Source Cross-Domain Sentiment Classification with Little Labeled Data," in 2020 IEEE/WIC/ACM International Conference on Web Intelligence (WI), IEEE, 2020, doi:10.1109/WIIAT50758.2020.00038.

[23] K. Shaukat, T. M. Alam, M. Ahmed, S. Luo, I. A. Hameed, M. S. Iqbal, J. Li, M. A. Iqbal, "A Model to Enhance Governance Issues through Opinion Extraction," in 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 0511–0516, 2020, doi:10.1109/IEMCON51383.2020.9284876.

[24] H. Gupta, S. Pande, A. Khamparia, V. Bhagat, N. Karale, et al., "Twitter Sentiment Analysis Using Deep Learning," in IOP Conference Series: Materials Science and Engineering, 012114, 2021, doi:10.1088/1757-899X/1022/1/012114.

[25] A. Badgaiyya, P. Shankarpale, R. Wankhade, U. Shetye, K. Gholap, S. Pande, "An Application of Sentiment Analysis Based on Hybrid Database of Movie Ratings," International Research Journal of Engineering and Technology (IRJET), **8**, 655–665, 2021.

[26] L. Zhang, S. Wang, B. Liu, "Deep learning for sentiment analysis: A survey," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, **8**(4), e1253, doi:10.1002/widm.1253.

[27] J. Blitzer, R. McDonald, F. Pereira, "Domain adaptation with structural correspondence learning," in Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, 120–128, 2006, doi:10.3115/1610075.1610094.

[28] S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, Z. Chen, "Cross-domain sentiment classification via spectral feature alignment," in Proceedings of the 19th International Conference on World Wide Web, 751–760, 2010, doi:10.1145/1772690.1772767.

[29] D. Bollegala, T. Mu, J. Y. Goulermas, "Cross-domain sentiment classification using sentiment sensitive embeddings," IEEE Transactions on Knowledge and Data Engineering, **28**(2), 398–410, 2015, doi:10.1109/TKDE.2015.2475761.

[30] K. Crammer, M. Kearns, J. Wortman, "Learning from Multiple Sources," Journal of Machine Learning Research, **9**(57), 1757–1774, 2008.

[31] S. Li, C. Zong, "Multi-domain Sentiment Classification," in Proceedings of ACL-08: HLT, Short Papers, 257–260, 2008.

[32] P. Luo, F. Zhuang, H. Xiong, Y. Xiong, Q. He, "Transfer learning from multiple source domains via consensus regularization," in Proceedings of the 17th ACM conference on Information and knowledge management, 103–112, 2008, doi:10.1145/1458082.1458099.

[33] H. Zhao, S. Zhang, G. Wu, J. M. Moura, J. P. Costeira, G. J. Gordon, "Adversarial multiple source domain adaptation," in Advances in Neural Information Processing Systems, 8568–8579, 2018.

[34] X. Chen, C. Cardie, "Multinomial Adversarial Networks for Multi-Domain Text Classification," in Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, **1** (Long Papers), 1226–1240, 2018, doi:10.18653/v1/N18-1111.

[35] D. P. Kingma, M. Welling, "Auto-Encoding Variational Bayes," in 2nd International Conference on Learning Representations, ICLR, 2014.

[36] H. Zhao, D. Phung, V. Huynh, Y. Jin, L. Du, W. Buntine, "Topic Modelling Meets Deep Neural Networks: A Survey," arXiv preprint arXiv:2103.00498, 2021.

[37] D. M. Blei, A. Y. Ng, M. I. Jordan, "Latent Dirichlet Allocation," Journal of Machine Learning Research, **3**, 993–1022, 2003, doi:10.1016/B978-0-12-411519-4.00006-9.

[38] M. Figurnov, S. Mohamed, A. Mnih, "Implicit Reparameterization Gradients," in Advances in Neural Information Processing Systems, 439–450, Curran Associates, Inc., 2018.

[39] W. Joo, W. Lee, S. Park, I.-C. Moon, "Dirichlet Variational Autoencoder," CoRR, **abs/1901.02739**, 2019, doi:10.1016/j.patcog.2020.107514.

[40] H. Zhang, B. Chen, D. Guo, M. Zhou, "WHAI: Weibull Hybrid Autoencoding Inference for Deep Topic Modeling," in International Conference on Learning Representations, 2018.

[41] C. Naesseth, F. Ruiz, S. Linderman, D. Blei, "Reparameterization Gradients through Acceptance-Rejection Sampling Algorithms," in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 489–498, 2017.

[42] S. Burkhardt, S. Kramer, "Decoupling Sparsity and Smoothness in the Dirichlet Variational Autoencoder Topic Model," Journal of Machine Learning Research, **20**(131), 1–27, 2019.

[43] Y. Ganin, V. Lempitsky, "Unsupervised Domain Adaptation by Backpropagation," in Proceedings of the 32nd International Conference on International Conference on Machine Learning, 1180–1189, 2015.

[44] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, "Distributed representations of words and phrases and their compositionality," in Advances in Neural Information Processing Systems, 3111–3119, 2013.

[45] D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[46] S. Ruder, B. Plank, "Learning to select data for transfer learning with Bayesian Optimization," in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 372–382, 2017, doi:10.18653/v1/D17-1038.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 6, No. 4, 13-20 (2021)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

A Reconfigurable Stepped Frequency Continuous Wave Radar Prototype for Smuggling Contrast, Preliminary Assessment

Massimo Donelli[*,1,2], Giuseppe Espa[2,3], Mohammedhusen Manekiya[1], Giada Marchi[1] , Claudio Pascucci[4]

[1]*Department of Information Engineering and Computer Science, University of Trento, Trento, 38123, Italy*

[2]*Center for Security and Crime Sciences (CSSC), University of Trento and Verona, Trento, 38120, Italy*

[3]*Department of Economy and Management, University of Trento, Trento, 38120, Italy*

[4]*Guardia di Finanza, Department, Comando Regionale Trentino Alto Adige, Trento, 38120, Italy*

ARTICLE INFO

Article history:
Received: 27 April, 2021
Accepted: 21 June, 2021
Online: 10 July, 2021

Keywords:
Stepped frequency continuous wave radar (SFCW)
Electromagnetic propagation
Microwave systems

ABSTRACT

A reconfigurable Stepped Frequency Continuous Wave (SFCW) radar prototype for supporting the italian financial police to contrast smuggling, is proposed in this work. In particular, the proposed radar can provide information related to the container contents and the presence of false bottoms speeding up the control operations at borders and ports. Moreover, it is able to reveal the presence of people hidden behind reinforced concrete hiding places. Radar resolution is improved by using suitable post-processing method such as the MUltiple Signal Classification (MUSIC) algorithm. Numerical as well as experimental results obtained considering realistic operative scenarios demonstrated the potentialities and capabilities of this system as an effective tool for smuggling contrast. The preliminary experimental results have been obtained using a compact radar prototype equipped with high gain and directivity antennas to cover all the different frequency bands.

1 Introduction

In recent years, globalisation has led to a sharp increase in the smuggling of goods and illegal materials such as drugs, tobacco, and weapons. To contrast smuggling, Guardia di Finanza (GdF), the Italian financial police, increases the controls at customs and ports. However, checking the contents of thousands containers or vehicles requires a lot of time, human resources, and money. New technologies and tools can effectively to support the GdF work and be effective to contrast smuggling. In such a scenario, radar technologies and, particularly, ground penetration (GPR) and through the wall radars (TTR) could represent effective tools for GdF. A scan-A (1D) analysis of a container can provide accurate information concerning the contents, presence of false bottoms or hidden humans without the need to empty the container or vehicle, saving time, money and resources. In order to be useful for these operations, it is necessary to reach a high resolution of about tens centimetres and an unambiguous range of less than ten meters. These performances can be easily obtained with an SFCW radar [1]. Due to their complexity and the high cost of microwave components, SFCW radars had less diffusion for through the wall and ground penetration radar applications than pulsed radars [2]. However, in the last decade, microwave components strongly improved their performances and reduced their dimensions and cost, making the realization of compact SFCW radars more feasible. Recently SWCW radars have been successfully adopted for GPR applications [3, 4], landmine detection [5], surveillance [6], through the wall inspection [7], the monitoring of the structural integrity of engineering structures [8]–[10] or for biomedical applications such as the monitoring of human or animal life signals [11]–[13]. The last application can be very useful for police operation since it can be used to detect fugitives hidden inside improvised hiding places such as bunkers and wall false bottoms or human traffic. SFCW radar capabilities to characterize the electric properties of materials as in [10] can be very useful to identify hidden goods, such as tobacco, drugs, and hidden weapons. To improve the radar performances, such as resolution, detection capabilities, clean the clutter and reduce the background noise, suitable post-processing algorithms can be applied [14]–[18]. This paper is an extension of work originally presented in 2nd Global Power, Energy and Communication Conference (GPECOM) [19] where only preliminary numerical results have been reported. In this work, a low-cost SFCW radar prototype as support, the GdF work

[*]Corresponding Author: Massimo Donelli, Email: massimo.donelli@unitn.it

is developed, fabricated, numerically and experimentally assessed in realistic scenarios. The device is based on the ability to operate at different frequency bands enclosed in a range 100MHz-12GHz.. Based on a Direct Digital Synthesis (DDS) generator, the proposed system is easily reconfigurable in terms of frequency bands, number of steps, and spatial resolution. The device can be programmed as ground, through the wall and life detection radar, and it is suitable for smuggling contrast and rescue operations. The prototype is based on the system reported in [19], and it is capable of taking a single scan-A deep profile measurements using 600 frequency steps in less than 2 seconds with a minimum spatial resolution of 0.005 m, simulating a pulsed radar or as continuous-wave radar (CW) to work like a Doppler radar. The device is too slow for moving targets, but it is perfect for GPR and TTW applications where the targets are fixed. The obtained numerical and experimental preliminary results, related to realistic scenarios and directly provided by the GdF, demonstrated the capabilities and potentialities of the proposed system.

2 Mathematical Formulation

This section reports the mathematical formulation of a stepped frequency continuous wave radar and the formulation of post-processing MUSIC algorithm. A typical SFCW radar schema is shown in Fig. 1. It consists of a sweep generator directly connected with a wideband directive antenna. The receiving section can receive the backscattered electromagnetic wave and aims to provide the in-phase and quadrature (I/Q) signals. The control system directly process the frequency domain signal into a synthesized time-domain signal. In particular, the received baseband I/Q signals are combined into a complex number as follows:

$$\Gamma_m = I_m + jQ_m = A_i e^{j\phi_m} \tag{1}$$

where $\phi_m = -2\pi f_m \tau$, whit f_m is the m-th frequency step and τ is the propagation delay, $I_m = A_i cos(\phi_m)$, and $Q_m = A_m sin(\phi_m)$.

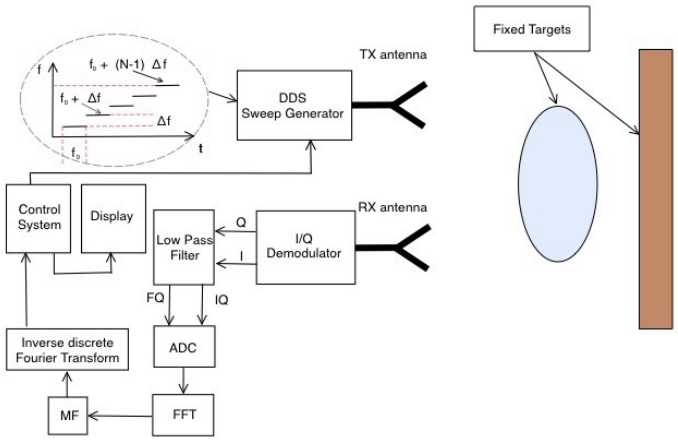


Figure 1: Schema of the SFCW radar.

The sweep generator generates M different sinusoidal tones. The receiver is aimed at collecting all the backscattered signals from the targets . In particular, the I_m/Q_m signals are organized into a complex vector $\hat{\Gamma} = [\Gamma_0, \ldots, \Gamma_{M-1}]$. The inverse discrete Fourier transformer reported in the following equation is then applied to the $\hat{\Gamma}$ vector in order to obtain the conversion from frequency to time domain and better reveal the presence of echoes produced by targets.

$$T_m = \frac{1}{M}\sum_{m=0}^{M-1}\Gamma_m = \frac{1}{M}\sum_{m=0}^{M-1} A_m e^{j\left(\frac{2\pi m_i}{M} - 2\pi\Delta f \frac{2R_t}{v}\right)} \tag{2}$$

SFCW radar strength is its ability to set the resolution properly and unambiguously adapt to the different operative scenarios. The resolution range is provided by the following equation:

$$\Delta R = \frac{v}{2M\Delta f} \tag{3}$$

where v is the electromagnetic wave propagation velocity in the medium. As it can be noticed from (3) the resolution range only depends by the frequency steps number M and the Δf. The following relation can estimate the unambiguous range:

$$R_{max} = \left(\frac{M}{2} - 1\right)\Delta R \tag{4}$$

similarly to equation (3), the unambiguous range can be changed by acting on the step number M and on the resolution range. In most of practical scenarios a A-scan (1D) range profile is enough and it can be obtained by processing the received signal for a given fixed position of the RX and TX antennas. However, it is worth noticing that for all the applications that require a B-scan (2D) representation, such as GPR applications, it is possible repeating the procedure for A-scan (1D) by using different antenna positions.

2.1 The MUSIC post processing algorithm

To properly resolve close targets, suitable post-processing algorithms must be applied to the SFCW radar signals. These signal are very weak and enveloped by the background or by other noise sources. Super-resolution algorithms such as the well-known MUSIC algorithm has been recently successfully used for different practical applications [20]–[24] and the MUSIC algorithm can be very useful to extract the echoes of targets from the background noise. In particular, the I/Q signals collected by the receiver can be expressed by the following vectorial relation:

$$\begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ . \\ . \\ . \\ s_n \end{bmatrix} = \begin{bmatrix} P & e^{-j2\pi f_{-1}\gamma_2} & . & . & . & e^{-j2\pi f_{N-1}\gamma_n} \\ e^{j2\pi f_1\gamma_1} & P & . & . & . & e^{-j2\pi f_{N-2}\gamma_n} \\ e^{j2\pi f_2\gamma_1} & e^{j2\pi f_1\gamma_2} & . & . & . & e^{-j2\pi f_{N-3}\gamma_n} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ e^{j2\pi f_{N-1}\gamma_1} & e^{j2\pi f_{N-2}\gamma_2} & . & . & . & P \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ y_n \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ n_3 \\ . \\ . \\ . \\ n_n \end{bmatrix} \tag{5}$$

that in compact form is $[\boldsymbol{s}] = [\boldsymbol{A}][\boldsymbol{y}] + [\boldsymbol{n}]$, where $[\boldsymbol{n}]$ is the noise vector, $[\boldsymbol{y}]$ is the reflection coefficient vector whose elements are related to the f_N frequency steps, $[\boldsymbol{A}]$ is the $n \times n$ delay matrix. The $n \times n$ signal covariance matrix is given by:

$$[\boldsymbol{R}] = \boldsymbol{S}\boldsymbol{S}^* = ([\boldsymbol{A}][\boldsymbol{y}] + [\boldsymbol{n}])([\boldsymbol{A}][\boldsymbol{y}] + [\boldsymbol{n}])^* \tag{6}$$

where $*$ represent the complex conjugate transpose. The signal covariance matrix $[\boldsymbol{R}]$ can be expressed as:

$$[\boldsymbol{R}] = [\boldsymbol{R}_s] + [\boldsymbol{R}_n] \tag{7}$$

where $[\boldsymbol{R}_n] = \sigma^2 [\boldsymbol{I}]$ is the noise autocorrelation matrix, σ is the noise variance, $[\boldsymbol{I}]$ is the identity matrix and $[\boldsymbol{R}_s]$ is the signal covariance matrix, which can be written as $[\boldsymbol{R}_s] = |A_n|^2 [\boldsymbol{e}_i] \left[\boldsymbol{e}_i^H\right] = P [\boldsymbol{e}_i] \left[\boldsymbol{e}_i^H\right]$, where $|A_n|^2 = P$ is the power of complex exponentials, $[\boldsymbol{e}_i] = \left[1, e^{j\omega_1\gamma_1}, e^{j\omega_1\gamma_2}, ..., e^{j\omega_1\gamma_n}\right]$, with $\omega = 2\pi f$. Since $[\boldsymbol{R}_s]$ is a hermitian matrix then the remaining eigenvectors υ_n will be orthogonal to $[\boldsymbol{e}_i]$ obtaining $\left[\boldsymbol{e}_i^H\right] \upsilon_n = 0; \quad n = 1, 2, ..., N$. The peak position in the time-domain can be obtained by searching the maximum value of the following function.

$$Fmusic(\gamma) = \frac{\boldsymbol{a}(\gamma)^* \boldsymbol{a}(\gamma)}{\boldsymbol{a}(\gamma)^* \boldsymbol{N}_n \boldsymbol{N}_n^* \boldsymbol{a}(\gamma)} \tag{8}$$

where $\boldsymbol{a}(\gamma)$ is the mode vector, obtained from the columns of matrix $[A]$ and $\boldsymbol{N}_n^*$ are the noise eigenvectors.

3 Prototype description

The SFCW radar prototype consists of a programmable digital signal generator (DDS), namely the TG124A (Signal Hound company), with a frequency range from 100KHz up to 12.5GHz and a power range from $-12dB$ up to $-3dB$. The receiver is a digital spectrum analyzer, the SA124B (Signal Hound company), with the following characteristics: frequency range 100KHz, 12.5 GHz, minimum detectable power $-150dBm$. The DDS generator is synchronized with the receiver by means of a coaxial cable that reports the reference local oscillator signal mandatory to obtain the baseband I/Q signals. The DDS generator is directly connected with a transmitting directive radiator, namely a log-periodic wideband antenna for the frequency band comprised between 500.0 MHz-6.0GHz and a high gain horn antenna for the frequency range 8.0GHz-13GHz, the antenna must be manually changed for different operations. The receiver can receive the backscattered electromagnetic wave and provide the I/Q signal that the control system can directly process. Moreover, the SA124B receiver can be easily programmed to sample the I/Q signals and transmit them to an elaboration unit (a high-performance laptop) with a high-speed USB connection. The radar prototype's RF sections are connected with three shielded semi-rigid coaxial cables equipped with sub-miniature type A (SMA) connectors. The SFCW radar prototype photo is shown in Fig. 2. The radar prototype can operate both in bistatic as well as monostatic configuration, as reported in Figs. 1 and 2 respectively. When the radar operates in monostatic configuration, a circulator is connected between the generator, receiver and antenna. The monostatic configuration is considered to reduce the weight of the prototype when operator is involved in through the wall operations (TTW). To reduce the perturbations on the electromagnetic field produsded by the mechanical support, the device was assembled on Teflon dielectric support and placed on a mechanical pedestal to steer the antenna properly. To properly set the radar functionalities, a graphical user interface (GUI) in Matlab language, has been developed. With the GUI, it is possible to customize the radar characteristics to fit with different operative scenarios. In particular, Fig. 3 reports a GUI screen snapshot, representing the tool for through the wall operations. On the left side panel, the radar signals are presented both in time domain and with their spectrogram.

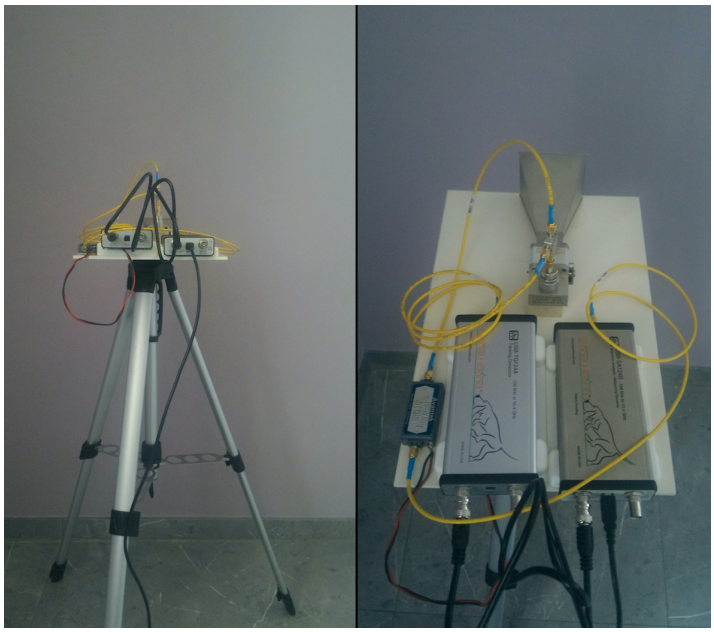


Figure 2: Photo of the assembled SFCW radar prototype.

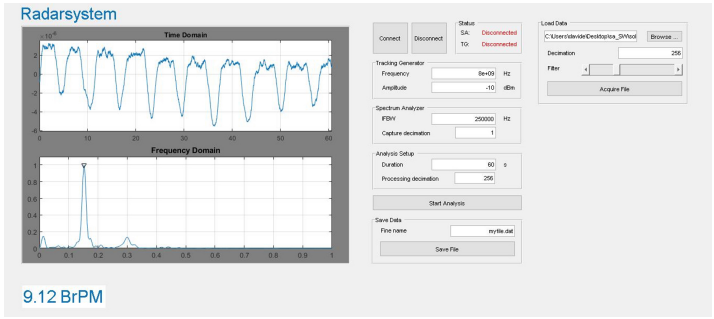


Figure 3: The SFCW radar graphical user interface (GUI).

4 Numerical and experimental assessment

This section is aimed at assessing the capabilities of the proposed SFCW radar. In particular, in sub-section 3.1 are reported the simulated results presented in [19] while sub-section 3.2 reports a preliminary experimental assessment aimed at detecting the human presence behind a wall. This experiment has been carried out to demonstrate the capability of this system to detect fugitives.

4.1 Numerical assessment

In this sub-section, the proposed SFCW radar has been assessed in realistic operative scenarios suggested by the GdF. In particular, the contents of different containers equipped with the false bottom are checked with a fast scan-A (1D) analysis. The radar system, the antenna and related scenarios are accurately simulated with a

customized numerical time-domain electromagnetic engine (FDTD). All the simulations have been performed with quad-cores personal computer, 16 GByte RAM. A gaussian white noise has been added to the original data to simulate better a realistic scenario. The signal to noise ratio of this experiment was $SNR = 5dB$. Different SNR ratio has been tested in a controlled environment, a noise above 5 dB produced very low effects on the simulations demonstrating the robustness of the proposed system. With reference to Figs. 4, 6, and 10, the blue colour represents the container's metallic walls, grey colour represents air-filled, brown colour relates to a homogeneous content of woods or granite blocks. The small orange lines in Fig. 6 represents empty space between wood blocks. The light pink colour in Fig. 10 represents a homogeneous block of tobacco.

4.1.1 *Empty container with false bottom at different distance*

Smugglers usually modify containers stored in ports by inserting false bottoms of different material inside them, the goal is to hide goods and illegal material. In the first experiment, an empty standard container of length 6.058 m, width 2.438 m, and height 2.591 m is considered. The container doors are opened, and the radar pointed precisely in the middle of the aperture at position $x_r = 1.219m$, $y = 1.2945m$.

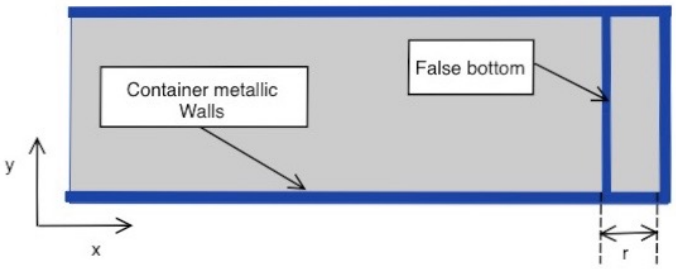


Figure 4: Empty container equipped with a metallic false bottom placed at different distances, A 0.353 m, B 0.453 m, C 0.554 m, D 0.653 m, and E 0.754 m.

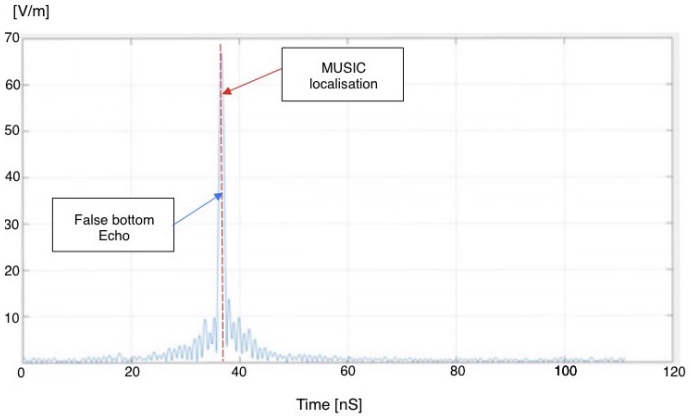


Figure 5: Empty container equipped with a metallic false bottom placed at different r=0.554 m, test case C. Time domain retrieved signal, blue line, MUSIC peak detection dotted red line.

Figure 10 shows the container section equipped with false metallic bottom placed at different distances. The radar parameters are the following: $f_{min} = 1.0$ GHz, $f_{max} = 2.071$ GHz, $M = 120$, $\Delta f = 90MHz$, bandwidth $B = 1.071$GHz. With these parameters an unambiguous range $R_{max} = 16.8m$ and a resolution of $\Delta R = 0.14m$ can be reached. To obtain the position of the false bottom, it is enough a single scan A radar trace. The synthetic data were obtained with an FDTD electromagnetic simulator, namely the gprMax, a well-known gnu software for ground penetration radar (GPR) simulations. To consider the defocusing effects, due to the reflections on container metallic walls, a set of simulations have been performed, considering an empty container and different frequency bands. The goal is to evaluate the effects of reflections and reduce the defocusing effect which can afflict the detection of the peak. Different false bottom positions were considered as reported in Tab. 1 and the synthetic data obtained with gprMax, for each frequency step, are combined and post-processed with the same procedure reported in [25] to obtain the response of a stepped frequency continuous wave radar. Tab. 2 reports the positions of false bottom obtained with a simple peak detection of the inverse discrete Fourier transformer in the time domain. For the sake of comparison, the false bottom position were also estimated with the MUSIC algorithm. Tab. 2 also reports the position errors obtained with the simple peak detection using the MUSIC algorithm. As it can be noticed from the data of Tab. 2 the localization error is very small. The MUSIC algorithm simplifies the data reading and interpretation, keeping the errors below five millimetres. An example of post-processed I/Q data converted into a scan-A time-domain diagram is reported in Fig. 5. As it can be noticed from the data reported in Fig. 5 the peak detected with the standard post processing procedure and with the music algorithm overlaps since both methods localize the peak with a high degree of accuracy. The next experiment considers an empty container with a non-metallic false bottom. This experiment's scope is to assess the capabilities of radar to detect low reflective materials such as wood and plastic wall. Three different commonly used materials, (dry wood, plexiglas, and pressed board wood, whose dielectric characteristics are summarized in Tab. 3) are considered for the wall false bottom realization. The geometry of the considered scenario is shown in Fig. 6. The scan-a diagram's simulation results obtained positioning the electromagnetic source in the middle of container aperture are reported in Figs. 7,8 and 9 respectively. The echo of the container end wall was filtered to visualize the false bottom's echo better. This filtering procedure is quite simple since the real dimensions of a standard container are known. For all considered scenarios, the false bottom made of dielectric material was successfully identified and localized with a high degree of accuracy. In particular, the wood dry false bottom (Fig. 7) is localized with a very low error $\gamma = 2.7mm$. The localization of plexiglas false bottom (Fig. 8) presents a slightly high error $\gamma = 6.8mm$ with respect to the wood dry false bottom wall. The last scenario (Fig. 9) provide the worst results: the localization error $\gamma = 16.1mm$, even though is quite high with respect to the two previous scenarios; still satisfactory.

Table 1: False bottom positions.

	A	B	C	D	E
r [m]	0.353	0.453	0.554	0.653	0.754

Table 2: Retrieved false bottom positions with peak detection and MUSIC algorithm.

Scenario	Peak Pos	MUSIC Pos.	Err. Peak	Err. MUSIC
A	5.700 m	5.701m	5.0 mm	3.5 mm
B	5.600 m	5.602 m	4.9 mm	3.3 mm
C	5.499 m	5.500 m	3.9 mm	3.4 mm
D	5.399 m	5.402 m	4.4 mm	3.3 mm
E	5.302 m	5.301 m	3.3 mm	3.0 mm

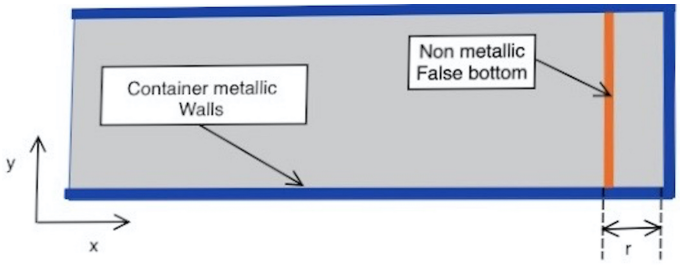


Figure 6: Empty container equipped with a non metallic false bottom placed at distance $r = 5.40m$.

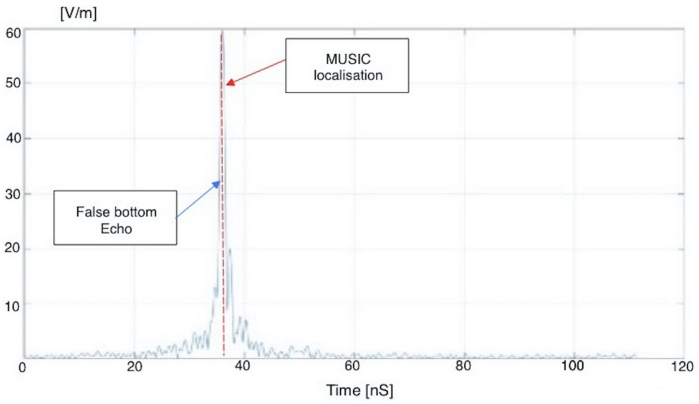


Figure 7: Empty container with a wood dry false bottom placed at distance $r = 5.40m$. Retrieved echo radar signal and MUSIC peak estimation.

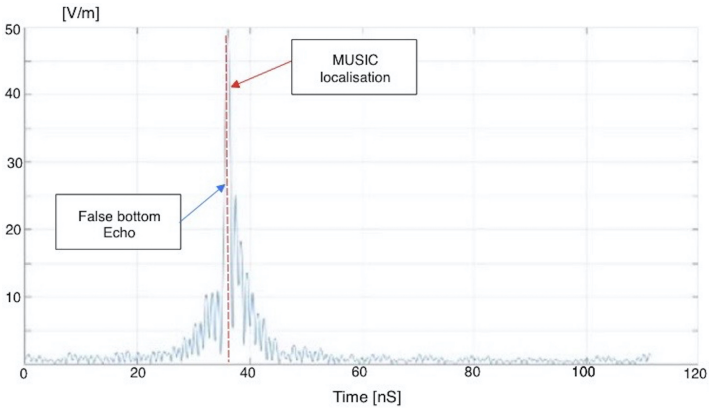


Figure 8: Empty container with a plexiglas false bottom placed at distance $r = 5.40m$.. Retrieved echo radar signal and MUSIC peak estimation.

Table 3: Dielectric characteristics of wall materials used for the false bottom.

Scenario	Material	ε_r
A	Wood, dry	4.5
B	Plexiglas	3.5
C	Wood, pressed board	2.0

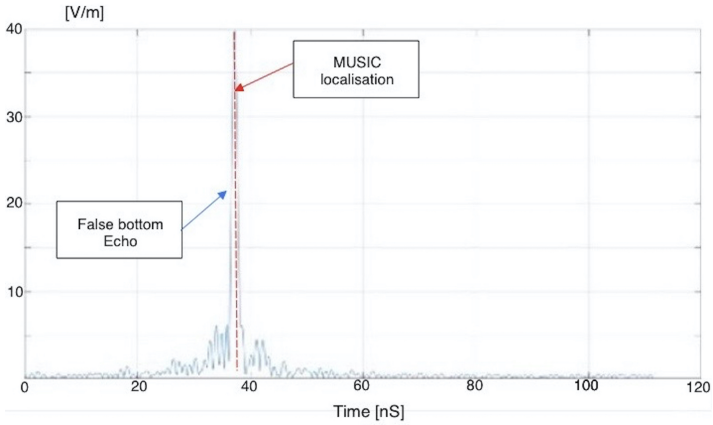


Figure 9: Empty container with a wood pressed board false bottom, placed at distance $r = 5.40m$.. Retrieved echo radar signal and MUSIC peak estimation.

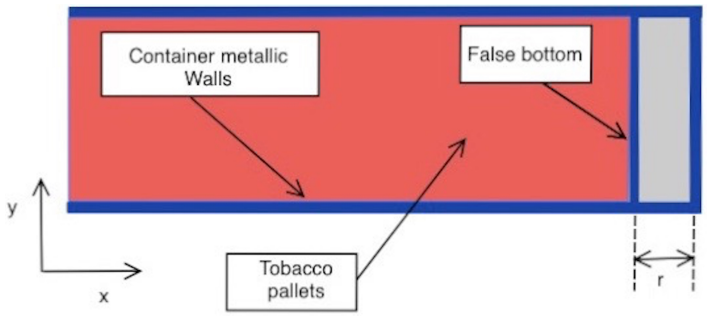


Figure 10: Container filled with tobacco pallets and a metallic false bottom placed at a distances $r = 5.30m$.

In the following experiments, different goods fill the container. The objective is to assess the radar capability to successfully detect the presence of false bottoms or other anomalies in goods without the necessity of unloading the container. This aspect is essential for police operations because it permits to check off more containers in less time. The shipper's shipping manifest permits estimating an electromagnetic wave's velocity and finding anomalies in the load. The first experiment considers a container filled with tobacco pallets and a metallic false bottom placed at a distance $r = 5.3m$ from the container door. The goal is to assess the capability to detect the false bottom despite the load presence. Scan-a result of the above scenario is reported in Fig. 11, as it can be noticed the echo is quite noisy due to the load presence. However, thanks to the MUSIC localization algorithm, the false bottom is successfully localized with a high level of accuracy with an error $\gamma = 12.5mm$. In the last experiment, the container is filled with blocks of granite. The dielectric permittivity of granite is $\varepsilon_r = 7.0$ quite high. This experiment is of particular interest because unload a container filled with granite

blocks requires much time and specific mechanical facilities. The scan-a results obtained from the simulation performed with gprMax are reported in Fig. 13. The echo is very well defined and presents a sharp peak easily detected by the MUSIC algorithm. The metallic false bottom is successfully detected with an error $\gamma = 2.3mm$. It is worth noticing that the results reported in Fig. 13are less noisy with respect to the data related to 11 a little bit noisier.

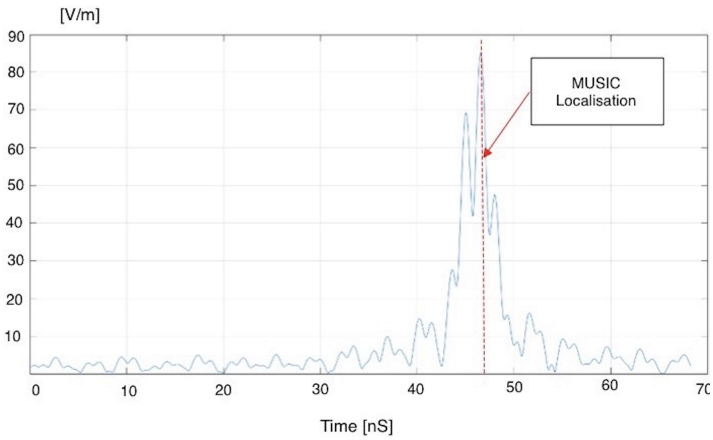


Figure 11: Container filled with tobacco pallets and a metallic false bottom placed at a distances $r = 5.30m$. Retrieved echo radar signal.

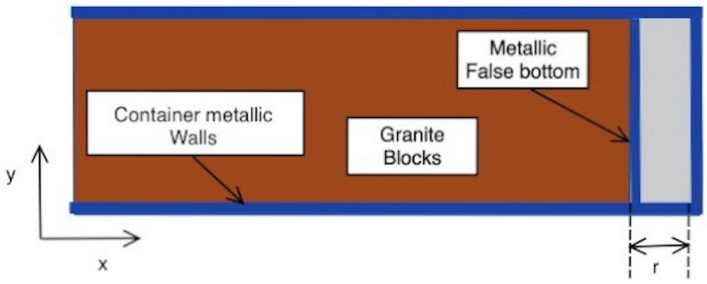


Figure 12: Container filled with granite blocks and a metallic false bottom placed at a distances $r = 5.30m$.

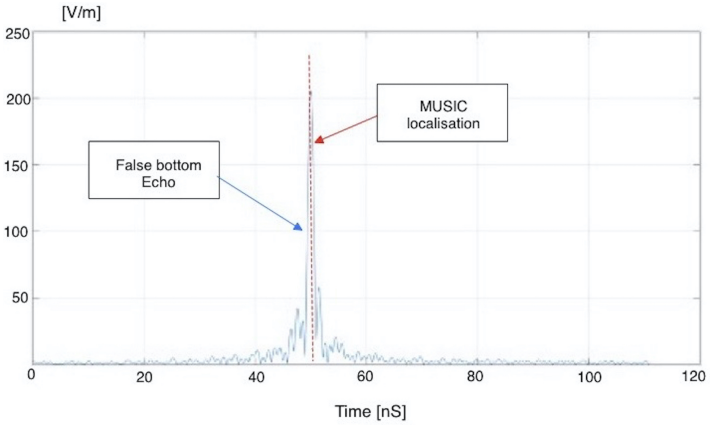


Figure 13: Container filled with granite blocks and a metallic false bottom placed at a distances $r = 5.30m$. Retrieved echo radar signal.

4.2 Experimental assessment

In this section, the SFCW radar prototype has been assessed in an operative scenario. The experimental setup refers to a typical through the wall application where a fugitive hides in a hiding place created behind a concrete wall. To detect humans hide behind walls, the radar prototype operates in continuous wave (CW) mode. The small chest and heart movements are detected by considering the Doppler effect carried on by the scattered electromagnetic wave. The wall thickness is $25cm$, the distance of the fugitive's chest from the wall is $d = 1.5m$, and he is sitting on a chair. The radar prototype is manually steered to the other side of the wall at a distance of $d_r = 1.0m$ with the goal of detecting the fugitive by revealing his signals life such as the breathing rate and the heartbeat. A photo of the considered through the experimental wall setup is reported in Fig.14.

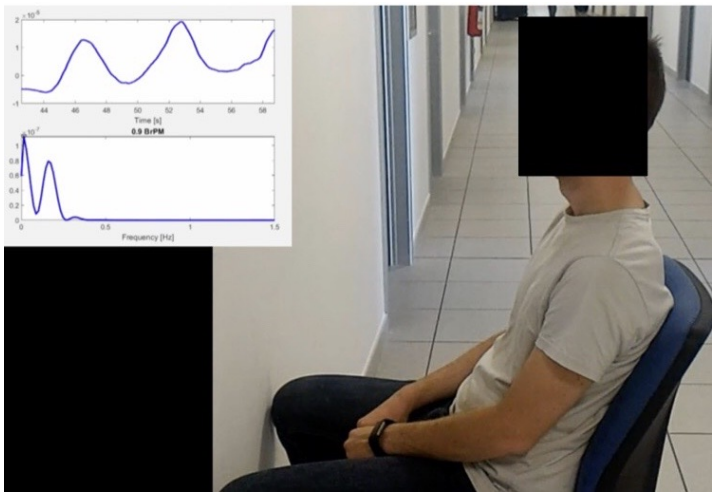


Figure 14: Experimental set-up. Radar in through the wall modality. Life signal detection, breathe rate and heartbeat, of a subject placed behind a 25 cm thickness wall.

The tool for the detection of human life signals has been selected in the radar GUI. The mechanical pedestal has been set at hight $h = 1.5m$ and then horizontally moved along the wall with a step size of about $s = 0.50m$, the total length of the wall was $2.0m$. The DDS generator's power has been set to the minimum value of $P_{tx} = -13dBm$. The electromagnetic wave power that reaches the fugitive chest is very low, and it cannot create health or interference problems with other devices. The following Fig. 15 reports the fugitive breathing rate detected at the second scan $1.0m$ far away from the sidewall. As it can be noticed from the data reported in Fig. 15, the fugitive shows normal breathing conditions, and it is quite relaxed. When the radar is moved on the first, third or fourth positions at $0.5m$, $1,5m$ and $2.0m$ respectively, far away from the sidewall, the system reports no human activities. In the next experiments, the fugitive simulated an accelerated breathing rate increasing from eight, with reference to Fig. 15, up to twenty breath cycles for minutes reported in Fig. 16. The radar prototype's resolution capabilities are quite accurate, and it can detect the small movements of the fugitive's heart. Fig. 17 (red line) reports the heart movement's radar signal. For the sake of comparison, the data reported in Fig. 17 has been compared with the data obtained

with a commercial optical heartbeat detector (blue line), as it can be noticed from Fig. 17 the agreement is quite accurate, and it demonstrate the capabilities of this prototype not only for police operations but also for the rescue operation.

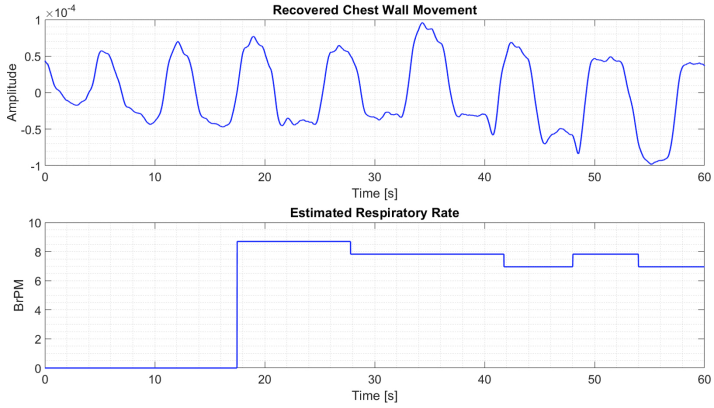


Figure 15: Experimental set-up. Life signals, breathing rate, of the fugitive detected behind the wall. Normal breathing conditions.

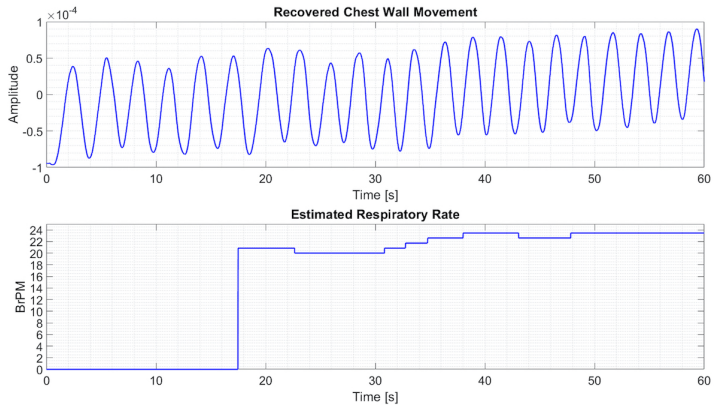


Figure 16: Experimental set-up. Life signals, breathing rate, of the fugitive detected behind the wall. Simulated accelerated breathing conditions.

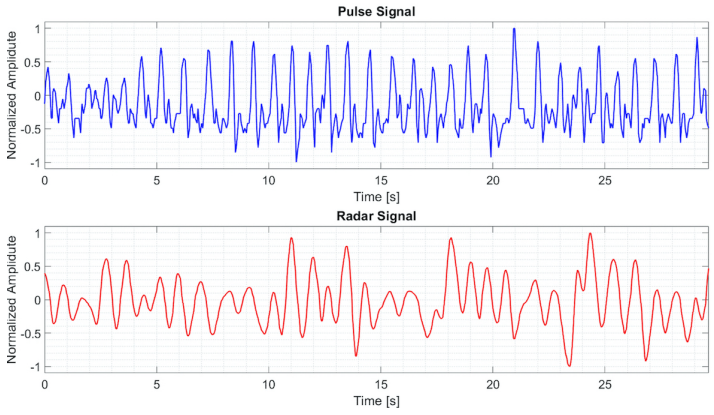


Figure 17: Experimental set-up. Life signals, heartbeat rate of the fugitive detected behind the wall with a more focused beam.

5 Conclusion

In this work, a Stepped Frequency Continuous Wave (SFCW) radar operating in the L band, suitable for fast inspection of foods and goods stored in metallic containers, has been proposed and numerically assessed in realistic scenarios. In particular, the proposed SFCW radar can estimate the dielectric characteristics of materials contained in a container and easily detect metallic or dielectric false bottoms commonly used for smuggling. A numerical assessment campaign, carried on considering different materials and simulating realistic scenarios, demonstrated the effectiveness and potentialities of the proposed system as a helpful tool for the contrast of smuggling.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The authors would like to thank Dr. E. Bortolotti for management the project and for the revision of manuscript, and the Autostrade Brennero S.P.A. A22 for funding the radar prototype.

References

[1] C. Nguyen, P. Joongsuk, Stepped-frequency radar sensors, Springer, 2016.

[2] G. Tronca, I. Tsalicoalou, S. Lehner, G. Catanzariti, "Comparison of pulsed and stepped frequency continuous wave (SFCW) GPR systems," in 17th International Conference on Ground Penetrating Radar (GPR), 1–4, 2018, doi: 10.1109/ICGPR.2018.8441654.

[3] A. Langman, M.-R. Inggs, "Pulse versus stepped frequency continuous wave modulation for ground penetrating radar," in Proceeding of International Geoscience and Remote Sensing Symposium, IGARSS 2001, 1533–1535, 2001, doi:10.1109/IGARSS.2001.976902.

[4] G. Gok, Y. K. Alp, O. Arikan, "A New Method for Specific Emitter Identification With Results on Real Radar Measurements," IEEE Trans. Inform. Forensic Secur., **15**, 3335–3346, 2020, doi:10.1109/TIFS.2020.2988558.

[5] L. Nicolaescu, P. V. Genderen, K. Palmer, "Calibration procedures of a stepped frequency continuous wave radar for landmines detection," in Proceedings of the International Conference on Radar (IEEE Cat. No.03EX695), 412–417, 2003, doi:10.1109/RADAR.2003.1278777.

[6] A. S. Turk, A. Kizilay, M. Orhan, A. Caliskan, "High resolution signal processing techniques for millimeter wave short range surveillance radar," in 17th International Radar Symposium (IRS), 1–4, 2016, doi:10.1109/IRS.2016.7884697.

[7] A. Kusmadi, A. Munir, "Simulation design of compact stepped-frequency continuous-wave through-wall radar," in International Conference on Electrical Engineering and Informatics (ICEEI), 332–335, 2015, doi:10.1109/ICEEI.2015.7352521.

[8] M. Pieraccini, M. Fratini, F. Parrini, G. Macaluso, C. Atzeni, "High-speed CW step-frequency coherent radar for dynamic monitoring of civil engineering structures," Electron. Letters, **40**, 907–908, 2004, doi:10.1049/el:20040549.

[9] M. Pieraccini, M. Fratini, F. Parrini, C. Atzeni, "Dynamic Monitoring of Bridges Using a High-Speed Coherent Radar," IEEE Trans. Geosci. Remote Sensing, **44**, 3284–3288, 2006, doi:10.1109/TGRS.2006.879112.

[10] S. S. Lambot, E. C. Slob, I. Van-den Bosch, B. Stockbroeckx, M. Vanclooster, "Modeling of ground-penetrating Radar for accurate characterization of subsurface electric properties," IEEE Trans. Geosci. Remote Sensing,, **42**, 2555–2568, 2004, doi:10.1109/TGRS.2004.834800.

[11] W.-C. Su, M.-C. Tang, R. E. Arif and, T.-S. Horng, F.-K. Wang, "Stepped-Frequency Continuous-Wave Radar With Self-Injection-Locking Technology for Monitoring Multiple Human Vital Signs," IEEE Trans. Microwave Theory Techn., **67**, 2555–2568, 2019, doi:10.1109/TMTT.2019.2933199.

[12] M. Caruso, M. Bassi, A. Bevilacqua, A. Neviani, "A 2–16 GHz 65 nm CMOS Stepped-Frequency Radar Transmitter With Harmonic Rejection for High-Resolution Medical Imaging Applications," IEEE Trans. Circuits Syst., **62**, 413–422, 2015, doi:10.1109/TCSI.2014.2362332.

[13] M. Donelli, F. Viani, "Life signals detection system based on a continuous-wave X-band radar," Electronics Letters, **52**, 1903–1904, 2016, doi:10.1049/EL.2016.2902.

[14] I. Nicolaescu, "Improvement of Stepped-Frequency Continuous Wave Ground-Penetrating Radar Cross-Range Resolution," IEEE Trans. Geosci. Remote Sensing, **51**, 85–92, 2013, doi:10.1109/TGRS.2012.2198069.

[15] M. Scherhaufl, F. Hammer, M. Pichler-Scheder, C. Kastl, A. Stelzer, "Radar Distance Measurement With Viterbi Algorithm to Resolve Phase Ambiguity," IEEE Trans. Microwave Theory Techn., **68**, 3784–3793, 2020, doi:10.1109/TMTT.2020.2985357.

[16] G. Cui, L. Kong, X. Yang, "Reconstruction Filter Design for Stepped-Frequency Continuous Wave," IEEE Trans. Signal Process., **60**, 4421–4426, 2012, doi:10.1109/TSP.2012.2197206.

[17] S. R. J. Axelsson, "Analysis of Random Step Frequency Radar and Comparison With Experiments," IEEE Trans. Geosci. Remote Sensing, **45**, 804–904, 2007, doi:10.1109/TGRS.2006.888865.

[18] A. N. Gaikwad, D. Singh, M. J. Nigam, "Application of clutter reduction techniques for detection of metallic and low dielectric target behind the brick wall by stepped frequency continuous wave radar in ultra-wideband range," IET Radar Sonar Navig., **5**, 416–425, 2011, doi:10.1049/iet-rsn.2010.0059.

[19] M. Donelli, M. Manekiya, G. Marchi, I. Maccani, C. Pascucci, "High Resolution L-band Stepped Frequency Continous Wave Radar for Smuggling Contrast," in 2020 2nd Global Power, Energy and Communication Conference (GPECOM), 327–332, 2020, doi:10.1109/GPECOM49333.2020.9247932.

[20] T. Yamakura, H. Yamada, Y. Yamaguchi, "Resolution improvement of the MUSIC algorithm utilizing two differently polarized antennas," in IEICE Transaction on Communications, 1827–1832, 1996, doi:10.1155/2019/8907685.

[21] S.-M. Shrestha, I. Arai, T. Miwa, Y. Tomizawa, "Signal processing of ground penetrating radar using super resolution technique," in IEEE Radar Conference, 300–305, 2001, doi:10.1109/NRC.2001.922995.

[22] K.-T. Kim, D.-K. Seo, H.-T. Kim, "Efficient radar target recognition using the MUSIC algorithm and inveriant features," IEEE Trans. Antennas and Propagation, **50**, 325–337, 2002, doi:10.1109/8.999623.

[23] S.-M. Shrestha, I. Arai, T. Miwa, "Signal processing of ground penetrating radar combining FFT and MUSIC algorithm for high resolution," in Tech. Rep. IEICE SANE2000 SAT2000-130, 31–43, 2003, doi:10.1155/S1110865703307036.

[24] S.-M. Shrestha, I. Arai, "Signal processing of ground penetrating radar using spectral estimation techniques to estimate the position of buried targets," EURASIP Journal on Applied Signal Processing, **12**, 1198–1209, 2003, doi:10.1155/S1110865703307036.

[25] V. Kafedziski, S. Pecov, D. Tanevski, "Target detection in SFCW ground penetrating radar with C3 algorithm and Hough transform based on gprMax simulation and experimental data," in 2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP), 2018, doi:10.1109/IWSSIP.2018.8439227.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 6, No. 4, 21-28 (2021)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Graph-based Clustering Algorithms – A Review on Novel Approaches

Mark Hloch[*,1], Mario Kubek[2], Herwig Unger[3]

[1]*Faculty of Electrical Engineering and Computer Science, University of Applied Sciences, Krefeld, 47805, Germany*

[2]*Central department I, FernUniversität Hagen, 58097, Germany*

[3]*Chair of Computer Engineering, FernUniversität Hagen, 58097, Germany*

ARTICLE INFO

Article history:
Received: 22 April, 2021
Accepted: 27 June, 2021
Online: 10 July, 2021

Keywords:
Graph-based clustering
Co-occurrence graph
SeqClu
DCSG

ABSTRACT

Classical clustering algorithms often require an a-priori number of expected clusters and the presence of all documents beforehand. From practical point of view, the use of these algorithms especially in more dynamic environments dealing with growing or shrinking corpora therefore is not applicable. Within the last years, graph-based representations of knowledge such as co-occurrence graphs of document corpora have gained attention from the scientific community. Accordingly, novel unsupervised and graph-based algorithms have been recently developed in order to group similar topics, represented by documents or terms, in clusters. The conducted work compares classical and novel graph-based algorithms, showing that classical clustering algorithms in general perform faster than graph-based clustering algorithms. Thus, the authors' focus is to show that the graph-based algorithms provide similar clustering results without requiring an hyperparamter k to be determined a-priori. It can be observed that the identified clusters exhibit an associative relationship reflecting the topical and sub-topical orientation. In addition, it is shown in a more in-depth investigation that the Seqclu (sequential clustering algorithm) can be optimized performance-wise without loss of clustering quality.

1 Introduction

Clustering is the process of grouping the most similar objects, e.g. images or text documents, in the same cluster in an unsupervised manner. In contrast to supervised classification, where the algorithm has been trained how to map its input to an according output, clustering only uses the provided input data and tries to find the best grouping of objects based on that information. Due to the sheer amount different data-types and according use-cases, dozens of architectural concepts, such as hierarchical, partitioning or graph-based clustering have been developed over the last decades[1]. Many of the classical, typically vector-based algorithms, such as the k-means[2], k-means++[3] or k-NN[4] algorithm come with the requirement of choosing the hyperparameter k, as the suggested number of expected clusters, a priori. This forces the user to estimate beforehand what number of result clusters are expected. This approach therefore softens the idea of an unsupervised algorithm providing the best possible result fully automatically, without user intervention. In addition, actual standard algorithms mentioned above, but also newer graph-based algorithms, e.g. Chinese Whispers[5] expose another weakness: they typically require a full set of documents beforehand and are not designed to adapt a growing or shrinking set of input data over time. In use-cases like building a web-engine clustering is known to improve the usability for the user: instead of having a long list of search results on a user's query the knowledge, that lies within the available document corpus, can be presented much more easily[6]. With the work of [7], [8] it is shown that the use of co-occurrence graphs is very useful for graph-based concepts on which novel clustering techniques can be built on. Recent work on graph-based clustering algorithms[9], [10] provide novel approaches in the field of clustering. This paper gives a comparison of these graph-based algorithms and shows their benefits over classical approaches, as well as requirements for further optimization.

2 Materials and methods

2.1 Classical clustering algorithms

2.1.1 K-means, K-means++, Mini Batch

Due to its simplicity the k-means algorithm and its variants are the most prominent representatives of vector based clustering algorithms used. As k-means assigns each input data point to exactly one cluster it is considered as a classical hard-clustering algorithm.

[*]Corresponding Author: M. Hloch, Reinarzstr. 49, 47805 Krefeld - Germany, mark.hloch@hs-niederrhein.de

The computational complexity, which is linearly proportional to the size of datasets, makes the k-means algorithm efficiently applicable even to larger datasets. Because each data point is represented numerically, the application field for k-means is wide: It can be used from document clustering up to other use cases, e.g. customer segmentation [11] or cyber-profiling criminals [12].

The general idea of the algorithm is partitioning the given data into k distinct clusters by iteratively updating the cluster centers and cluster associated data points. The k-means algorithm mainly performs two steps: Firstly, the user has to define manually the value k, which determines how many clusters should be a result of k-means. Secondly, a loop of two repeating steps of assigning the data set points to one of the clusters with lowest distance to the cluster's centroid and the calculation of a new centroid for each cluster is performed. Mathematically it can be said that the k-means is an optimization problem where the objective function that is employed is the Sum of Squared Errors (SSE). During the assignment and update steps the k-means algorithm tries to minimize the SSE score for the set of clusters. A more detailed overview on the mathematical implications and technical origins can be found at [13].

Due to its impact on the clustering result the distance measure has to be chosen carefully. In general, the most popular choice for estimating the closest centroid to each of the datapoints is the Euclidean distance. Other distance measures, such as the Manhattan distance or cosine similarity can also be used [13]. The k-means algorithm comes with two major disadvantages:

1. choosing the initial cluster centers
2. estimating the k-value.

The clustering result of k-means highly depends on the **initialization of the cluster centers** [14]. In order to optimize the results provided by k-means especially the initialization of the cluster centers has been subject of research over the years and lead to several approaches [15]–[16]. In contrast to the classical k-means algorithm as described in [2], the k-means++ more carefully determines the initial cluster centers and subsequently uses a weighted probability score to improve the finding of cluster centers over time.

Estimating the k-value and therefore the number of output clusters k a priori also is a big disadvantage in contrast to fully unsupervised clustering algorithms. From the practical point of view this would require the user to estimate a good k value before actual clustering can be performed. Especially in cases where the input data are very large or growing over time the manual selection of k is therefore not applicable.

Several approaches such as the Silhouette Coefficient [17] or Calinski–Harabasz Index [18] can be used to suggest the k-value automatically. In addition to the above-mentioned improvements, other algorithms like the Mini-batch k-means or fast k-means [19], [20] aim to improve the scalability and performance of k-means for large datasets such as web applications. In case of the mini-batch k-means algorithm small random batches of data are chosen and assign each of the sample points to a centroid. In a second step the cluster centroid is then updated based on the streaming average of all of the previous samples assigned to that centroid.

2.1.2 *Chinese Whispers*

The Chinese Whispers (CW) algorithm [5] is a randomized graph-based hard-clustering algorithm. Due to its simplicity and linear time properties, it performs very fast even for larger datasets. As it does not require a preliminary k-value it can be considered as an unsupervised clustering algorithm. The type of graph on which the CW algorithm is applied can be weighted, unweighted, undirected or directed. The CW is there-fore applicable to a wide range of use cases in natural language processing such as language separation or word sense disambiguation. The CW algorithm works in a bottom-up manner by first setting a random class label for each node of the graph and then merging class labels with those local neighborhood classes with the biggest sum of edge weights. In case of multiple winning classes, one is chosen randomly. Over time regions of the same class will stabilize and grow till they connect to another class region.

As shown in [5], CW scales are very well even for large datasets in linear time. It is also shown that the clustering quality is comparable to standard algorithms using the vector space model. As the CW algorithm has randomized properties its output changes on each run of the algorithm. This makes it hard for practical applications where the data change over time or in cases where the clustering process has to be repeated for the same data. Another disadvantage of the CW algorithm is its tendency of forming a large amount of often very small clusters. In order to extract the dominant clusters, e.g. to obtain an overview of related topics for each cluster, filtering is required which may result in unwanted information loss.

2.2 *Novel graph-based clustering algorithms*

2.2.1 *Dynamic clustering for segregation of co-occurrence graphs (DCSG)*

The DCSG [10] algorithm is a novel clustering algorithm inspired by the human's brain learning process. As the human brain develops from child to adulthood it continuously learns new words and categorizes them forming the entire knowledge of the human being. Transferring this method to the concept of co-occurrence graphs, clustering can be applied in order to identify topical related regions (clusters) formed by the terms within the graph. DCSG imitates the learning process by reading each document on sentence base while building up the co-occurrence graph. The nodes represent the terms, edges the relation of the words. While adding each term to the co-occurrence graph clustering is applied by measuring the distance between each new term and its distance to the existing cluster centers. The distance between two terms is determined by the inverse of the DICE-coefficient [21], [22]. The cluster center itself is represented by its centroid term as defined in [23], [24]. A cluster center is formed by the node with the shortest average distance Δd to every other node in the co-occurrence graph. In addition, the standard deviation μ is determined. Based on this information the cluster range $r_{cluster_i}$ is defined as

$$r_{cluster_i} = \Delta \mathrm{d} + 3\mu \tag{1}$$

and is used to determine which term belongs to which cluster. While processing new terms two major steps are performed

1. **Insertion of terms**

   New terms will be added to the graph as nodes and edges to the according co-occurrent terms. If a term is already existing the related connections will be updated.

2. **Clustering**

   Each term t_{new} will be merged into existing clusters if the distance $t_{new} \leq \Delta d + 3\mu$. In case that $t_{new} > \Delta d + 3\mu$ a new cluster will be created. In addition, the algorithm will check if due to the insertion any relations may have changed in order to update the cluster centers correctly.

Adding a new term that already exists in the graph will result in a change of the relation weights and cluster centers may change. In this case $\Delta d + 3\mu$ will also change. The membership of the according terms with that cluster therefore needs to be re-evaluated in case of exceeding the distance threshold might be moved to different clusters.

DCSG is a novel brain inspired algorithm that works in contrast to the classical approaches like k-means on dynamic data growing over time. As presented in the results on clustering text data show that this knowledge can be represented very well. Due to the high amount of recalculation steps, especially when the cluster centers move frequently, having a large amount of initial input data the DCSG algorithm will perform very slow. If the data is growing over time, just similar to the human brain's learning process, this disadvantage will be compensated as the graph will converge over time and recalculations will decrease.

2.2.2 *Sequential Clustering using Centroid Terms (SeqClu)*

The SeqClu-algorithm [9] is a co-occurrence graph-based hard-clustering algorithm that is capable of clustering documents in a sequential manner using the concept of centroid terms as described in [22]–[24].

In contrast to many of the standard clustering algorithms it does not require an a-priori definition of the number of clusters. As the algorithm works sequentially a set of feature vectors F of documents is processed incrementally at once or just as they appear over time. SeqClu is therefore applicable in cases where the input vector might change over time such as in a Web-Engine, where documents appear or disappear over time.

The general idea of the algorithm is to compare each new feature vector f (a document) against an existing set of clusters containing previously clustered documents. For each existing cluster the algorithm considers the membership of a new document by a distance determination process. If a closest cluster can be found without exceeding a certain threshold the document will join the according cluster. If not, a new cluster will be formed by the document extending the cluster model.

There are mainly three crucial parts that influence the behaviour of the algorithm:

- initialization,
- cluster membership value and
- threshold of membership (winning cluster determination).

The **initialization** can be performed in various ways and influences the quality of clustering result. If a set of one or more clusters are predefined it will be used as a reference for all further arriving documents to be clustered. The simplest initialization would be using a single cluster having a randomly chosen document as the first cluster. In combination with an inaccurately chosen membership threshold this approach most likely will tend to merge new documents, especially a low number of documents, into the initial cluster and imprecise the overall clustering result. Instead, it is suggested to form at least two initial clusters formed by the most distant documents (antipodean documents) existing. This requires an additional preprocessing step as for each of the initially existing documents the distance between each of the documents centroid must be determined. It is known that the co-occurrence graph will converge at about 100 documents and the process of distance determination can be very time consuming as the conducted experiments show. It is therefore suggested to consider only 2 to 100 randomly chosen documents and is also suspect of the experiments in this publication.

In order to determine a document's **cluster membership** each newly arriving document needs to be matched against the existing clusters. For this purpose, the average distance between the new documents centroid and all existing documents in each of the clusters is determined by using Dijkstra's shortest path algorithm. In order to speed-up the calculation process all previous path calculations are cached having a linear time complexly for reoccurring path determinations.

Regarding the judgement whether a document shall be merged into an existing or form a novel cluster, a **threshold** is required. As the threshold determination process directly influences the clustering result it must be chosen very carefully. To avoid having the user to choose this threshold manually, it is chosen dynamically for each of the clusters using the local connections between the cluster's centroid and its nearest neighbors. In order to reduce the amount of computation time, only neighbors within a certain radius are considered using a breadth-first search with limited depth.

As first experiments have shown the SeqClu-Algorithm is able to cluster documents unsupervised in a sequential manner providing good results without performing any optimization regarding the process of threshold determination. The initial step of finding the two initial clusters, especially having a large initial set of documents, is very time consuming. On the other hand, the calculation results of the initialization phase can be reused for the clustering process itself having a lookup table. In this case the disadvantage of slow initialization becomes an advantage for the clustering process itself and will speed-up this phase of the algorithm, which will be shown in the experiments conducted in this publication.

2.3 *Conceptual differences of tested algorithms*

Table 1 provides an overview over the six tested algorithms from conceptual point of view. All algorithms work with different requirements for initialization and are applicable to different use cases. In general, the classical algorithms are flexible regarding the type of input data, which can be numerical or textual data. From the user point of view the biggest disadvantage is the requirement to provide the number of output clusters (k-value) a-priori. Therefore, user

intervention is required in order to update the suggested number of output clusters or the use of additional algorithms that determine the k-value automatically. The classical algorithms are therefore especially inflexible in cases, where the corpus size changes over time. In contrast, the graph-based algorithms fully run unsupervised and therefore do not require any further change in initialization even if the corpus grows or shrinks. From clustering point of view the algorithm will adapt and provide automatically an according number of output clusters.

Table 1: Conceptual differences of tested algorithms

Algorithm	Type	Supervised/ unsupervised	A-priori no. of clusters	Growing corpora
K-means	vector-based	supervised	k-value	no
K-means++	vector-based	supervised	k-value	no
Minibatch	vector-based	supervised	k-value	no
CW	graph-based	unsupervised	none	no
SeqClu	graph-based	unsupervised	none	yes
DCSG	graph-based	unsupervised	none	yes

3 Results and discussion

3.1 Setup of experiments

For the conducted experiments, natural language preprocessing in form of sentence extraction, stop word removal and baseform reduction has been applied to all of the used documents. The according output was used as standardized input for all algorithms tested. In addition, the co-occurrence graph G has been built on sentence-based co-occurrences measuring the distance between terms by the reciprocal significance value. The significance value itself is determined using the DICE-coefficient. All small sub-graphs from the original co-occurrence graph have been removed finding the largest connected sub-graph within G in order to obtain a consistent single graph G. The graph is stored in an embedded graph database using Neo4j [25] where each node represents a term and the edges are annotated with their significance and respective distance value. As the vector-based algorithms don't use a co-occurrence graph the TF-IDF [26] matrix was created based on the documents term-vector, which was extracted during natural preprocessing. The implementation of k-means, k-means++ and mini-batch were realized by using the python-based machine learning programming library scikit-learn version 0.23. The Chinese Whispers algorithm was taken from the ASV Toolbox [27] from the university of Leipzig and is written in Java. SeqClu and DCSG algorithms were both implemented using the Java programming language. All experiments have been performed on ten workstations with identical hardware specs.: Intel Core i7-7700K CPU @ 4.20GHz and 16 GB Memory

3.2 Used corpora and parametrizations

The conducted experiments were performed over corpora consisting of 40,60,. . . ,260,280 documents and for each corpus the tested algorithms did run 100 times in order to obtain statistical relevant results. The corpora each contained a random number of equally distributed documents of the categories politics, cars, money and sports of the German newspaper "Der Spiegel". The documents themselves were each tagged by the author with their respective text category and therefore can be used as a gold standard in order to evaluate whether a document has been clustered correctly or not. Table 2 shows the parametrizations that were used for the individual algorithms.

Table 2: Parametrization of used algorithms

Algorithm	Parametrization
K-means, K-means++	Number of output clusters (k-value): 4 Initial cluster center: random Number of iterations: 100
Minibatch	Identical to K-means Batchsize: 100 Number of samples (init): (3 ∗ batchsize)
CW	Number of iterations: 100 Mutation rate: 0.0
DCSG	No initialisation required
SecClu	Initialisation: antipodean documents Dynamic threshold

3.3 Clustering Quality

Evaluating different clustering algorithms is difficult. Commonly used external evaluation measures, e.g. the f-measure or rand index, penalize false positive and false negative decisions. If the number of output clusters exceeds the number of classes it will result in a quality trade-off. Especially the tested graph-based clustering algorithms don't require a pre-defined k-value and break this condition by exceeding the number of output clusters. In order to limit the effect of quality trade-off the purity is preferred over other evaluation measures as it does not penalize if the number of output clusters is bigger than the number of class labels. The purity was calculated for each of the tested corpora and algorithms by

$$purity(C, M) = \frac{1}{N} \sum_{k} \max_{j} |c_k \cap m_j| \tag{2}$$

with N as the total number of documents, the set of clusters $C = \{c_1, c_2, \cdots, c_k\}$ and $M = \{m_1, m_2, \cdots, m_j\}$ as the set of classes. For each cluster c_k we determined the class m_j with the most members $n_{k,j}$ in c_k and then subsequently $n_{k,j}$ is summed up and divided by N.

Figure 1 shows the total average purity of all algorithms. The classical algorithms like k-means reside within a range of 0.7 to 0.8 exhibiting a good purity value. The graph-based algorithms, even with lower purity values, still are close having average values between 0.6 to 0.7.

It can be observed from Table 3 that for the standard algorithms the purity in average is almost constant and independent of the number of clustered documents.

The graph-based algorithms SeqClu and DCSG show, that their purity values increase in quality at about 100 to 160 documents per corpus. For SeqClu this behavior reflects that the co-occurrence graph converges at about 100 documents. Each additionally added document causes lesser changes to the co-occurrence graph resulting in lesser changing centroid terms, that are used by this clustering

algorithm for cluster determination. The DCSG also shows this tendency as the knowledge of the graph that is used for clustering also improves over time and stabilizes.

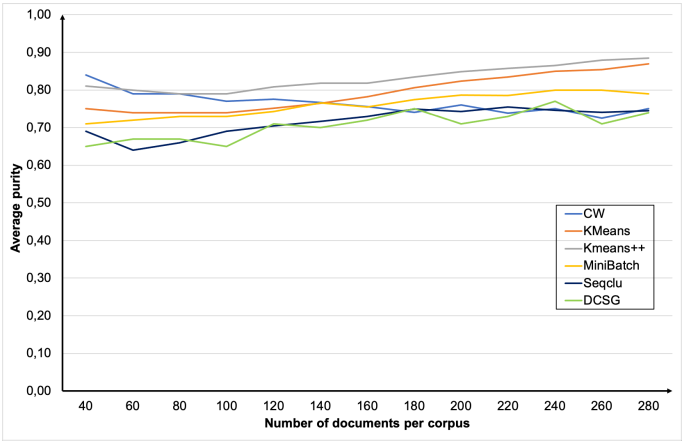


Figure 1: Comparison of average purity for different corpora sizes

Table 3: Average purity for each of the tested algorithms

# of doc.	K-Means	K-Means++	Minibatch	CW	SeqClu	DCSG
40	0.75	0.81	0.71	0.84	0.69	0.65
60	0.74	0.80	0.72	0.79	0.64	0.67
80	0.74	0.79	0.73	0.79	0.66	0.67
100	0.74	0.79	0.73	0.77	0.69	0.65
120	0.75	0.81	0.74	0.78	0.70	0.71
140	0.76	0.82	0.77	0.77	0.72	0.70
160	0.78	0.82	0.75	0.76	0.73	0.72
180	0.81	0.83	0.77	0.74	0.75	0.75
200	0.81	0.85	0.79	0.76	0.74	0.71
220	0.83	0.86	0.79	0.74	0.75	0.73
240	0.85	0.87	0.80	0.75	0.75	0.77
260	0.85	0.88	0.80	0.72	0.74	0.71
280	0.87	0.88	0.79	0.75	0.75	0.74
Avg. total	0.79	0.83	0.76	0.77	0.72	0.71

3.4 Cluster-sizes and number of clusters

As the purity can be influenced by having a large number of clusters - in worst case one cluster per document - it is required to take a closer look at the number of clusters and documents per cluster. This effect mainly has an effect to the graph-based algorithms as they expose in contrast to the standard algorithm ($k = 4$) a larger number of output clusters. As the SeqClu algorithm is in the authors focus of research the investigations were focused on this algorithm and resulted in the following two main observations:

1. The average cluster size in average is almost constant between 4 and 5 documents per cluster and

2. the number of clusters is slowly increasing from 20 to 50 for the tested cluster sizes from 40 to 280 documents.

In general, it can be concluded that the knowledge growth at this point seems to be not converged or the parametrization needs to be finer grained, as all the tests were using standard parametrization. In order to reveal possible indicators and starting-points for further optimization, the tested corpora were manually examined regarding the clustering result and the individual correctness of the clusters, even if the gold standard suggested number of output clusters is exceeded.

After examination the following commonly occurring observations have been made for the major number of tested corpora:

1. **Homogenous Clusters**

   Homogenous Clusters are found 4 to 5 times in average, having 10-20 documents with 100 percent accuracy and reflect an actual topic the documents focus on. Those clusters contain exactly one category and in addition, the topical relatedness within that category is identifiable. For example, a cluster of category “sports” can be clearly related to the topic formula 1 racing by the document centroids (number of occurrences in brackets): Season (3), Schumacher (8), Race (5)

2. **Heterogeneous Clusters**

   Heterogeneous Clusters occur in 1 to 2 clusters at a size of 10-20 documents of mainly two mixed topics. It could be observed that in some cases the centroid term that mainly influences the categorization of a document is generic. For example, terms like Euro, Zahl or Percent are very hard to map to a certain category. The according document could be related to Money, Politics or even Sports. In general, it seems difficult to identify the main scope of a cluster automatically for heterogeneous but also homogenous clusters. An according algorithm must be able to identify the topic and in case of having a heterogeneous cluster, differentiate between the mixed topics, which could occur on term level, which is focus on further research.

3. **Border Clusters**

   Border clusters occur at a varying number of times, depending on the topics within the corpus. These cluster types have an average size of 5-6 documents dominated by exactly one topic, but also having a single document of a different topic within. For example, the cluster may contain the centroid terms {War(2), Irak(2), USA(1)} of category politics, but one document with centroid {Beginning}. The politic topic (War, Irak, USA) can be clearly identified as the main topic for this cluster. In this example the document with the centroid Beginning cannot be accurately assigned to a single topic. The actual document refers to the beginning of a voting process regarding enrolment fees at the beginning of the semester. As the term {Beginning} also makes sense if related to the beginning of war between US and Irak the cluster has its border to a different topic marked by this term. Border clusters are interesting due to the fact that they mark the border between somehow topical related areas in the graph. This might be used e.g. within a search engine to lead the searching user to adjacent fields of knowledge.

4. **Single document clusters**

   Single document clusters are clusters formed by single documents that could not be assigned to any other cluster by exceeding the dynamic threshold used in SeqClu. In very early steps there might be not as many clusters with clearly defined topics that could lead into the creation of these new

clusters. As SeqClu is a single linkage algorithm it is prone to outliers. Further research in order to mitigate this problem has to be done in future work, e.g. by merging the single document cluster with a larger cluster within a short distance.

Putting things together it can be concluded that the graph-based clustering algorithm performs very well regarding purity. As the number of output clusters is greater than the gold standard suggests, the differentiation of different topics is much closer to real world applications. In contrast, the standard algorithms gain a high purity. But when manually evaluating the clustering results topical differences are not detected.

3.5 *Pure clustering performance*

Figure 2 shows the pure execution time of clustering different corpus sizes for all standard and graph-based clustering algorithms, except DCSG.

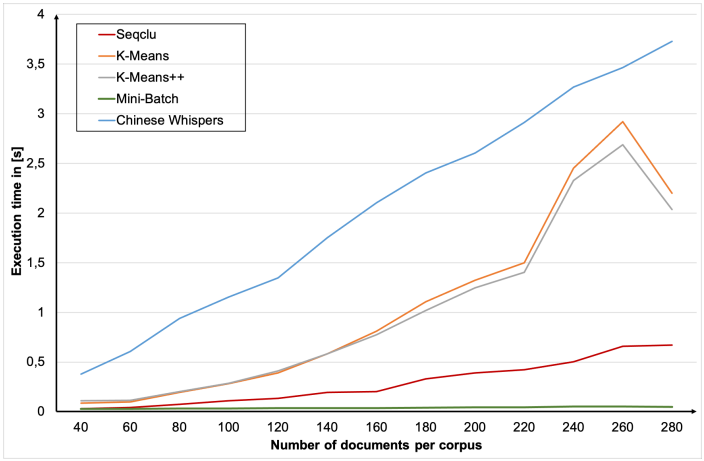


Figure 2: Pure execution time for clustering

In general, it can be concluded that the Minibatch algorithm performance is outstanding in comparison to the other algorithms even for larger corpora. As the Minibatch algorithm is designed especially for larger datasets, it is not surprising that it performs that well and faster than k-means and k-means++.

The graph-based SeqClu algorithm shows a linear increasing execution time for the pure clustering. It is also very fast in comparison of the other algorithms as it benefits from a caching mechanism of shortest path calculations that is used during the initial cluster initialization.

Figure 3 shows the overall execution time including all preliminary initializations for DCSG and SeqClu. In case of SeqClu and DCSG require, due to their focus on non-real-time applications and their heavy use of graph-based distance calculations, additional computations while clustering or during initialization. In case of the DCSG algorithm, that works on a sentence-based manner, a huge amount of distance determinations has to be performed which results in a high computation time of up to approx. 17 hours on 280 documents. As the scope of DCSG is not to be tremendously fast but trying to find an accurate representation of knowledge this performance drawback is not essential for the entire quality of the algorithm. Further investigations are beyond scope of the authors work and therefore not subject of this paper. SecClu's execution time in comparison to DCSG is about 2 hours in case of 280 documents. It is mainly influenced by the calculation of the nearest neighbors and number of documents used during the determination of initial clusters (antipodean documents). Especially the determination of antipodean documents leaves space for improvements.

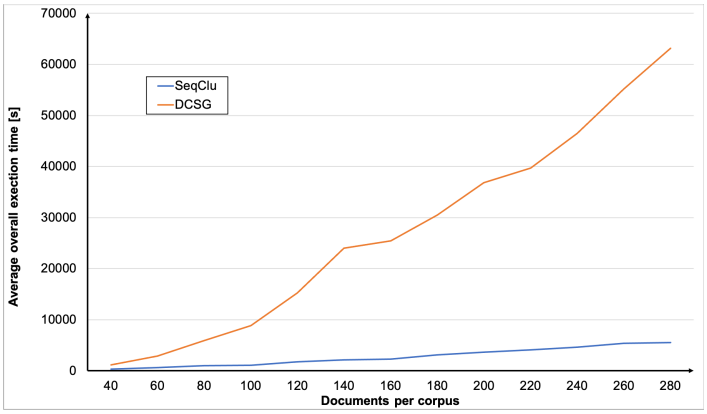


Figure 3: Execution time SeqClu and DCSG (initialization and clustering)

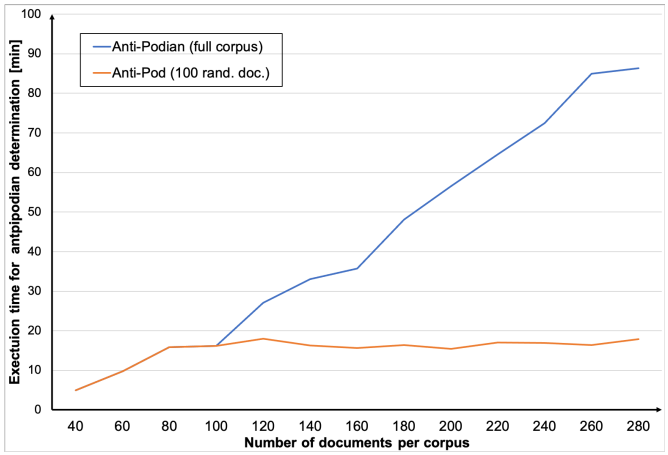


Figure 4: Execution times for initialization of SeqClu and SeqClu-100

The conducted experiments therefore have been performed by comparing the execution time for SeqClu using all available documents during initialization versus SeqClu-100, using only the first up to 100 randomly chosen available documents instead. It can be concluded (see Figure 4) that depending on the corpus size up to approx. 70 minutes of computational time could be saved for the tested corpora using SeqClu-100.

It is expected that co-occurrence graph converges at about 100 documents and can be considered as stable. When using 100 initial documents – even if randomly – chosen, the graph therefore is considered as stable and side-effects, e.g. on clustering quality should be negligible. Therefore, the purity has been also determined (Figure 5) concluding that there is no overall negative impact of using only 100 random documents for finding the two initial clusters on clustering quality.

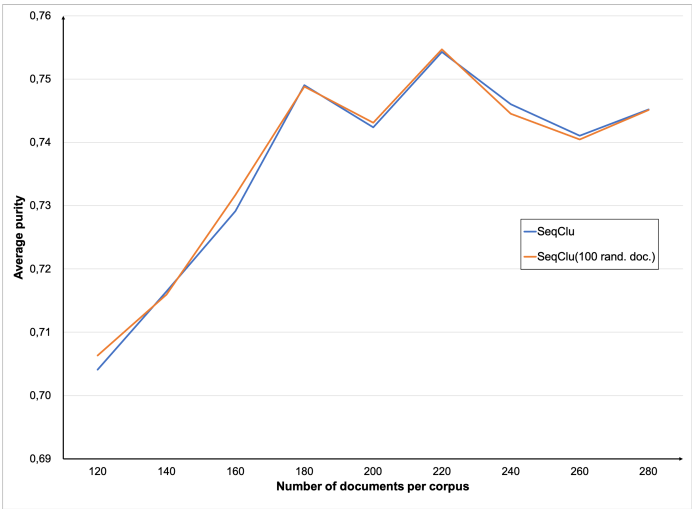


Figure 5: Comparison of purity for SeqClu and SeqClu-100

Further experiments also showed that the average cluster-sizes and number of output clusters do not change significantly and limiting the initialization of SeqClu to 100 documents is a performance enhancement which is reflected by having a better overall performance.

4 Conclusions

The provided work compares novel graph-based clustering algorithms against well-known vector-based algorithms. All algorithms were investigated regarding their clustering quality and general performance. The results show that the vector-based algorithms generally perform at a higher speed in contrast to the examined graph-based clustering algorithms. In contrast, the classical approaches such as the k-means algorithm forces the user's intervention a-priori, which limits the use cases where the user is able to investigate into the input data before actually clustering is applied. In contrast, the graph-based clustering algorithms show that a good categorization without the preliminary of a k-value requirement is possible. In addition, it can be concluded that graph-based clustering provides the property of having an associative representation - similar to the human brain – of the clustered data. A topical differentiation between individual topics subtopic is therefore much closer to the actual way of thinking of the user in contrast to classical approaches. The in-depth investigation into SeqClu's clustering structure and relations between the result clusters gained in further starting-points for optimization. Additionally, it was shown by limiting the number of documents during initialization that the performance of SeqClu can be significantly improved without negative impact on the overall clustering quality.

References

[1] V. Estivill-Castro, "Why So Many Clustering Algorithms: A Position Paper," SIGKDD Explor. Newsl., **4**(1), 65–75, 2002, doi:10.1145/568574.568575.

[2] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, 281–297, University of California Press, Berkeley, Calif., 1967.

[3] D. Arthur, S. Vassilvitskii, "K-means++: The Advantages of Careful Seeding," in Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07, 1027–1035, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2007.

[4] N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression," The American Statistician, **46**(3), 175–185, 1992, doi:10.1080/00031305.1992.10475879.

[5] C. Biemann, "Chinese Whispers: An Efficient Graph Clustering Algorithm and Its Application to Natural Language Processing Problems," in Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing, TextGraphs-1, 73–80, Association for Computational Linguistics, Stroudsburg, PA, USA, 2006.

[6] M. Christen, et al., YaCy: Dezentrale Websuche, Online Documentation on http://yacy.de/de/Philosophie.html, 2017.

[7] M. Kubek, Dezentrale, kontextbasierte Steuerung der Suche im Internet, Ph.D. thesis, Hagen, 2012.

[8] M. Kubek, Concepts and Methods for a Libarian of the Web, FernUniversität in Hagen, 2018.

[9] M. Hloch, M. Kubek, "Sequential Clustering using Centroid Terms," in Autonomous Systems 2019: An Almanac, 72–88, VDI, 2019.

[10] S. Simcharoen, H. Unger, "Dynamic Clustering for Segregation of Co-Occurrence graphs," in Autonomous Systems 2019: An Almanac, 53–71, VDI, 2019.

[11] M. Bacila, R. Adrian, M. Ioan, "Prepaid Telecom Customer Segmentation Using the K-Mean Algorithm," Analele Universitatii din Oradea, **XXI**, 1112–1118, 2012.

[12] M. Zulfadhilah, Y. Prayudi, I. Riadi, "Cyber Profiling using Log Analysis and K-Means Clustering A Case Study Higher Education in Indonesia," International Journal of Advanced Computer Science and Applications, **7**, 2016, doi:10.14569/IJACSA.2016.070759.

[13] C. Aggarwal, C. Reddy, DATA CLUSTERING Algorithms and Applications, 2013.

[14] J. Pena, J. Lozano, P. Larranaga, "An Empirical Comparison of Four Initialization Methods for the K-Means Algorithm," 1999.

[15] E. Forgy, "Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classification," Biometrics, **21**(3), 768–769, 1965.

[16] S. Khan, A. Ahmad, "Cluster center initialization algorithm for K-means clustering," Pattern Recognition Letters, **25**, 1293–1302, 2004, doi:10.1016/j.patrec.2004.04.007.

[17] L. Kaufman, P. Rousseeuw, "Finding Groups in Data: An Introduction to Cluster Analysis," 1990.

[18] T. Caliński, H. JA, "A Dendrite Method for Cluster Analysis," Communications in Statistics - Theory and Methods, **3**, 1–27, 1974, doi:10.1080/03610927408827101.

[19] D. Sculley, "Web-Scale k-Means Clustering," in Proceedings of the 19th International Conference on World Wide Web, WWW '10, 1177–1178, Association for Computing Machinery, New York, NY, USA, 2010, doi:10.1145/1772690.1772862.

[20] A. Feizollah, N. Anuar, R. Salleh, F. Amalina, "Comparative Study of K-means and Mini Batch K- means Clustering Algorithms in Android Malware Detection Using Network Traffic Analysis," 2014, doi:10.1109/ISBAST.2014.7013120.

[21] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," Ecology, **26**(3), 297–302, 1945, doi:10.2307/1932409.

[22] M. Kubek, T. Böhme, H. Unger, "Empiric Experiments with Text Representing Centroids," in 6th International Conference on Software and Information Engineering (ICSIE 2017), 2017.

[23] M. Kubek, H. Unger, "Centroid Terms and their Use in Natural Language Processing," in Autonomous Systems 2016, VDI-Verlag Düsseldorf, 2016.

[24] M. Kubek, H. Unger, "Centroid Terms As Text Representatives," in Proceedings of the 2016 ACM Symposium on Document Engineering, DocEng '16, 99–102, ACM, New York, NY, USA, 2016, doi:10.1145/2960811.2967150.

[25] A. Vukotic, N. Watt, T. Abedrabbo, D. Fox, J. Partner, Neo4j in Action, Manning, 2015.

[26] G. Salton, A. Wong, C. S. Yang, "A Vector Space Model for Automatic Indexing," Commun. ACM, **18**(11), 613–620, 1975, doi:10.1145/361219.361220.

[27] C. Biemann, U. Quasthoff, G. Heyer, F. Holz, "ASV Toolbox: a Modular Collection of Language Exploration Tools," in Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), European Language Resources Association (ELRA), Marrakech, Morocco, 2008.

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 37-43 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

A Statistical Description of Students Admitted to Higher Education Institutions, Public and Private, in Albania for the Academic Year 2017-2018

Feruze Shakaj[1,*] Markela Muça[2] , Klodiana Bani[2]

[1]*Ministry of Education and Sports (MES)., Street "Durrësit" Nr.23, 1001 Tiranë, Albania*

[2]*University of Tirana, Faculty of Natural Sciences, Department of Applied Mathematics, " Zogu I " Boulevard, Nr. 25/1,* 1060, *Tiranë, Albania*

ARTICLE INFO

Article history:
Received: 04 May, 2021
Accepted: 20 June, 2021
Online: 10 July, 2021

Keywords:
Study programs
Gender
State matura
Bayesian Criterion
Two steps cluster method
Albania

ABSTRACT

This paper makes a statistical analysis of some indicators that characterize the environment of students admitted to public and private Institutions of Higher Education i.e. university level institutions, presenting an overview to the distribution of these students in the key areas of study that these institutions offer such as: Arts, Agriculture, Natural Sciences, Social Sciences, Medical Sciences and Sports.

The distribution is studied based on gender, residential area (city or village) and high school average. The study is undertaken to know and better understand the trends related to these indicators in the main areas of study mentioned above. A careful description of the figures from the stidy reveals typical features and explains better the situation. The purpose of the study is to see the impact of these factors on the study programs where these students have been declared winners. In this paper it will be introduced the two-step method (two steps cluster method) and it will be illustrated with an application on a database obtained from the State Matura (Center for Educational Services, Ministry of Education, Sports and Youth).

The two-step cluster analysis identifies the clusters by first executing pre-clustering and then applying the hierarchical method until the final p clustering. Because of the ability to use a fast-clustering algorithm in advance, it can handle large data sets. To evaluate the quality of the groups we used the value of silhouette measure of cohesion and separation. The software used to perform the analysis is SPSS, V 25. The criterion used to determine the groups is Schwarz's Bayesian Criterion (BIC).

1. Introduction

Cluster analysis is an analytical technique used to classify or group data into finite and small number of groups, based on two or more variables. This analysis is a convenient method to identify homogeneous groups of objects or people who share the same characteristics [1]. There are several algorithms for cluster analysis and each of them aims to form groups based on calculating the measure of distance (measured distance) between observed individuals and observation groups. Grouping can be done according to individuals or variables. The data set consists of numeric, categorical or mixed variables, and different types of algorithms are constructed for each type of variable. Categorical data can be obtained from quantitative data or qualitative [2]. Cluster analysis applications appear in various fields. The data collected in the real world often contains both types of data, hence the mix. Traditional CA (Cluster Analysis) methods are difficult to apply directly to these types of data. In this paper we will use a method which can be applied simultaneously to mixed data [3]. Our paper focuses on the application of the two-step method of Cluster Analysis, which was developed in [4] to handle these types of data. The cluster analysis used is a two-step aggregation procedure in SPSS 25.0, which gives the user the ability to determine the appropriate number of groups, and then classify them using a non-hierarchical routine. This procedure is useful in

*Corresponding Author: Feruze Shakaj, Email: shakajferuze@gmail.com

this particular situation due to the sample size and the large number of variables being analyzed. Also, Garson in 2009 encourages the use of the two-step method for large data sets, using continuous and categorical data with three or more levels. The two-step aggregation method offers a particular advantage to leadership educators because of its ability to handle categorical variables such as gender, class level, and level of involvement,[5]. Like the K-means method this procedure can effectively handle databases that contain many records with data. From the name "Two-Step Cluster Analysis" it shows that the algorithm is based on two stages of approximation. The algorithm used in the first step is very similar to that of the K-means method. Based on these results, two-step cluster analysis leads to a hierarchical agglomeration method of CA, which combines objects sequentially to form homogeneous clusters [6]. Specifically, the two-step method of Cluster Analysis involves performing the following steps[7]:

- Preliminary collection
- Data collection in subgroups

Pre-grouping individuals or records into smaller subclusters, this step uses a sequential cluster approximation. In this step the data is scanned one by one and it is decided whether the current record should be merged with the previously formed group or classified into a new group based on the distance criterion [8]. SPSS implements this procedure by constructing a modified Cluster of features a (CF) tree according to (Zhang et al., 1996). This (CF) tree consists of node levels, where each node contains a number of inputs determined by the variable modality. Exactly one of these entries represents the desired sub-cluster [9].
The model assumes that the continuous variables x_j *(j = 1, 2, . . ., p)* are within cluster *i* independent normal distributed with means μ_{ij} and variances σ^2_{ij} and the categorical variables a_j are within cluster *i* independent multinominal distributed with probabilities π_{ijl}, where *(jl)* is the index for the *l-th* category *(l = 1, 2, . . ., m_l)* of variable a_j *(j = 1, 2, . . ., q)*. Two distance measures are available: Euclidean distance and a log-likelihood distance. The log-likelihood distance can handle mixed type attributes. The log-likelihood distance between two clusters i and s is defined as:

$$(i, s) = \xi i + \xi s - (i, s) \quad (1)$$

where

$$\xi_i = -n_i \left(\sum_{j=1}^{p} \frac{1}{2} \log\left(\hat{\sigma}_{ij}^2 + \hat{\sigma}_j^2\right) - \sum_{j=1}^{q} \sum_{l=1}^{m_j} \hat{\pi}_{ijl} log\left(\hat{\pi}_{ijl}\right) \right) \quad (2)$$

$$\xi s = -n_n \left(\sum_{j=1}^{p} \frac{1}{2} log\left(\hat{\sigma}_{sj}^2 + \hat{\sigma}_j^2\right) - \sum_{j=1}^{q} \sum_{l=1}^{m_j} \hat{\pi}_{sjl} log\left(\hat{\pi}_{sjl}\right) \right) \quad (3)$$

$$\xi_{(i,s)} = -n_{(i,s)} \left(\sum_{j=1}^{p} \frac{1}{2} log\left(\hat{\sigma}_{(i,s)j}^2 + \hat{\sigma}_j^2\right) - \sum_{j=1}^{q} \sum_{l=1}^{m_j} \hat{\pi}_{(i,s)jl} log\left(\hat{\pi}_{(i,s)jl}\right) \right) \quad (4)$$

ξv can be interpreted as a kind of dispersion (variance) within cluster *v (v = i, s,*$\langle i, s \rangle$). ξv consists of two parts. The first part $-n_v \sum \frac{1}{2} log\left(\hat{\sigma}_{vj}^2 + \hat{\sigma}_j^2\right)$ measures the dispersion of the continuous variables x_j within cluster v. If only $\hat{\sigma}_{vj}^2$ would be used, d(i, s) would be exactly the decrease in the log-likelihood function after merging cluster i and s. The term is added to avoid the degenerating situation for $\hat{\sigma}_{vj}^2 = 0$
ξ_i might be interpreted as a distribution within the group. Similarly, agglomeration methods, hierarchical collection methods as well as clusters with the shortest distance are joined in the same stream. The log-likelihood function for step k is calculated as:

$$l_k = \sum_{v=1}^{k} \xi v \quad (5)$$

This function lk might be interpreted as distribution within groups, but not exactly as the log-likelihood function. Where only non-geometric variables are used, lk becomes entropy within the number k of the groups.

The number of groups is given in advance and at each step of the procedure, this number is evaluated through two criteria which at the end automatically determine their number. Akaike Assessor Information Criterion (AIC) and Bayesian Information Criterion (BIC), which are determined by equations (6) and (7):

$$AIC_k = -2l_k + 2r_k \quad (6)$$

$$BIC_k = -2l_k + r_k \log n \quad (7)$$

where r_k represents the number of independent parameters, lk is the distribution within groups, k is the number of intermediate groups and n is the number of individuals [10].

Since the database with which we will work contains both continuous data and categorical data in our study we will use exactly the log-likelihood distance, which assumes that continuous variables have normal distributions and categorical variables have multinomial distributions. All variables are assumed to be independent just as individuals are. Most applications include variables that are somewhat related, so the Two Steps Cluster method is the best approximation to reality [11].

2. Purpose

Since 2006 in Albania with the development of educational reform the admission of students to Higher Education Institutions (IALs) has become more massive and universities, both public and non-public, have opened their doors to a large number of students by implementing thus the constitutional right of every Albanian citizen to be educated. From that year onwards, in addition to the existing fields of study, with the increased demand over time, a number of new fields of study have opened up and gained momentum, which have arisen as a need to the labor market. The purpose of this paper is to provide an overview of students admitted to IAL-s of the Republic of Albania, not only according to the respective densities in each program or university, but deepening further to simultaneously look at other characteristics of these students. One such example is the way they are distributed by gender, origin: by city or rural areas or even by the results achieved during secondary education, which in the database we will use, are reflected through their average for all years of high school. The intention is not only to give numerical data, but also to have a detailed look to see if there is

any connection between these indicators or if this distribution is random.

In other words, the purpose is to study if there is any connection between the student's gender and the program in which he/she is admitted. The secondary aim to this paper can be to help the relevant policy-making bodies, to improve secondary and higher education policies: the first by increasing the quality of curricula and infrastructure and the second by enriching departments with contemporary curricula, in order to meet the requirements of all students. This in turn would be seen as a challenge by students motivating them to see not only academic achievement and school learning, but also their satisfaction, and commitment to lifelong learning. This would also motivate them to open new doors and provide the right resources to achieve unquestionable academic success [12]. At last, this study can help all students studying outside Albania, giving an incentive to come and contribute to their country.

3. Database description

The database used in this study was obtained from the Center for Educational Services, part of Ministry of Education, Sports and Youth. This database contains all students enrolled in the first cycle of studies in 2017, in public and non-public universities of the Republic of Albania. The database consists of 22418 individuals and each individual is described by 4 independent variables, which are: study program, gender, high school average and type of residence, city or village.

The study programs are grouped according to the main fields of study such as: arts, agriculture, medicine, natural, social, sports. This collection is made in order for higher education institutions that offer study programs in each of these fields to be able to improve strategies to attract as many students as possible.

4. Experiment Results

As a start, the two steps method was used to identify the possible groups that are created by combining the modalities of gender, study program, student background (city or village) and the average achieved by them during the years of preuniversity education. The number of groups was given in advance equal to 15 and the Schwarz'Bayesian criterion was used to estimate the number of groups. Table 1 presents the results provided by SPSS

4.1. Auto-Clustering

Table 1: Self-collection table

Number of Clusters	Schwarz's Bayesian Criterion (BIC)
1	128265.515
2	104392.007
3	84160,103
4	67257.631
5	56414.898
6	47470,884
7	40938.323
8	35713,380
9	31882.708
10	28820.460
11	26626,089
12	24793.313
13	23030.864
14	21421,162
15	19901.124

From the table it is noticed that the values obtained by the criterion starting from cluster 1 and then for the second cluster 2 to cluster 10 have a significant difference. Starting from cluster 11 to cluster 15it can be noticed that the difference between the values that this criterion takes passing from group to group is negligible. Based on this criterion, the method itself proposes that the division into 10 clusters is the most appropriate.

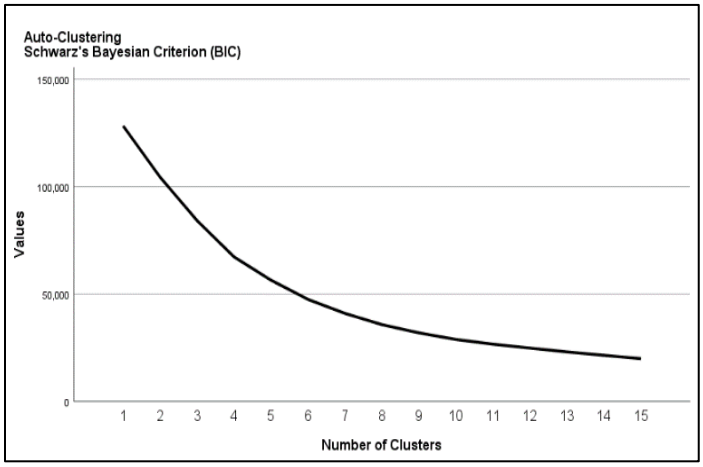


Figure 1: Number of groups based on grouping criteria

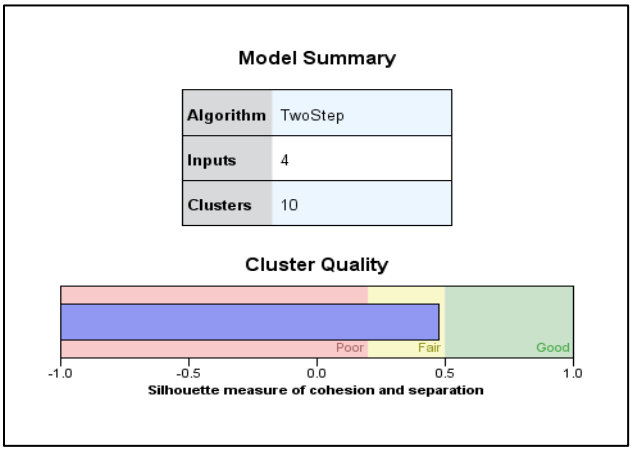


Figure 2: Number of groups based on grouping criteria

Another analysis, screen plot (see figure 1), confirms the above conclusions: the graphical representation of the pairs (number of gropuos - Criterion BIC) is in the same order as in table 1. Usually a suitable r can be found from the points where the gradient of the curve (curve) begins to become "flat" (See the position where this graph starts and becomes "flat"). In figure 1 a full screen can be seen, which graphically represents the variability of BIC values and serves to select the appropriate number of groups. It can be noticed that the best number of groups

is equal to 10 because the values of the BIC criteria differ very little per $k \geq 10$.

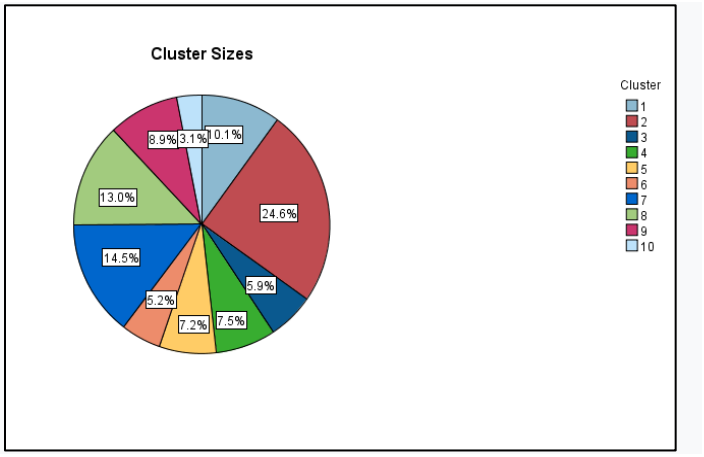


Figure 3: Group size, pizza graph for relative density of students in groups

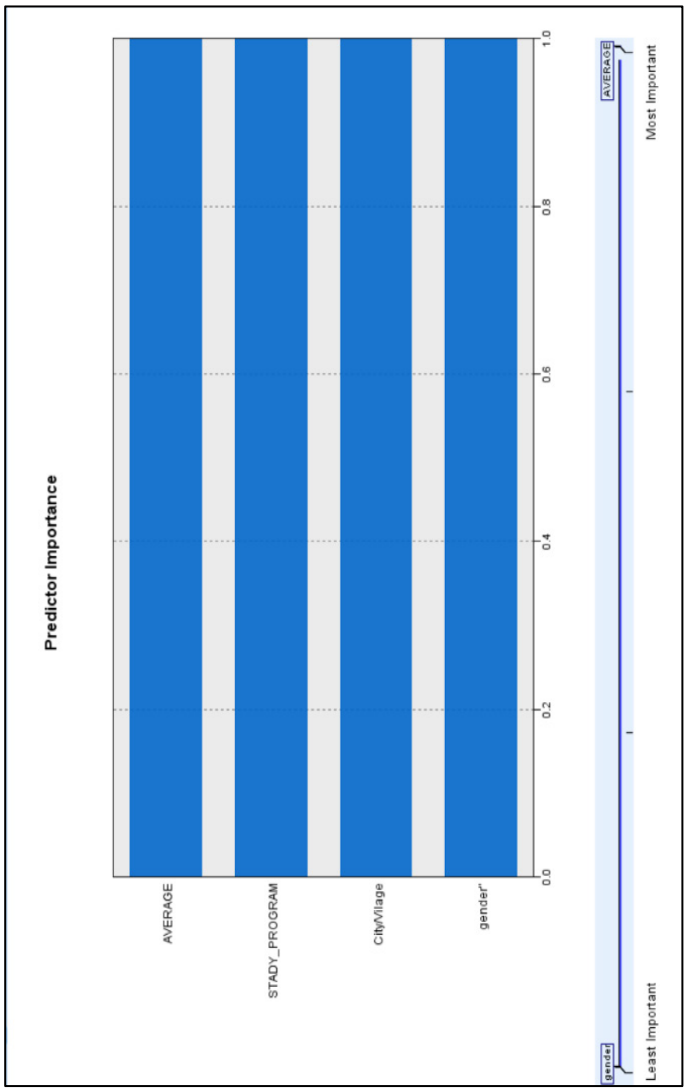


Figure 4: Weight of variables

Using the two-step method resulted in another important graph, figure 2 which shows that the individuals are divided into 10 groups based on 4 factors; gender, program of study, background and grade average. Also, the value of the silhouette indicator (the value of silhouette measure of cohesion and separation) which is less than equal to 0.5, shows that the 10 groups are well distinguishable from each other. The clustering can be called fair if the Silhouette measure of cohesion and separation is between 0.2 and 0.5. So, it can be noticed that the quality of the cluster is good because the mass of the silhouette of cohesion and case separation is close to the coefficient 0.5. It is orderly to emphasize that, if this indicator were less than 0, then this grouping would not make sense. Remember that the silhouette measure of cohesion and separation is an important indicator. Figure 3 shows the distribution of students expressed as a percentage in each group.

From figure 4, it can be noticed that the average grade has influenced more for the formation (difference) of the groups while the other three variables (study program, origin and gender) are ranked lower according to their importance in the writing order.

Using the step method resulted in another important figure, figure 6 which presents some characteristics for the 4 factors that describe the individuals of each group.

Some characteristics for each group are presented as follows:

- **cluster 1**-consists of 2269 students (or 10.1%), where all students are male. These students have chosen social field study programs and all come from the city with a grade average from high school equal to 7.29.
- **cluster 2**- consists of 5512 students (or 24.6%), where all students are female. These students have chosen socially oriented study programs and all come from the city with a grade average from high school equal to 7.74.

From this result it can be concluded that most of the female students coming from urban areas are oriented towards fields of study with social profile, and that their results during secondary education are not among the highest.

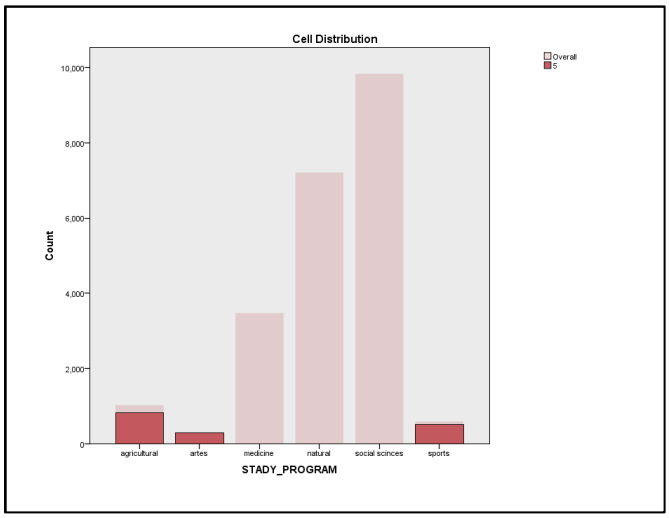
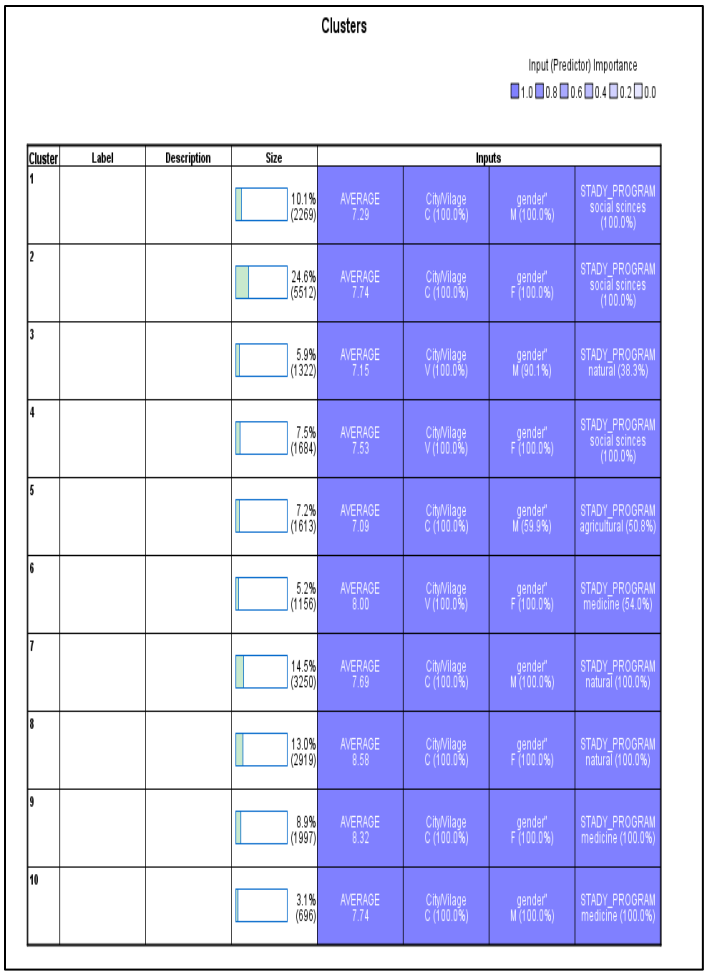


Figure 5: Distribution of variables in cluster 5



Cluster	Label	Description	Size	Inputs			
1			10.1% (2269)	AVERAGE 7.29	City/Vilage C (100.0%)	gender" M (100.0%)	STADY_PROGRAM social scinces (100.0%)
2			24.6% (5512)	AVERAGE 7.74	City/Vilage C (100.0%)	gender" F (100.0%)	STADY_PROGRAM social scinces (100.0%)
3			5.9% (1322)	AVERAGE 7.15	City/Vilage V (100.0%)	gender" M (90.1%)	STADY_PROGRAM natural (38.3%)
4			7.5% (1684)	AVERAGE 7.53	City/Vilage V (100.0%)	gender" F (100.0%)	STADY_PROGRAM social scinces (100.0%)
5			7.2% (1613)	AVERAGE 7.09	City/Vilage C (100.0%)	gender" M (59.9%)	STADY_PROGRAM agricultural (50.8%)
6			5.2% (1156)	AVERAGE 8.00	City/Vilage V (100.0%)	gender" F (100.0%)	STADY_PROGRAM medicine (54.0%)
7			14.5% (3250)	AVERAGE 7.69	City/Vilage C (100.0%)	gender" M (100.0%)	STADY_PROGRAM natural (100.0%)
8			13.0% (2919)	AVERAGE 8.58	City/Vilage C (100.0%)	gender" F (100.0%)	STADY_PROGRAM natural (100.0%)
9			8.9% (1997)	AVERAGE 8.32	City/Vilage C (100.0%)	gender" F (100.0%)	STADY_PROGRAM medicine (100.0%)
10			3.1% (696)	AVERAGE 7.74	City/Vilage C (100.0%)	gender" M (100.0%)	STADY_PROGRAM medicine (100.0%)

Figure 6: Groups

- **cluster 5**- it can be seen that the only students who have chosen to study in the fields of study of agriculture are gathered. It turns out that all these students are males from urban areas and it is also worth noting that this is the group with the lowest average. This conclusion is somewhat in contradiction with reality, as it would be more natural for students coming from rural areas to be oriented towards agricultural programs. Furthermore, for the group (cluster 5) as noticed in the cell distribution for the study program variable, they take the graph shown in Figure 5. What can be noticed is that, in this group there are also students who have chosen to study for arts and for sports. One of the reasons that these students are included in this group is that their number is very small and their averages during higher education are generally low (see figure 5).
- **Cluster 10**- consists of 696 students (or 3.1%) where the largest number are male students. These students have chosen medical-oriented study programs and all come from urban areas with a high school grade point average of 7.74. Naturally the question arises, why is this number so small?

From the results presented in figure 6 there can be distinguished some similarities or differences between the groups formed. It can be seen that cluster 1 and cluster 4, consist of female students who have chosen to study in socially oriented programs. These two groups differ by origin (city / village) and the latter come with a grade average (7.53) higher than those of group 1 (7.29). What is worth noting is the fact that the average during higher education in the fourth group is higher than in the first group. (see box plot, figure 7)

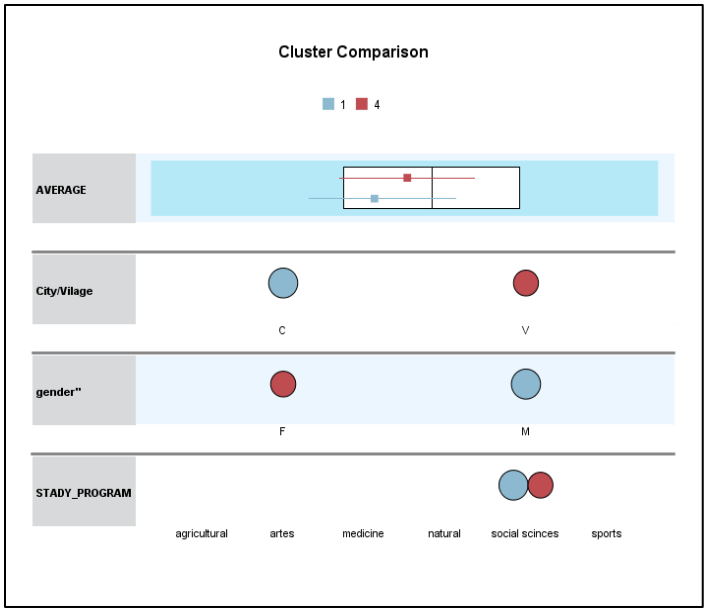


Figure 7: Comparison of clusters 1 and 4

It can be noticed that the females that make up cluster 4 have a higher average grade than the average grade of the males of cluster 1. The average grade of cluster 1 is closer to the overall grade average (7.76). This means that, the fourth group has higher performance than the first group in relation to this variable. Some of the reasons that can lead to this result are: higher accountability and demand from the teachers in urban areas, fictitious grading in rural areas, or simply the personal inclination of these students. Based on this result, it is up to the authorities, to make further studies to see if this fact is influenced or is natural.

In figure 6 it can be seen a cluster, which is cluster 3, where male students are gathered, who have chosen to study in study programs with natural direction and who come from rural areas, but there is not a such group with female students. So, none of the female students coming from rural areas have preferred or succeeded in winning these programs.

There is also cluster 6, where female students who have chosen to study in medical study programs originating from rural areas are gathered. It is worth noting that there is not a such group of male students. So, none of the male students coming from rural areas have preferred or succeeded in winning these programs. Another interesting fact to note is that the number of female students who have chosen to study in socially oriented study programs whether they come from urban or rural areas is significantly higher than the number of male students who have chosen to study in study programs in this regard. To prove this fact, it is enough to compare cluster 2 and cluster 4 with cluster 1 and see that the number of cases that have been collected in groups 2 and 4, the sum of cases is almost three-fold the number of cases

collected in group 1. This result is a fact that shows the natural tendency of women towards social sciences.

Let's remain in figure 6 and cluster 7 with cluster 8 the number of cases included in each group does not change much (respectively 3250 with 2919 cases). In these groups there are gathered male students who have chosen to study in study programs with natural direction, who come from urban areas (cities) and have an average of group 7.69, in cluster 7 and all female students who have chosen to study in study programs with natural direction and coming also from urban areas, but with a group average of 8.58. So, it can be observed that group averages have a significant difference (see figure 8). The same phenomenon is observed if a comparison between cluster 9 with cluster 10 is done (see figure 9).

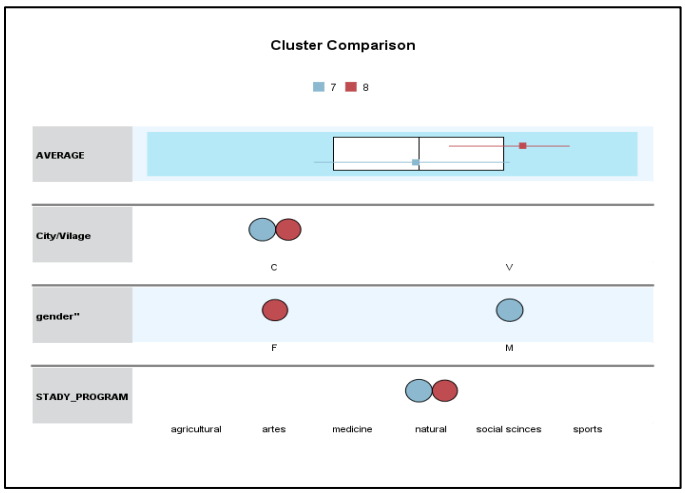


Figure 8: Comparison of clusters 7 and 8

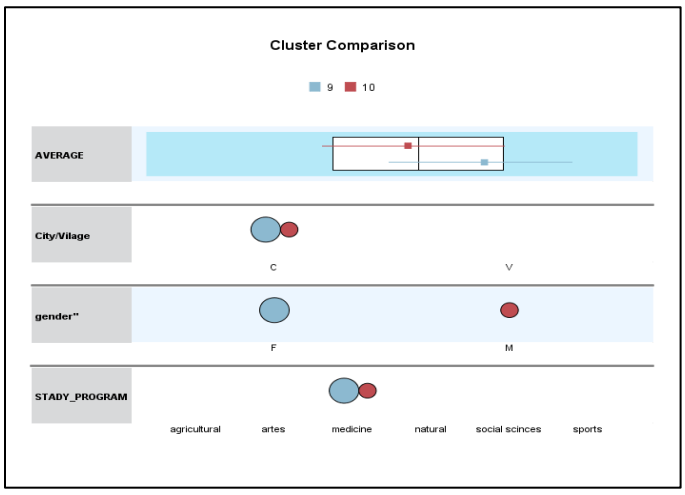


Figure 9: Comparison of clusters 9 and 10

In the first there are gathered female students who have chosen to study medical study programs, coming from urban areas, with an average of 8.32 and in the second there are gathered male students who have also chosen to study in medical study programs, with average 7.74. Again the difference in the means of these two groups is considerable. Can it be concluded that women are more intelligent than men? Based on the database we have taken in the study, this fact can be confirmed. To reach a more grounded conclusion it will be necessary to perform the analysis for several years in a row, and if a similar result is obtained, then the statement that "men are more intelligent than women" would be rejected. In the following, figure 9 and figure 10 graphically give what was just described in words above.

5. Conclusions

As mentioned above, the purpose of this paper is to provide an overview of students admitted to IAL-s of Albania not only according to the respective densities in each program or university, but further deepening to simultaneously see the other characteristics of these students, for example how they are distributed by gender, origin: from the city or from rural areas or even according to the results achieved throughout secondary education

At the end of the analysis, it can be reached to some conclusions which may be interesting:

The most important variable in creating clusters is the average. Female students from urban areas are oriented towards socially oriented study programs. Male students coming from urban areas do not, based on the analysis, prefer medical-oriented study programs. Female students who come from rural areas and have chosen socially oriented study programs have better high school results than same-sex students who have chosen the same major but who come from urban areas. All students who have chosen to study in agricultural study programs are male and come from urban areas. These students make up the group of the least grade average from high school. In natural and medical study programs, female students coming from urban areas have a higher average than male students coming from the same areas. No female students coming from rural areas have chosen or managed to win in natural study programs.

No male students from rural areas have selected or succeeded in earning medical degree programs. Female students coming from both rural and urban areas have chosen to study mostly in socially oriented study programs. Thus, showing the natural inclination of women towards the social sciences. It can be noted that all these results are valid for the database taken in consideration in this study. To see if these results can be generalized or not, the study should continue with the admissions in the IAL-s of the Republic of Albania in different years in other time spans or continuous years. This broader study is also authors's goal for further work and study.

References

[1] C.Matsimbe, A.A. Mmbanze, K. Gelb, J.A. Jonas, and G. Nhapuala " Use of Two-Steps Cluster Analysis to Understand How Emotions Affects Night Shift Teenagers' Students, "Journal of Education, Society and Behavioral Science, **24** (3): 1-15, 2018; Article no. JESBS.39264, 2018. DOI: 10.9734/JESBS/2018/39264

[2] M. Muça, K. Bani and F. Shakaj, "Combining the hierarchal and non-hierarcal methods for acluster analysis: a case study for classification of students accordin to their results.," In SPNA, Tirana, 2014.

[3] M. Shih, J. Jheng and L. Lai, " A Tw-Step Method for Clustering Mixed Categroical and Numeric Data.,Tamkang Journal of Science and Engineering, **13**(1), 2010. DOI: 10.6180/jase.2010.13.1.02

[4] T. Chiu, D. Fang, J. Chen, Y.Wang, and C. Jeris, "A robust and scalable clustering algorithm for mixed type attributes in large database environment.," KDD '01: Proceedings of the seventh ACM SIGKDD

international conference on Knoëledge discovery and data mining., p. 263–268, 2001.

[5] T.M. Facca, S.J. Allen, "Using Cluster Analysis to Segment Students Based on Self-Reported Emotionally Intelligent Leadership," Journal of Leadership Education,**10**(2), 2011. doi: 10.6180/jle.2011.13.1.02

[6] Online,"https://docplayer.net/12782167-Chapter-9-cluster-analysis.html".

[7] D. Şchiopu, "Applying TwoStep Cluster Analysis for Identifying Bank Customers' Profile," BULETINUL UniversităŃii Petrol - Gaze din Ploieşti, **LXII**, 66 - 75, 2010.

[8] M. Yaghini, "Two-Step Clustering Algorithm," 2010.

[9] T. Zhang, R. Ramakrishnan and M. Livny, "BIRCH: an efficient data clustering method for very large databases," ACM SIGMOD Record, **25,** 2, 1996.

[10] N.D. Moroke, "A TWOSTEP CLUSTERING ALGORITHM AS APPLIED TO CRIME DATA OF SOUTH AFRICA," Corporate Ownership & Control, **12**(2), 2015.

[11] "https://www.ibm.com/support/pages/how-log-likelihood-distance-method-applied-twostep-cluster-analysis#: ~: text = The% 20log-likelihood% 20distance% 20measure, be% 20independent% 2C% 20as% 20are% 20cases.) ".

[12] J.M. Marron, D. Cunniff, "What Is An Innovative Educational Leader ?," CIER, **7**(2), 145-150, 2014.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 44-51 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Remote Patient Monitoring Systems with 5G Networks

Antonio Casquero Jiménez*, Jorge Pérez Martínez

Signal System and Radiocommunication Department, Polytechnic University of Madrid, Madrid, 28040, Spain

ARTICLE INFO

Article history:
Received: 03 May, 2021
Accepted: 15 June, 2021
Online: 10 July, 2021

Keywords:
5G
IoT
Remote patient monitoring
eHealth

ABSTRACT

The new generation of mobile communications and the recent advances in data management are going to enable a fast transformation in the health sector of many countries. 5G networks, with superior technical characteristics, would allow the development of a new set of application and services gathered under the concept of eHealth. In this article we propose a remote monitoring system based on 5G networks that would allow to provide a varied set of medical services from long distance. However, for achieving an optimal performance, the network must guarantee high bandwidths and low latencies, at the time that a massive number of devices and its corresponding generated data are handle efficiently. Consequently, an appropriated system architecture and data model structure are proposed, taking into consideration the high security requirements that any health-related application or service inherently implies.

1. Introduction

Today, health systems in many countries are facing some important sanitary challenges regarding an increasingly aging population, the rise of chronic diseases and the Covid-19 pandemic. In this context, the digital transformation of the health sector with the development of new services supported by the latest technological advances is considered as an indispensable way of saving expenses, optimizing the resources of the sector, and obviously improving the populations welfare. As a result, concepts such eHealth and technologies as 5G or IoT have become essential elements in this digitalization process.

Most of the new and innovative applications and services that are expected to appear in the following years would arise from the combination of 5G and IoT and would be generally supported on Big Data and Artificial Intelligence techniques. Among them, remote patient monitoring systems would be one of the most popular and deployed applications.

This paper is an extension of the paper "5G networks in eHealth services in Spain: remote patient monitoring system", originally presented by the authors in the 2020 IEEE Engineering International Research Conference (EIRCON) [1]. In addition to the results and achievements presented in [1], in this current article we deepen in the theorical concepts behind the solution proposed, emphasising the role of 5G networks. Although the general system architecture is the same as the one presented in [1], this paper extends its study providing a more detailed description of the involved elements and their function within the system. Finally, and in addition to the initial results gathered in [1], the security and data privacy of the remote patient monitoring system proposed is analysed.

1.1. The remote patient monitoring service: definition and communication requirements

Remote patient monitoring (RPM) is an eHealth application that consists in the monitorization of a patient's health state with sensors, wearables or medical devices that measures vital and contextual parameters such as temperature, pulse, sugar or oxygen in the blood, among others [2]. Moreover, the term RPM also includes the applications and platforms that allows the analysis of medical data by means of BigData and artificial intelligence (AI) techniques to provide self-consultation and evaluation by doctors. In general, these applications are especially designed for the monitoring and follow-up in real time of patients with different illnesses, although the uses of RPM systems can be directed to other use cases such as the follow-up of post-operative processes.

According to the predictions done by STL Partners [3] remote patient monitoring applications would be one of the main eHealth services in the next decade, experimenting a tremendous growth, following the trends of a more personalized and continuous home centric health care. In addition, and thanks to the use of 5G networks, the forecasts determine that the adoption of these kind

*Corresponding Author: Antonio Casquero, Polytechnic university of Madrid, Madrid, 28040, Spain, +34649034991, acasquerojimenez@gmail.com

https://dx.doi.org/10.25046/aj060406

of applications could lead to more than 50 Billion USD of global annual cost savings for the health sector by the year 2030 [3].

However, these kinds of applications impose important communications requirements being the availability of a wide coverage area the most important one, as it is mandatory for a remote patient monitoring system to connect all the agents that might be involved such as patients, doctors, caregivers, or even patients´ family members [4]. Nevertheless, it is also crucial to consider the network's capacity to efficiently manage a large number of connected devices, as the network that supports the system must be able to handle an increasing deployment of devices without impacting the service performance. In this line, the security and reliability of the data transmitted must be guaranteed, as well as the continuity of the service so that patients are monitored at all times. Special attention must also be paid to coverage in indoor environments, as most of the sensors and devices deployed will be located in these environments [4]. On the other hand, both energy consumption and battery life of the devices are critical aspects to consider in order to have self-sustaining connected devices, whose batteries are adapted to the duration of the medical processes. Likewise, remote monitoring aplications imposes requirements related to high-speed mobility, especially in the case of emergency situations.

In the following table the main communication requirements can be summarized. However, it is worth mentioning that these would considerably depend on the medical service derived from the system.

Table 1: Technical requirements of RPM systems [5]-[7]

Use case attribute	Value
Throughput	< 1Mbps
Latency	<50ms
Reliability	99,99%
Number of devices	10-10^4/km^2
Battery duration (use case dependant)	10 years
Security	Critical
Mobility	0-500 km/h
Coverage	Important, including indoor
Indoor coverage	No critical 1-10 m horizontal, < 3m vertical

1.2. Organisation of this paper

This paper is structured as follows. In section 44 we have provided the definition of remote patient monitoring as well as an analysis of the requirements imposed by these types of systems. In section 2 we study the main technologies involved in the provision of the connectivity required in RPM systems whereas in section 3 we present a RPM system architecture based on 5G networks. Afterwards, in section 4, a data model for the adequate treatment and management of the data in the system is proposed. In the next section, two crucial aspects such as the security and data privacy of the system are reviewed playing special attention to the advantages of 5G in this context and the possible additional measures that could be taken. Finally, in section 6, the conclusions of the paper are presented.

1.3. Related projects and papers

The development of 5G networks as well as the possible services derived from them are still in an early stage, moving from the commercial verification period to the small-scale deployment of certain solutions. Consequently, to the best of our knowledge, there are no commercial systems or solutions such as those proposed in this paper, although it is true that in the specialized literature, we do find truly interesting proposals.

In [8], the authors present a continuous monitoring system based on 5G networks using wearables and sensors. The devices measure certain vital parameters and via Bluetooth, they send the captured data to the smartphone, which then forwards the data to external servers through 5G networks. In this project, they also propose an intelligent algorithm for decision making and alert generation based on the evaluation of the measured parameters. On the other hand, in [9] they propose a remote monitoring system for smart environments based on IoT and on the existing 4G network infrastructures. In this article, they analyse the communication requirements of the system, especially those referred to the bandwidth, and perform a detailed study of the communication protocols. In this context [10] designs a home monitoring system based on 5G networks and Edge computing with the scope of treating remotely chronic diseases and to promote active aging, in line with the developments and advances reported in [11]. From another perspective, in [12] they also propose the joint use of short-range networks and mobile networks, as an optimal way to develop a remote monitoring system, putting in this case special emphasis on the security of the system and its vulnerabilities. Finally, and with a more general approach, in [13], in addition to a complete review of 5G technology and its applicability to the healthcare sector, they propose a 5G network architecture capable of supporting multiple medical services, including patient monitoring. The system is characterized by the joint use of small cells and macro cells and the deployment of computing and processing servers in the edge. The requirements that eHealth applications impose on telecommunication networks and, in particular, the suitability of 5G networks to meet them are also presented in [13].

2. Connectivity Technologies

The development of eHealth services and applications, like the one we are analysing in this paper, would appear from the combination of 5G and IoT. However, it is worth mentioning the important role that Big Data and Artificial Intelligence techniques and algorithms would play regarding the intelligent analysis of the data generated by these systems. Indeed, the role of data treatment and management platforms is considered indispensable, as they provide a secure and ubiquitous source of information for the consultation of the generated data among all the agents involved in a particular service [2].

However, in this section we focus on the main connectivity technologies involved in the provision of remote patient monitoring systems. These technologies must take into

consideration the requirements defined in section 1.1 and must satisfied them properly.

2.1. 5G

5G is the new paradigm of wireless communications designed to support a wide variety of services. 5G networks improve considerably the performance of the previous generations of mobile communications thanks to the introduction of a new radio interface known as "the New Radio" (NR) and the redefinition of the core of the network resulting in the 5GCN (5G Core Network) [14].

In addition, the 5G system is defined as a service-based architecture (SBA) that through the recent advances in network function virtualisation (NFV) and software-defined networking (SDN) allows a flexible usage and configuration of network functions. As a consequence, different use cases with very diverse requirements can be defined by means of network slices and consequently multiple verticals can develop their services within a single physical infrastructure [15].

The new frequency bands, which include the ultra-high capacity mmWaves, the new waveform used, the utilization of massive MIMO techniques, or the use of heterogeneous access networks, which supports non 3GPP standardised access networks, jointly with the improvements in the network architecture, allows 5G networks to offer the following characteristics [1]:

- Guaranteed user data rate of 100Mbps in DL and 50 Mbps in UL with peak rates of 20Gbps.
- 1/10 X in end-to-end latency, reaching delays of 1 ms.
- Service transmission reliability of >99.999%
- 1000 X in number of IoT devices reaching a density of 1 million terminals/km^2
- 1/10 X in energy consumption

Finally, it is worth mentioning the recent advances in Multi-access edge computing (MEC) technology regarding 5G networks. This concept enables to bring cloud computing processes and storage (typically performed in the core of the network) closer to the final users. Running these processes in close proximity to end users, for example in the 5G base stations, reduces network congestion and improves applications performance by providing faster and more reliable connections with lower latencies [16].

2.1.1. *Why 5G?*

The characteristics and advantages of 5G networks in the provision of medical services can be summarized in the following points:

- Reliability and security. The ultra-reliable connections of 5G networks are the main driver for the utilization of this technology for the development of eHealth applications as patient´s data privacy and service continuity can be guaranteed. Network slices isolation, advanced data encryption techniques and the new and improved authentication mechanisms contribute to this objective.
- High performance. As we have stated in the previous section, 5G networks provide high average data rates and ultra-low latencies, fulfilling the requirement of a wide majority of medical services, including the one we are analysing. Massive MIMO with 3D beamforming and Edge computing and MEC (Multi-access Edge Computing) platforms are some of the innovations included in this field of action.
- Ultra-high Capacity. With up to one million of devices per Km^2,5G networks can efficiently manage the simultaneous connections of multiple medical devices from multiple users. Beside this, these networks provide 90% reduction in energy consumption with 10 years of battery life for low power IoT devices.

2.2. IoT

IoT ("Internet of things") consists in the grouping and interconnection of devices and objects through a network. As we can infer from this definition, RPM systems can be classified as an IoT application, in which the technology chosen for the provision of such connectivity would be 5G [1]. In 5G networks, we can define different categories of IoT according to their requirements [17]: Massive IoT, Broadband IoT, Critical IoT and Industrial Automation IoT. The characteristics of RPM system can be included within the framework of the massive IoT and Broadband IoT.

2.2.1. *Massive IoT*

Massive IoT intends to provide connectivity to a very large number of devices that transmit and/or receive small volumes of data. These devices, that frequently rely on battery power supply, are usually low-cost and can be located in remote places with little coverage [18].

In this context, LTE-M (LTE-MTC) and NB-IoT (Narrowband-IoT), both standardized by the 3GPP [19], fulfil all 5G requirements from both ITU and 3GPP for massive machine type communications. Although they are both LTE technologies, they are designed to coexist within the new 5G NR, thanks to techniques such as Dynamic spectrum sharing (DSS), defined in Release 15, and the numerology used by the sub-carriers, compatible with the NR.

Table 2: Technical comparison of NB-IoT and LTE-M [18].

Characteristics	NB-IoT	LTE-M
Bandwidth	180 KHz	1.4 MHz
Subcarrier bandwidth	15 KHz	15 KHz
Peak data rate	250 Kbps	1 Mbps
Latency	1.5-10 s	50-100 ms
Battery duration (use case dependant)	+10 years	10 years
Operation mode	FDD	FDD/TDD
Voice support	No	Yes
Coverage DL (MCL)	164 dB	164 dB

Coverage UL	Gain: 20dB	Gain: 15 dB
Indoor coverage	Excellent	Good

2.2.2. *Broadband IoT*

Broadband IoT is a category of IoT developed from mobile broadband communications (MBB). The introduction of 5G with the NR and the new frequency bands, allows offering IoT services with higher data rates and lower latencies than those offered by mIoT technologies, reaching peak values of tens of Gbps in transmission and latency values of 5ms [17]. Currently, many broadband IoT applications are being developed due to the already mentioned combination of MBB features (high bandwidth, high transmission speed, low latencies) with the characteristics of IoT services (wide coverage range, energy efficiency).

In this context, it is worth mentioning that the 3GPP as part of Release 17 [20], has introduced a new concept defined as the 5G NR-Light. It consists in a set of modifications of the 5G NR in order to reduce its capabilities to efficiently develop a set of IoT applications that will require devices more complex than those used in mIoT (massive IoT) communications but less complex than those used in 5G NR. These devices are usually going to be wireless industrial sensors (with low latencies and moderate transmission rates), video systems, high-capacity wearables, and patient monitoring systems.

3. RPM system architecture

Remote patient monitoring systems (RPM), as previously discussed, allows the monitoring of an individual's health status using sensors and/or wearables thanks to the control and measure of certain relevant medical indicators and parameters. Currently, we can find some remote monitoring systems models and proposals (section 1.3) in which these devices use personal area networks (PAN) such as Bluetooth and local area networks as Wi-Fi to connect to a hub (e.g. a smartphone) or a gateway. The data are collected and processed in mobile applications for smartphones that allow the patient to monitor and manage relevant medical information.

However, these applications are generally not standardized and therefore the monitoring processes performed does not have the required medical rigor. The systems should be able to send, periodically, or in real time, this medical data to authorized third parties, in order to obtain a more exhaustive analysis involving healthcare professionals and the use of Big Data techniques. This will facilitate diagnosis and follow-up tasks without the patient having to travel to the hospital or primary care center. In addition, it would be possible to detect anomalous patterns in the data and activate alarms in emergency situations.

Taking into consideration these aspects, an once we have described what is a RPM system and the main technologies involved, we present a possible system architecture based on 5G networks [1].

3.1. Functional high-level architecture

The system we propose is an open multiplatform system, based on European standards, which will allow the continuous monitoring of the health state of patients, as well as the possible detection of new pathologies and new therapeutic approaches.

The design of the network, responsible of providing the required connectivity, must be in line with the requirements of the proposed system where we highlight the massive creation of data by the end users (directly or indirectly) from multiple sources. In this context, the system and the network must guarantee the availability and security of the data so that the agents involved in the provision of health services can make use of the multiple applications and services that can be offered from the exploitation of these.

In order to obtain an open and scalable solution, both the data model and structure and the IoT architecture are based on the ETSI reference architecture, particularly on the ITU-T Recommendation Y.2060 [21], which offers the description of a reference model for IoT applications.

In the Figure 1 and taking the multilayer ETSI reference model as a reference, the functional high-level architecture of the system proposed is presented.

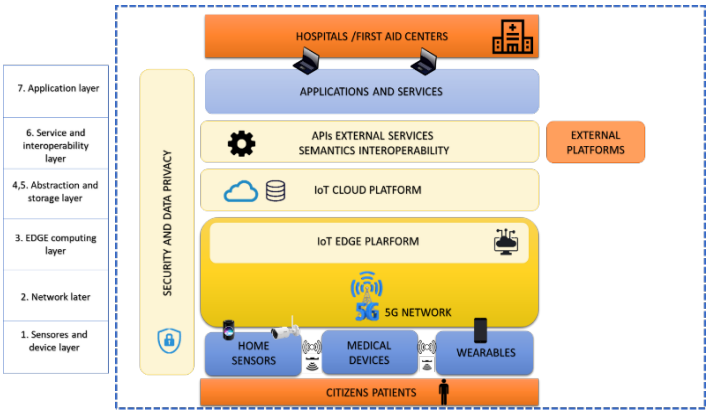


Figure 1. Architecture of the proposed RPM system together with the ETSI reference layer model [1]

3.2. Architecture description

In this subsection we provide a detailed description of the different layers that define the system. The security and data privacy module, which covers all the layers of the system, will be studied in detail in section 5.

3.2.1. *Sensors and device layer*

This is the physical layer of the system that encompass a set of devices and sensors responsible of collecting information about the health status of the patient and of its environment [10]. In this layer we can find contextual sensors, medical devices and wearables that transmit and/or receive information directly over the mobile network without the need of using local networks (LAN OR WLAN) or gateways (routers or smartphones). However, we must also highlight the role that smartphones or virtual assistants can play in the system, due to the considerable number of sensors that they have integrated.

3.2.2. *Network layer*

5G networks will be the telecommunications network in charge of providing the connectivity in the system.

In our solution, we would mainly use low-cost devices and sensors that will perform periodic measurements of certain vital and contextual parameters, using narrow bandwidths. These would deal with very small amounts of data and will not require low latency values. Therefore, the best solution, as this scenario matches the characteristics of mIoT, is to use LTE-M or NB-IoT networks [1].

On the other hand, and in order to provide high performance monitoring applications, we can find another group of devices (advanced medical devices or high capacity wearables) that will perform measurements and monitoring processes of higher capacity. Consequently, they will require higher transmission rates and generally lower latencies. These devices would work with higher data volumes, and their requirements would be similar to those associated with Broadband IoT applications. As a result, we will use the 5G NR or 5G NR-Light [20].

Finally, we must highlight the possibility of using other wireless technologies, such as Wi-Fi 6, for connectivity at indoor environments. The new Wi-Fi standard, officially named as 802.11ax, is capable of providing theorical peak rates of 9.6 Gbps and can quadruplicate the number of simultaneous connected devices [22]. This new Wi-Fi standard, thanks to new security mechanisms and to the definition of network energy efficiency techniques, can effectively provide mIoT communications. The main advantage of this technology is the greater variety of sensors and devices that can be used, since the Wi-Fi 6 access points (AP) allows the integration of other communication modules (by means of external cards) associated to IoT communications such as RFID, ZigBee, and Bluetooth [22]. In this particular case, Wi-Fi 6 could provide indoor connectivity and 5G networks would act as backhaul technology.

This hybrid solution, which offers an optimized performance, is technically feasible thanks to the fact 5G network support multiple access technologies, even if they are not 3GPP standardized.

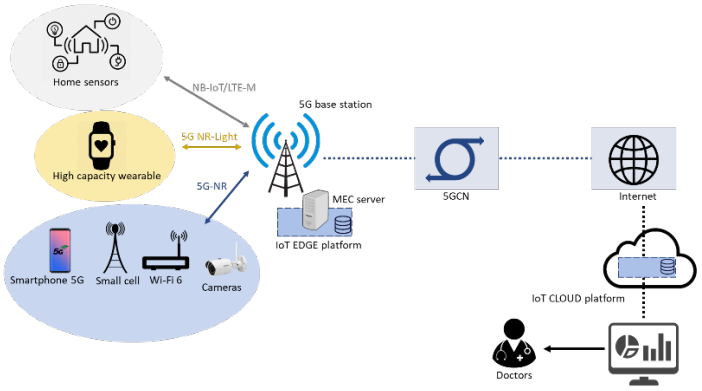


Figure 2: Representation of the global system architecture

3.2.3. *Edge computing layer*

The main function of IoT Edge platform is the integration and aggregation of the data collected in the previous layers and the transmission of the data to the cloud. The platform also carries out filtering and processing tasks in order to reduce the volume of data to be transmitted to the upper layers and consequently the bandwidth required. In this layer, a first analysis of the data can be made as analysing it in the proximity of the users makes it possible to improve the reaction or intervention in real time in cases of risk or emergencies and allows the execution of AI processes based on the collected data with very low latencies.

3.2.4. *Abstraction and storage level*

The abstraction and storage layer is encompassed by the IoT cloud platform, which is in charge of collecting the information processed by the previous levels, gathering the information from all the IoT Edge platforms deployed in a health region (south-bound) and integrating the semantic interoperability layer (north-bound).

Moreover, this platform is responsible of performing cloud computing tasks from the pre-processed data in the previous layers. At this level of the system is where the management of identities is performed as well as storage and administration tasks. The proposed platform is also where we carried out the intelligent analysis of the data [1]. In addition, thanks to the use of data models and information collection standards, the IoT cloud platform can manage information from various different sources such as IoT Edge platforms (user-generated data) or health systems (clinical history databases).

3.2.5. *Semantic interoperability layer*

The concept of semantic interoperability layer was introduced and developed in the ACTIVAGE project [11], a European project which develops IoT solutions for Smart Living Environments. However, the role of this layer is crucial for the proper development of the system and therefore it must also be included in our design.

The semantic interoperability layer allows the described IoT platforms to share data and exchange information with other external platforms thanks to the definition of interfaces and to the conversion of the data to a common model. Data from different sources are harmonized in using HL7 and the SNOMED vocabulary[10]. This layer also includes security and access control functions to the upper layers.

3.2.6. *Application and intelligent services layer*

The application layer enables the provision of a wide range of intelligent services from the captured and processed data. They are categorized as intelligent because in all of them there may be data processes that use artificial intelligence and/or big data techniques. The offered services are intended for patients but also for healthcare personnel or even informal caregivers of the patients. The final services can be adapted to the socioeconomic context of the region in which the system is deployed.

The most relevant medical services that can be derived from the proposed system are now presented. These services can address the main sanitary challenges considered in section 1:

- Chronic care services. Chronic diseases are progressively increasing every year over the population of many countries [23]. The RPM system proposed is expected to allow a personalized management and administration of patients with these conditions thanks to a rigorous and continuous follow up of the medical and contextual parameters that reflect their condition. In addition, the system could favor the capacity for

early detection of diseases thanks to the predictive analysis of data running both in the Edge and in the cloud.

- Elderly care services. In this category we can include any service aimed at maintaining and improving the quality of life of the elderly. In this line monitoring services can facilitate the independence of the elderly allowing them to maintain an autonomous life away from health institutions for a longer period.
- Personal health self-management services. These are services aimed at people whose objective is the prevention of possible chronic diseases based on the detection of one or more conditions but in which there is still no organ damage.

4. System data model

The proposed RPM system, as can be inferred from the functional description of it, can be considered as data centric. The massive amount of data that the system has to handle jointly with the demanding security requirements that any health-related application imposes, demands a reorganization of the traditional data management procedures. The changes required in this respect, side with those introduced in 5G networks.

The aforementioned shift of paradigm implies that beyond the management of the devices involved in the monitoring processes, the system must establish mechanisms and tools for the efficient management of the data itself. In this context, and following the indications reflected in [24], the system should consider the following aspects:

- Data ownership. Patients should have control over the data generated by their monitoring devices. To this end, we must establish an ownership structure for the data and the devices, that allows to establish a relationship and association among them. This structure must allow the patient to manage, locally or remotely, the generated data.
- Data provenance. It is essential to create an immutable record of the source of the data, thus establishing a unique version of it. The system must be able to identify the device that generated the data and at what time it was generated. This information makes possible to guarantee the authenticity and veracity of the information collected.
- Data governance. Patients should be able to manage the access to their data, being able to grant permissions to third parties for external consultation.

Therefore, we can conclude that for data management in 5G networks, and by extension in our system, we must manage the data generated and the users with access to them. One possible solution consists in the utilization of two separate platforms, one in charge of property management and the other of data storage [24].

These concepts are particularly relevant in certain processes and management tasks that the system must perform. In the Figure 3, we show an example which reflects the registration of a device in the system and the access of an authorized third party to the data generated by it. In this scenario we have three entities: the patient, owner of the medical device, the medical device itself, which generates the data, and the healthcare personnel, who wishes to consult the generated medical data.

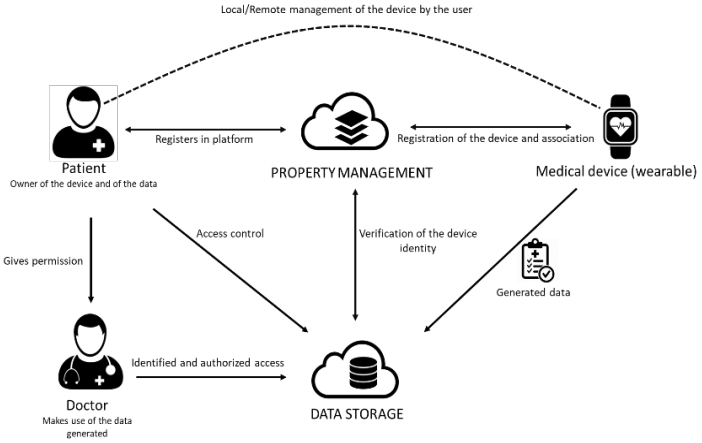


Figure 3: System´s data management model. Based on [24]

As part of this registration process, the patient must register in the system in order to be able to add and associate to his/her identity different medical devices. This process will allow the user to manage the device and control the data generated by the it. Once the registration is completed, and as part of the monitoring process, the data generated will be stored in a data storage platform. This platform will verify the identity of the device and the authenticity of the stored data, which can be consulted by the patient or by authorized third parties.

5. Security and data privacy in the system

The proposed system is responsible of the creation, maintenance and exploitation of the data generated by the patients. In this context, as we are dealing with an application framed within the health sector, there is a series of legal obligations related to data security and privacy that must be considered and therefore, the system must be designed in line with the General Data Protection Regulation (GDPR) established by the European Union. Consequently, the Security and Privacy (S&P) module defined in our system architecture must cover all the layers, from the sensors and devices to the final applications, ensuring three main security principles: confidentiality, availability, and integrity.

In this context, 5G networks introduce a series of improvements in the security domain compared to previous generations of mobile networks. Moreover, we would define a set of specific security measures to enhance the system overall security.

5.1. 5G security improvements

5G networks offers better capabilities than other mobile technologies not only in terms of bandwidth, latency or density of devices that can be supported, but also in terms of security and data privacy. In this context, the main improvements are related to authentication procedures and to the encryption of the data in the network.

Authentication and identity management are fundamental aspects in cellular networks. In 4G networks, authentication tasks are performed establishing a mutual authentication process in which, by means of the authentication and key agreement (AKA)

protocol, both the user's terminals and the network core validate their identities [25]. However, these processes do not include the encryption of the data in the access network in certain signalling processes, for example, when sending the user's identity. To this end, 5G networks encrypt the international mobile subscriber identity (IMSI) and extends the length of the session keys from 128 to 256 bits, thereby improving the protection of communication between the device and the network core [25]. In addition, fifth generation networks introduce a number of specific network functions designed to improve the security of authentication processes, such as the AUSF (Authentication Server Function) or the SEAF (Security Anchor Function) and defines a unified framework with new authentication methods: 5G-AKA, EAP-AKA and EAP-TLS [26]. This framework enables the authentication of a user's equipment independently of the access technology used, whether it is standardized by the 3GPP or not. This fact responds to the heterogenous access capabilities of 5G networks and is a concept extremely important in our system as we propose different access technologies as part of the solution, some of them non 3GPP standardized. Consequently, and in all situations, the user's data privacy would be guaranteed.

On the other hand, and considering the wide variety of medical devices and sensors that can that take part in the monitoring processes, the aforementioned authentication framework includes alternative authentication methods, such as the EAP-TLS, formed by an extensible authentication protocol (EAP) with transport level security (TLS). This method is expected to be considerably used in IoT applications as it does not require the use of (U)SIM cards for identity management and authentication processes. This could facilitate the use of low cost devices with different sizes and shapes [26].

Nevertheless, there are some other important aspects regarding 5G networks security, such as the isolation between the access network and the 5GCN. In this context, in the 5G system there is a clear separation between these two parts, in which the access network performs radio management tasks, and the core manages and controls network resources and security[26]. This separation makes it possible to easily identify and isolate network elements in case they are under attack. As a result, the core is the responsible of ensuring the user authentication and the encryption of the data and signalling traffic, with the advanced encryption standard (AES), whereas the access network sends the encrypted traffic between the user's equipment and the core, thus protecting traffic at the radio interface. In addition, the definition of a service-based architecture in the network core allows protection mechanisms to be applied at higher layers, e.g. transport and application [27].

Finally, we must mention the advantages of defining network slices to provide these kinds of services, not only in terms of network resources allocation, but also in terms of security, as possible failures or attacks in one slice will not affect the services being provided in others. To this end, mechanisms and tools must be developed to allow isolation between slices and to prevent unauthorized users from accessing both the assigned network resources and the information they carry.

5.2. Specific measures

However, and despite of the improvements in terms of security and data privacy that 5G networks include, the highly regulated healthcare sector also imposes the need of the definition of an appropriated data model structure. This model, which has been studied in the previous section, responds to the Edge computing paradigm, so that part of the data, before being send to the cloud, is processed, and managed in a secure space. Therefore, part of the management, processing and storage tasks of the data will be carried out on the IoT Edge platform. This platform will be able to manage, independently and autonomously, the users and the devices associated to them [16].

On the other hand, and as the system will rely on the cloud, the communications between the different platforms and cloud services must be performed securely. In this context, a research group of the Polytechnic University of Madrid called the Life Supporting Technologies group, recently proposed the use of distributed identity systems for this kind of purposes, with the scope of integrating different healthcare services [10]. The proposed distributed system will be the one we will use in our system and will be in charge of managing user identities anonymously. In addition, it will also incorporate smart semantic contracts (SSCs) to control data access, its usage and conditions. The SSCs will be able to unlock encrypted data only when the conditions of the contract are met, so it will be possible to transmit encrypted data between different systems without affecting data privacy.

6. Conclusions

In the following decade, and accelerated by the current situation, the health sector of many countries would experience a deep digitalization process. As part of it, 5G networks would play an important role as they would allow the development of new applications and services. Among them, we highlight remote patient monitoring systems that would permit to offer a more personalized and continuous healthcare from long distance.

In this article, we have presented a possible RPM system architecture based on 5G networks. In this context, the new mobile communication generation offers not only a far more superior performance but also a more secure and reliable environment for the management and treatment of the generated data. In order to improve the general system performance, we also make use of other important technological advances in other fields such as IoT or MEC computing. The development of IoT systems, based on 5G networks, with more accurate and enhanced characteristics and the improvement of MEC computing would allow to considerably upgrade the final performance of the system. As a result, a wide flexible variety of medical devices can be offered, as the system is inherently flexible and adaptable.

Finally, the new enhancement in 5G networks security and data management together with additional security and data privacy measures would allow to give the system the required security. This aspect is crucial as 5G networks can help to achieve the severe regulations and security measures that health systems and applications suffer.

References

[1] A. Casquero, J. Perez, "5G networks in eHealth services in Spain: Remote patient monitoring system," In IEEE Engineering International Research Conference, EIRCON 2020, 1–4, 2020, doi:10.1109/EIRCON51178.2020.9254013.

[2] IDATE, "Ehealth Market trends, players & outlook," DigiWorld Interactive Platform, 1–38, 2018.
[3] D. Singh, "5G'S HEALTHCARE IMPACT: 1 BILLION PATIENTS WITH IMPROVED ACCESS IN 2030," STL Partners, 1–39, 2019.
[4] IDATE, "5G IoT in the healthcare sector," DigiWorld Interactive Platform, 56–67, 2019.
[5] Next Generation Mobile Networks Ltd, Perspectives on Vertical Industries and Implications for 5G, NGMN P1 Requirements & Architecture - Verticals Requirements, 1–41, 2016.
[6] Next Generation Mobile Networks Ltd, Recommendations for NGMN KPIs and Requirements for 5G, NGMN P1 WS#3 Key Performance Indicators to 3GPP TSG RAN, (2016), 1–19, 2016.
[7] 3GPP ETSI 3rd Generation Partnership Project, "5G; Service requirements for next generation new services and markets (Release 15)," 3GPP TS 22.261 Version 15.5.0 Release 15, 1–52, 2018.
[8] R. Singh, R. Garhewal, "Smart M-Health Continuous Monitoring System Using 5G Technology," Research Project University of Ottawa, 1–7, 2018, doi:10.13140/RG.2.2.15946.21440.
[9] Ngo Manh Khoi, S. Saguna, K. Mitra, C. Ahlund, "IReHMo: An efficient IoT-based remote health monitoring system for smart regions," 2015 17th International Conference on E-Health Networking, Application and Services, HealthCom 2015, 563–568, 2015, doi:10.1109/HealthCom.2015.7454565.
[10] MYSPHERA, Life SupportingTechnologies Group, Universidad Politécncia de Madrid, "Public consultation for the remote monitoring of chronic patients," Red.Es, 1–35, 2019.
[11] P. Barralon, B. Charrat, I. Chartier, V. Chirié, G. Fico, S. Guillen, N. Homehr, O. Horbowy, R. Kamenova, F. Lamotte, A. Peine, "IoT for Smart Living Environments Recommendations for healthy ageing solutions," WG5 - Smart Living Environment for Ageing Well – Recommendation Paper Alliance for Internet of Things Innovation, (April), 37–67, 2019.
[12] P.K.D. Pramanik, G. Pareek, A. Nayyar, "Security and privacy in remote healthcare: Issues, solutions, and standards," Telemedicine Technologies: Big Data, Deep Learning, Robotics, Mobile and Remote Applications for Global Healthcare, 201–225, 2019, doi:10.1016/B978-0-12-816948-3.00014-3.
[13] A. Ahad, M. Tahir, K.-L.A. Yau, "5G-Based Smart Healthcare Network: Architecture, Taxonomy, Challenges and Future Research Directions," IEEE Access, **7**, 100747–100762, 2019, doi:10.1109/ACCESS.2019.2930628.
[14] S. Redana, O. Bulakci, "View on 5G ArchitectureView on 5G Architecture - 5G PPP Architecture Working Group," 5G Public Private Partnership (5G PPP), 1–182, 2020, doi:10.5281/zenodo.3265031.
[15] G. Brown, "Cloud RAN & the Next-Generation Mobile Network Architecture," Heavy Reading White Paper Huawei Technologies Co. Ltd., 1–9, 2017.
[16] Y.C. Hu, M. Patel, D. Sabella, N. Sprecher, V. Young, "ETSI White Paper #11 Mobile Edge Computing - A key technology towards 5G," ETSI White Paper No. 11 Mobile, 1–16, 2015.
[17] A. Zaidi, A. Bränneby, A. Nazari, M. Hogan, C. Kuhlins, "Cellular IoT in the 5G era," Ericsson White Paper, 1–17, 2019.
[18] C. Kuhlins, B. Rathony, A. Zaidi, M. Hogan, "Cellular networks for Massive IoT," Ericsson White Paper, 1–16, 2019.
[19] 3GPP Technical Specification Services Group, "Release 15 Description-TR 21.915," 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects, 1–118, 2019.
[20] 3GPP, "New SID on Support of Reduced Capability NR Devices," 3GPP Work Item Description, 1–4, 2019.
[21] UIT-T, "UIT-T Rec. Y.2060 General description of Internet of things," ITU Telecommunication Standardization Sector, 1–20, 2012.
[22] H. Fangming, "Wi-Fi 6 and 5G and Their Application Scenarios," Huawei Campus Network Marketing Support Module, 1–47, 2019.
[23] R. Hanno, R. George, K. Taylor, "Vital Signs: How to deliver better healthcare across Europe," Deloitte Centre for Health Solutions, 1–56, 2016.
[24] G. Horn, "5G End-to-End Data Management," Qualcomm Technologies, 2020.
[25] Huawei Technologies Co LTD, "5G Security: Forward Thinking," Huawei White Paper, 1–12, 2016.
[26] D. Basin, S. Radomirovic, J. Dreier, R. Sasse, L. Hirschi, V. Stettler, "A formal analysis of 5g authentication," Proceedings of the ACM Conference on Computer and Communications Security, 1383–1396, 2018, doi:10.1145/3243734.3243846.
[27] P. Teppo, K. Norrman, "Security in 5G RAN and core deployments," Ericsson White Paper, 6–13, 2020.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 52-61 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Advanced Physical Failure Analysis Techniques for Rescuing Damaged Samples with Cracks, Scratches, or Unevenness in Delayering

Yanlin Pan*, Pik Kee Tan, Siong Luong Ting, Chang Qing Chen, Hao Tan, Naiyun Xu, Krishnanunni Menon, Hnin Hnin Win Thoungh Ma, Kyaw Htin

GLOBALFOUNDRIES Singapore Pte. Ltd., QRA-EFA, Singapore, 738406, Singapore

ARTICLE INFO

Article history:
Received: 04 May, 2021
Accepted: 26 June, 2021
Online: 10 July, 2021

Keywords:
Delayering
PFA
Sample damage
Sample rescue

ABSTRACT

This paper is an extended version of work published in IPFA 2020. In the previous paper, advanced physical failure analysis (PFA) techniques for rescuing damaged samples with cracks, scratches, or unevenness in delayering are introduced. In the present work, the techniques will be further exploited and summarized for the potential applications in general devices. The three typical rescue cases will be fully discussed through comprehensive analysis on the failure mechanism and the rescuing process. Compared to the conventional PFA techniques that normally require back-up samples, the novel rescue techniques offer more alternative solutions for coping with sample damage problems in delayering without starting over with a new sample that would waste machine time and human resources. These new PFA techniques involve only basic failure analysis (FA) skills that could be easily manipulated and FA equipment that is commonly available in FA labs, and would extend the scope and capability of the tradition PFA to help the FA engineers deliver FA results with high quality and high success rate in the daily work, especially for handling "one of a kind" devices.

1. Introduction

Delayering by finger polishing is one of the commonly used physical failure analysis (PFA) techniques in the failure analysis (FA) labs. This technique is simple, direct, and flexible, which is basically mechanical polishing using fingers to press the sample against a rotating cloth plate with polishing slurry in between the cloth and the sample [1]. Before the polishing of each layer, reactive ion etch (RIE) is usually used to expose the metal structure by removing the inter-metal dielectric (IMD) for a faster removal rate and a more even surface of the polished sample. For monitoring the polishing progress, FA engineers will then use the scanning electron microscope (SEM) and the optical microscope (OM) to inspect the sample surface. A typical delayering workflow in PFA is shown in Figure 1. Normal process of the delayering completes when the target layer is reached, followed by the defect identification and analysis. However, if the sample comes with the accidental damage or the naturally generated edge rounding during the delayering, the problems have to be fixed to restore the sample to the former condition before the polishing is continued to the target layer.

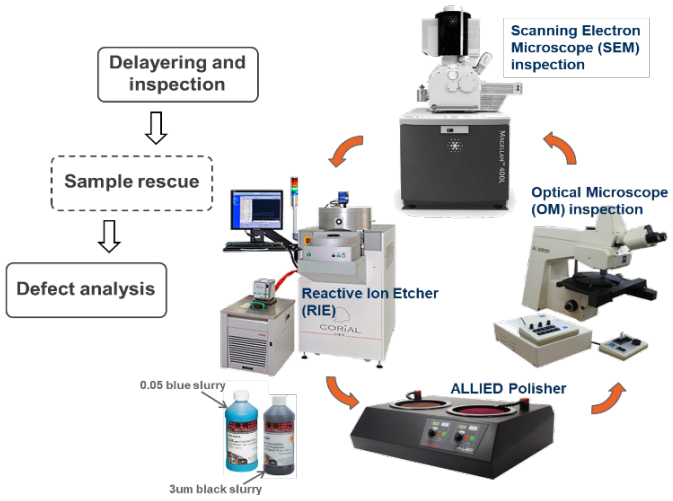


Figure 1: Illustration of a typical delayering workflow in PFA.

The common problems encountered in the delayering include sample cracks, polishing scratches, and surface unevenness [2-6]. The sample cracks refer to the sample die breaking into pieces due to chipping, dropping or crushing, which would pose great challenges to the following polishing if the crack lines extend

*Corresponding Author: Yanlin PAN, 60 Woodlands Industrial Park D Street 2, Singapore 738406. Phone: +65 9114 9926. Email: yanlin.pan@globalfoundries.com

https://dx.doi.org/10.25046/aj060407

across the defect area. The polishing scratches on the sample surface are usually caused by aggregated slurry particles mixed in the slurry solution or diamond lapping films when doing pre-polish sample preparation. The surface unevenness of the samples results from the edge effect which is a normal phenomenon in the delayering by both finger polishing and machine polishing [7-10]. For the small targets such as single memory bits or localized hotspot locations from electrical fault isolation (EFI), the unevenness will not lead to severe problem. But for the big inspection area or the defect location at the edge of the sample, the unevenness must be specifically treated to maintain the feasibility of the FA process. These problems could even become worse with the continuous development of the semiconductor devices, which introduces more layer stacks, smaller transistors, and softer low-k dielectrics. Moreover, accidents due to human error are considered inevitable and hard to control in the real FA work place. Hence, compared to the preventive practice for reducing the sample damage in delayering, finding rescue solutions to the problems is a more pressing need.

In the previous studies, sacrificial dummy as a polishing balance, platinum (Pt) as an etching mask, and lateral delayering with adapted polishing plate were used to counteract the edge effect [11-13]. Other studies focused on the impact of pad properties [14-15], cloth roughness [16], slurry particle size [17-18], and interaction mechanism [19-20] on the material removal rate and polishing uniformity. However, the unevenness problems have not been fully resolved because of the complexity of the issue. Unevenness is related to numerous factors such as force of pressing samples, speed of plate rotation, polishing angle, polishing orientation, slurry type and sample thickness, which are all dependent on the experience of the engineers. These variable parameters and different conditions of different devices would make it nearly impossible to operate in a standard procedure. In addition to surface unevenness, polishing scratches and sample cracks in the delayering are also dealt with in the daily FA work, but most of time there is no effective method of bringing the sample back to the normal condition. In order to fill the skill gaps, new techniques need to be developed.

In this paper, we will introduce three advanced PFA techniques for rescuing damaged samples with cracks, scratches, or unevenness in delayering by finger polishing, through their according typical FA cases. The first case uses the diamond lapping film and the sacrificial dummy samples to fix the cracked sample for continuing the FA process with no effect from the cracks. The second case uses e-beam Pt deposited in SEM machine to repair the scratch pits in the region of interest (ROI). With the polishing progresses to the lower layers, the scratch features gradually diminish until thoroughly removed from the sample surface. In the third case, an innovative technique combining controlled slurry polishing and partial RIE was developed for creating an ultra-large inspection area (> 50000 μm^2) with negligible unevenness. Compared to the conventional PFA techniques that normally require back-up samples, the novel rescue techniques offer more alternative solutions for coping with sample damage problems in delayering without starting over with a new sample that would waste machine time and human resources. These techniques would extend the scope and capability of the tradition PFA, and help the engineers deliver FA results with high success rate, especially for handling "one of a kind" devices.

2. Experiments

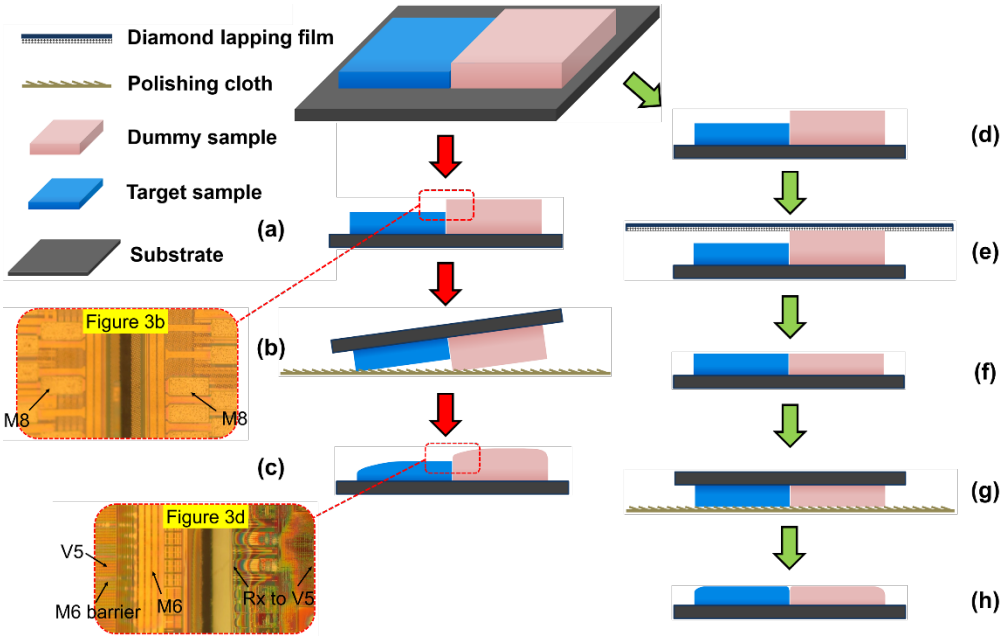
The experiments of the three cases were performed on a 40 nm node logic device, a 40nm node static random access memory (SRAM) device, and a 40 nm node "snake" metal line electrical test (ET) device, respectively. A mechanical polisher (ALLIED HighTech TwinPrep 5) with a rayon flock polishing cloth (ALLIED Spec-Cloth) and 3 μm polishing slurry (ALLIED water based diamond suspension) were used for slurry polishing. A 3 μm and a 15 μm diamond lapping films (ALLIED) were used for the selective polishing on the dummy samples. A SEM system (FEI Magellan™ 400L) equipped with a gas injection system was used for the topography inspection on the sample surface and the e-beam Pt deposition on the scratch pits. An optical microscope (ZEISS Axiotron) was used for the fast inspection during the mechanical polishing and the sample height determination. A RIE system (Corial 200 IL) was used to remove the IMD materials. TEM sample preparation was performed using a FIB-SEM dual-beam system (FEI Helios NanoLab 450S). TEM analysis was performed using a 200 kV Field Emission TEM (JEOL JEM-2100F).

3. Results and Discussion

3.1. Case 1: Using Diamond Lapping Film and Sacrificial Dummy to Save Cracked Sample with Target Defect Area Close to Gaps/Crack Lines

The first case is about a 40 nm node logic device which suffers scan failure with a diagnosed trace path across an area of 283 μm × 71 μm. The sample was delayered halfway (at M8) in the PFA process when it accidentally dropped broken into multiple pieces. Directly continuing the delayering on the piece that contains the defect area would come with the difficulty of polishing and the high risk of missing the defects that could locate anywhere along the whole trace path through each layer. Instead, the broken pieces were glued together with two dummy samples as polishing balance on a substrate using wax (Figure 5a). However, when the normal slurry polishing on the joined sample reached V5, obvious unevenness was observed near the crack line and the wax gap (Figure 5b). The mechanism of the unevenness generating during slurry polishing is illustrated in Figure 2a-2c. The height difference between the dummy sample and the target sample without treatment aggravated with the progress of delayering and caused surface unevenness of the sample.

Figure 3 shows a failed case of delayering where the dummy sample was originally higher than the target sample at M8, according to the focusing conditions in OM. After the sample was delayered from M8 to V5, severe unevenness had generated in the defect area and the dummy side was still higher than the sample side (Figure 3c, 3d), which means that the unevenness would further extend to the lower layers in the following polishing process. In the rescued case, we had to prevent the unevenness from extending to the defect area that would introduce the risk of missing the defect when the defect location or the layers were misjudged due to the unevenness. The unevenness in the defect area also posed great challenges to the layer-by-layer inspection for finding the defects, especially at the critical layers (V5 in this case) which are the interfaces between the low-k IMD layers (M5



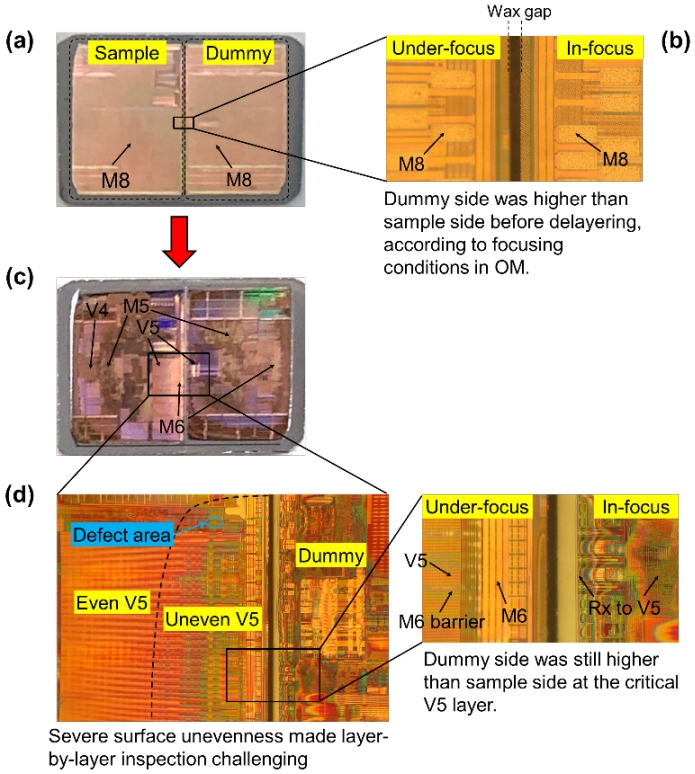


Figure 3: Failed case of delayering where (a) the dummy sample was originally higher than the target sample at M8, according to (b) the sample focusing condition in OM. When the target sample reached V5, (c, d) severe unevenness had generated in the defect area.

uniform surface at the critical layer, the unevenness would significantly deteriorate due to the great difference (up to 10 times) of the removal rate between the low-k materials and the non-low-k materials. Moreover, the sample inspection and the defect finding in SEM would be tediously prolonged because the defect area may cover multiple layers so that the staggered inspections on the different portions of the different layers in the trace path would have to be performed through repeated polish-and-view processes.

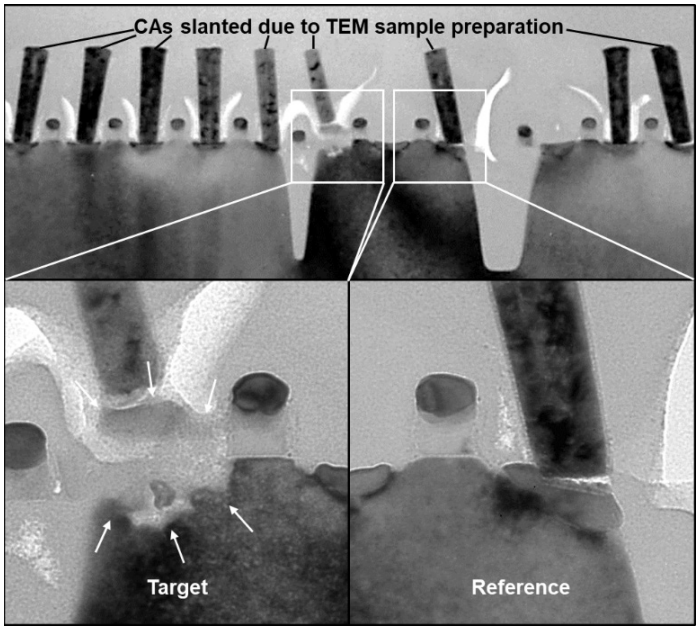


Figure 4: TEM (across PC) images of the target location and the reference location. Damaged Rx and PC-CA short were observed at the target location.

magnification used on the OM was 100× and the numerical aperture (NA) was 0.9. The measurement error regarding to Depth of Field (DOF) was ~ 0.22 μm [21], far smaller than the particle size ~ 3 μm of the polishing slurry. It is worth to mention that the chosen dummy sample should not be thinner than the target sample in the first place.

After removing the height difference between the target sample and the dummy sample, the unevenness was slowly eliminated in the subsequent slurry polishing within the V5 layer (Figure 5d). The sample was restored to its former evenness from V4 downwards, and the defect of damaged Rx and PC-CA short were found at the PC/CA layer (Figure 5e). The mechanism of the rescue process is illustrated in Figure 2d-2h. Figure 4 shows the TEM (across PC) images of the target defect location and the reference location, from which we can see that CA landing on the PC residue caused PC-CA short and the Rx below was damaged. It was suspected that PC residue blocked the PC etching at the defect location. In summary, to rescue the cracked samples with the defect area close to the gap/crack lines, dummy samples as polishing balance need to be used and the surface of the joined samples must be levelled first using diamond lapping films, before the slurry polishing is continued.

3.2. Case 2: Using E-beam Pt Deposition to Repair Scratched Sample with Scratch Pits in the ROI

The second case involved a 40nm node SRAM device. Electrical bench measurement on the failed unit showed source-drain leakage failure in an ET SRAM structure. EFI hotspot analysis was then performed to narrow down the ROI for the root cause finding in PFA. According to the detected defect location, the sample was delayered and viewed from the top layer. To avoid potential unevenness problem subject to the density difference between the target device region and the adjacent scribe lines, the thick top metal layers were removed using 3 μm diamond lapping film. Unfortunately, multiple scratch pits were accidentally created in the ROI during the lapping process (Figure 9a), which would introduce potential damage to the defect. The failure mechanism is illustrated in Figure 6a-6c. The scratch feature as a pit will extend from the scratched layer to the lower layers if the slurry polishing is directly continued without treatment.

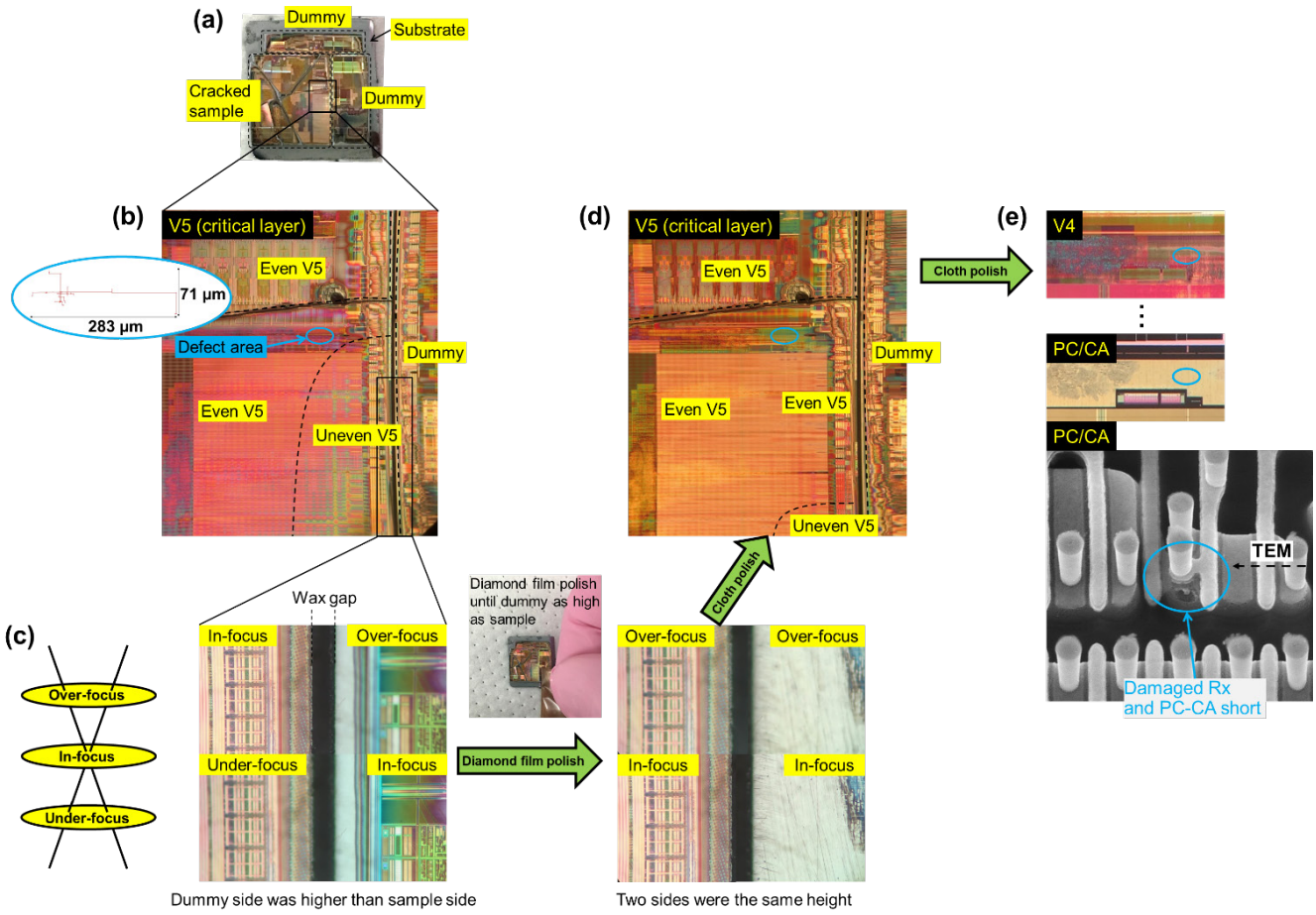


Figure 5: Successful case of sample rescue by using diamond lapping films and sacrificial dummy samples to save the cracked sample with the target defect area close to the crack lines. (a) The cracked samples (halfway at M8) were stuck onto a substrate together with two dummy samples as balance before the slurry polishing is continued. (b) When the joined sample reached V5, obvious surface unevenness was seen near the defect area, generated from the height difference between the dummy sample and the target sample. (c) By selective polishing on the higher dummy side using the diamond lapping films until it was the same height as the target sample, (d) the unevenness was slowly removed in the subsequent slurry polishing within the V5 layer. (e) The sample was restored to its former evenness from V4, and damaged Rx and PC-CA short were found at the PC/CA layer.

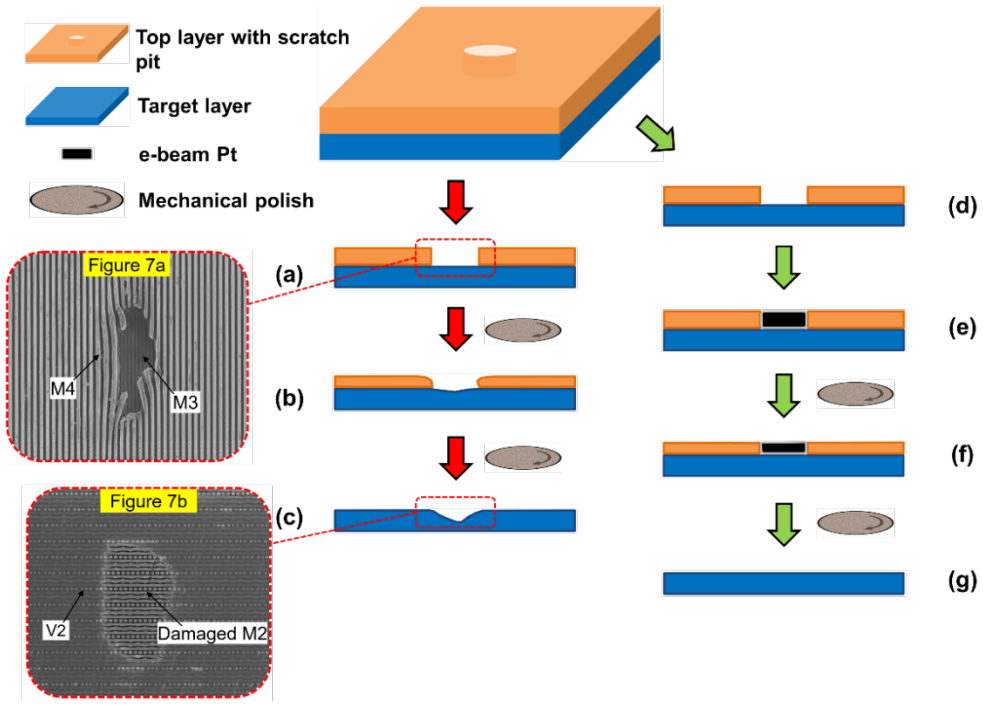


Figure 6: Schematic illustration of the failed case and the successful case of rescuing the scratched samples. (a-c) In the failed case, the scratch feature as a pit will extend to the lower layer if the slurry polishing is continued without treatment, which may damage the defect location. (d-g) In the successful case, by selectively filling the scratch pit with e-beam Pt, the defect location is protected from the damage of over-RIE and over-polish. Insets are the example SEM images from the failed case.

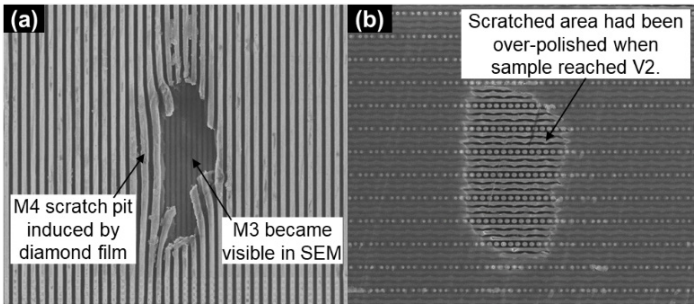
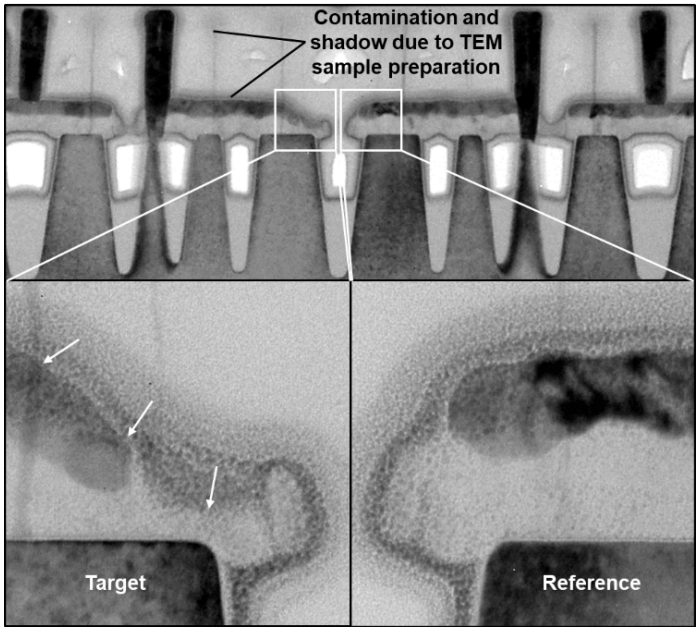


Figure 7: Failed case of delayering where (a) the sample was scratched at M4 by diamond lapping films, and (b) the scratched area had been over-polished when the normal slurry polishing continued and reached V2.

To solve the problem, we used e-beam Pt deposition to fill the scratch pits in SEM. The concept is using Pt metal patch as substitute for Cu to block the RIE and resist the slurry polishing on the next layer. The critical part of this method is the Pt thickness. Over-deposited Pt will leave Pt residue at the ROI resulting in surface unevenness in another way. Therefore, in the beginning we used high kV SEM to deposit thinner Pt film compared to the metal layer of which the thickness was determined by e-beam penetration depth in the SEM imaging. If the Pt were fading faster than the adjacent metal under slurry polishing, we would deposit extra layer of Pt on the ROI. For 40nm node device in this case, the Pt thickness was fixed to 0.1 μm. Voltage of 3 kV and current of 0.8 nA were used for SEM. The deposition rate is 10 μm (length) × 10 μm (width) × 10 μm (thickness) per 10 mins.



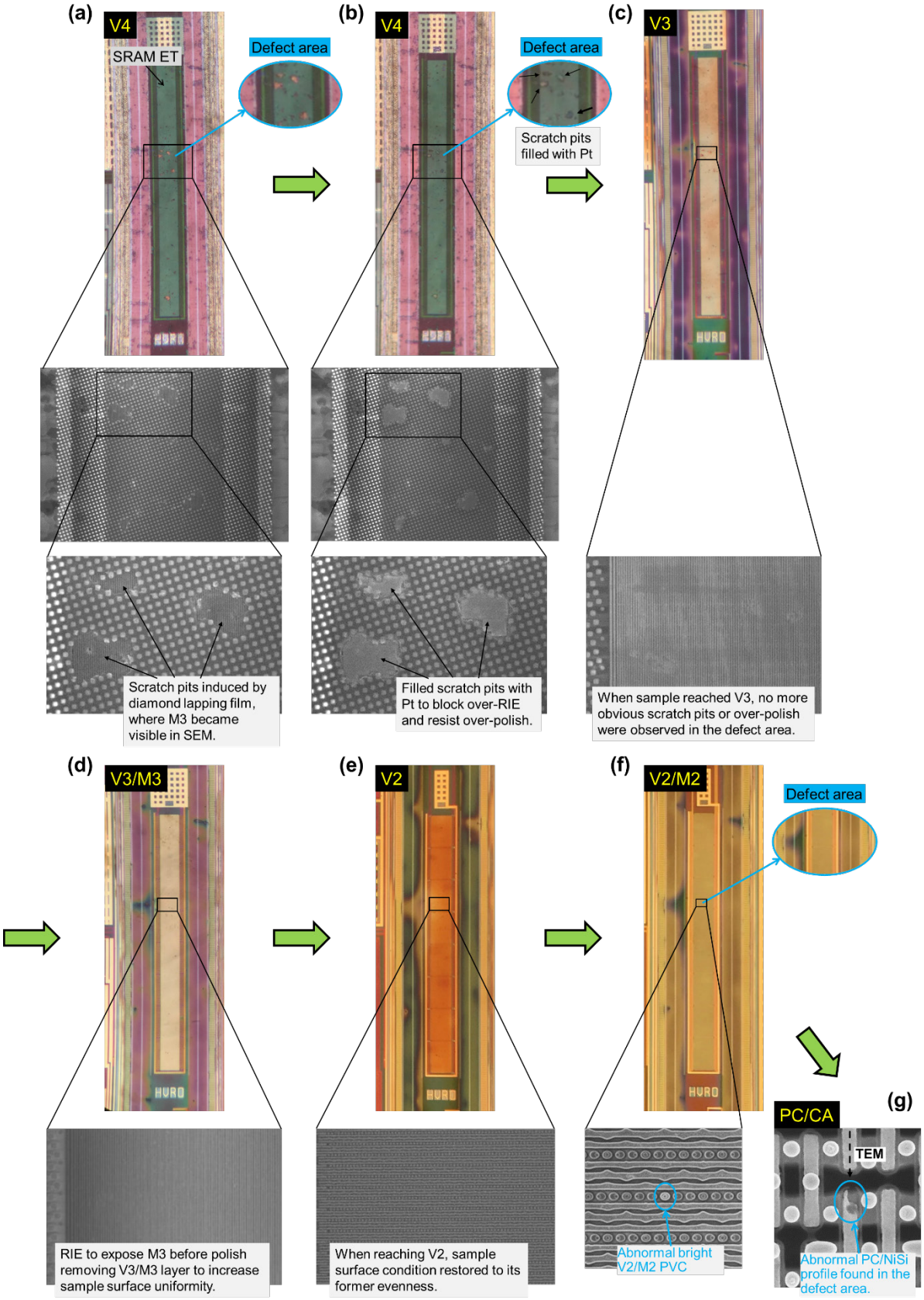


Figure 9. Successful case of sample rescue by using e-beam Pt deposition to repair scratched sample with scratch pits in the ROI. (a) The defect area of the sample got scratch damage when using diamond lapping films to remove top metal layers. (b) To block over-RIE and resist over-polish, e-beam Pt was deposited at the scratch pits to cover the damaged area. (c, d) After the normal RIE and slurry polishing removed V4/M4, the effect of the scratch damage on the V3/M3 layer showed greatly mitigated. (e) At V2 level, the sample had restored to an un-damaged condition, and (f) abnormal bright V2/M2 PVC was observed in the ROI. (g) The defect of abnormal PC/NiSi profile was finally found at the EFI spot location.

With protective Pt filling the scratch pits (Figure 9b), the ROI slowly restored to its former evenness and no more damage feature was seen when the sample reached V2 (Figure 9c-9e), as shown in the successful rescue case. After the M2 was exposed by RIE, optimal surface condition for SEM inspection was obtained and abnormal bright V2/M2 passive voltage contrast (PVC) was observed at the spot location (Figure 9f). The rescue mechanism is illustrated in Figure 6d-6g. The delayering on the sample was then continued from M2 downwards by normal delayering process and finally found the abnormal PC/NiSi at PC/CA level (Figure 9g). TEM images (along PC) are shown in Figure 8. Compared to the reference location, the defect location has abnormal PC/NiSi profile and missing portion of PC which are correlated with the source-drain leakage. PC formation issue in the wafer process is suspected as the root cause. In summary, this case demonstrated simple method of e-beam Pt deposition for tackling the common sample scratching problems induced by diamond film lapping or other incidents in delayering.

3.3. Case 3: Using Controlled Slurry Polishing Combined With Partial RIE to Remove Sample Unevenness and Create Ultra-Large (> 50000 μm^2) Inspection Area

The third case is about an ultra-large "snake" metal line ET structure in a 40 nm node device that failed M4 high resistance. To perform PFA on the failed ET, we tried to polish the sample to V4 and identify the physical defect that resulted in the high resistance. The area to be inspected is as big as > 50000 μm^2 where a single metal line "snake" runs through the whole ET structure in the wafer scribe line. PVC inspection in SEM is needed to locate the defect position which requires intact connection in the metal line without any damage such as metal bridging or metal broken from the sample preparation. However, when the normal slurry polishing on the sample reached V5, obvious evenness at the ET corners was observed, which would potentially lead to metal smear or break if we continued to polish until V4.

A failed case is shown in Figure 11. The "snake" metal line ET sample had uneven IMD at V5 due to the edge effect, while slurry polish directly proceeded upon full time RIE removing IMD. After M5 was removed, severe M4 unevenness and damage were generated. At the corner areas, the sample was even over-polished to M3. The failure mechanism is shown in Figure 10a-10c, where the IMD unevenness is consider as the root cause. Without intact metal line connection, PVC inspection was not possible. Even if the defect was not damaged, tedious inspection in SEM would be needed to check every piece of metal through the whole large area for the defect.

Figure 10d-10h illustrates how we coped with the issue in the successful case. By smearing the metal lines to form an etch-stop area after the first partial RIE (Figure 10f), the IMD was evened

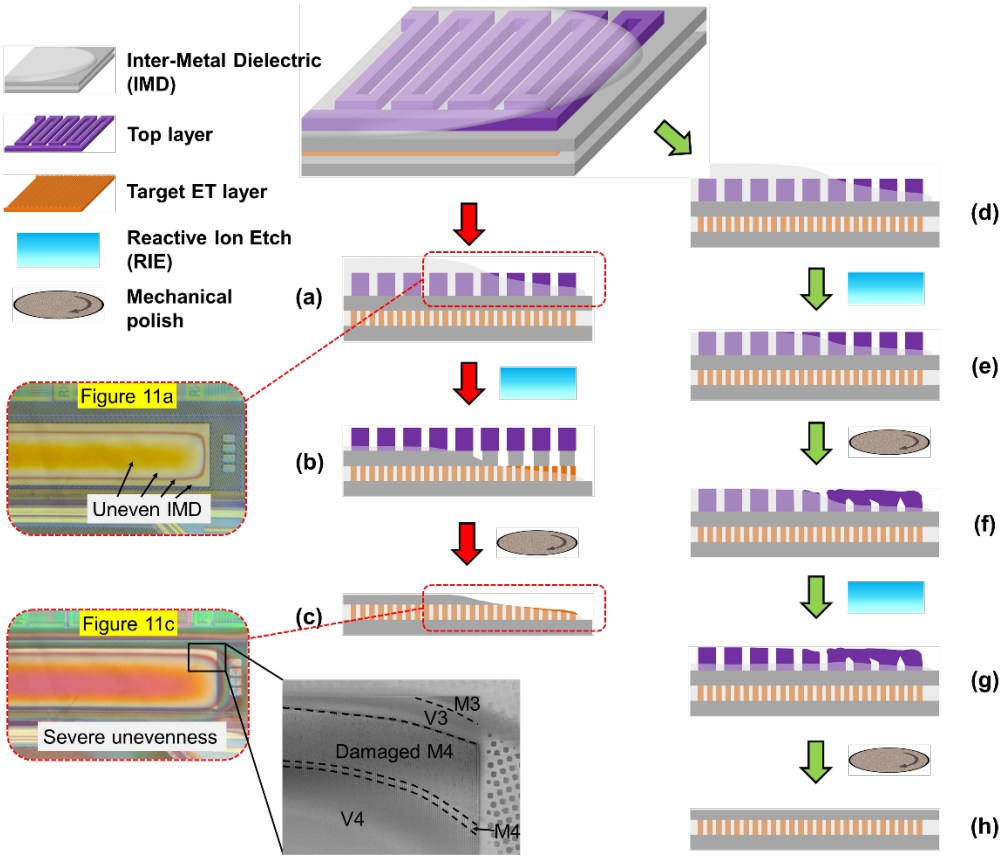


Figure 10: Schematic illustration of the failed case and the successful case of rescuing the uneven samples. (a-c) In the failed case, the edge effect induced unevenness will aggravate along with the slurry polishing, especially for the structure in the scribe lines. (d-g) In the successful case, by smearing the metal lines to form an etch-stop area after the first partial RIE, the unevenness of the IMD will be removed, which will result in an even surface in the next layer. Insets are the OM and SEM images from the failed case.

out by another partial RIE (Figure 10g) to obtain an even surface in the next layer (Figure 10h). The SEM and OM results of the successful rescue case are shown in Figure 13. When the uneven IMD was spotted as circular halo at the ET edge, partial RIE using half of the full RIE time was performed to expose the outer M5 with thinner IMD (Figure 13a). The different division of the RIE time is dependent on different sample conditions including via length and degree of unevenness. More divisions deliver better results in flattening the IMD. In this case, the full RIE time was equally divided into two halves (e.g., if the full time is 20 s, each half will be 10 s). The first partial RIE shrank the IMD halo by around 50% (Figure 13b), which means half of the M5 was exposed. Then we slightly slurry polished the surface until the exposed M5 metal lines collapsed and smeared together. The halo area further shrank smaller because the V5 was polished thinner or nicely fully removed. As a result, three different regions (smeared M5, exposed M5, and un-exposed M5) were generated (Figure 13c). Since the RIE for oxide etching doesn't chemically react with metal structure, the smeared M5 worked as a mask to block the RIE impact on the IMD4. After the second partial RIE was applied onto the sample, the IMD5 in the areas of exposed M5 and un-exposed M5 was etched down until the same level as that in the smeared M5 area (Figure 13d). The cross-sectional view is illustrated in Figure 10g.

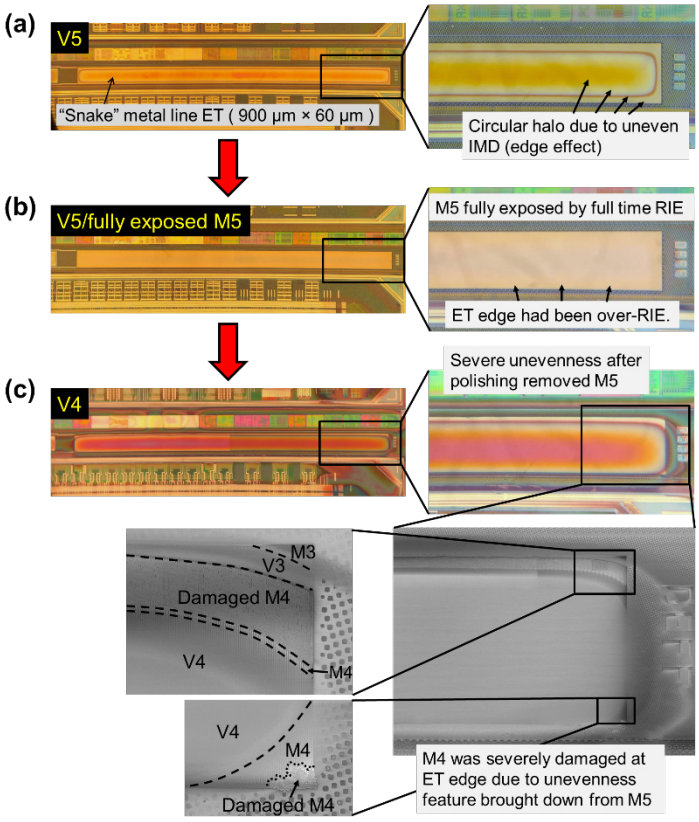


Figure 11: Failed case of delayering ultra-large "snake" metal line ET where (a) the uneven V5/M5 was performed with (b) full time RIE which induced over-exposure of M5 and in turn (c) over-polish on M4.

Finally, after removing M5 by normal slurry polishing, the sample achieved great evenness at V4 across the whole ET structure without any damage to M4. PVC inspection in SEM was then successfully performed and the bright signal was traced along the single ET metal line to quickly locate the defect site where the PVC stopped (Figure 13e). TEM (across the metal lines) analysis on the defect location showed abnormal M4 pattern which was responsible for the high resistance failure, and lithography issue in the wafer process was suspected (Figure 12). In summary, slurry polishing combined with partial RIE is able to generally remove the unevenness induced by edge effect within large areas.

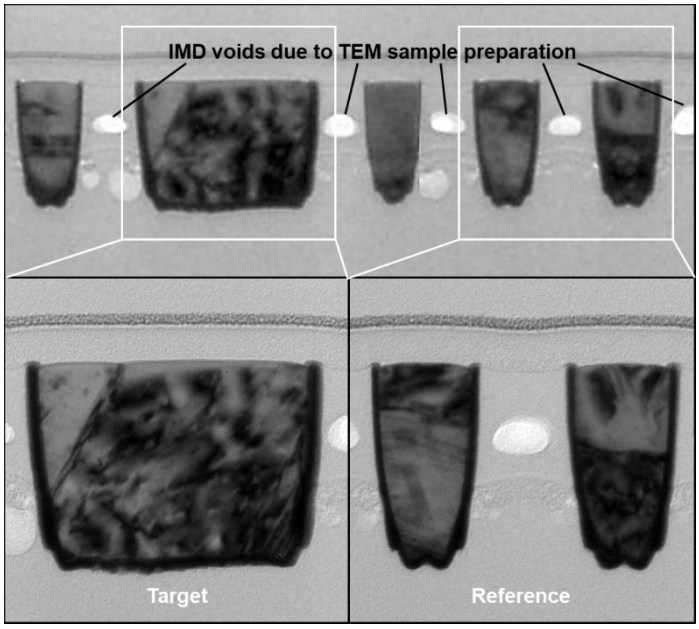


Figure 12: TEM (across metal line) images of the target location and the reference location. Abnormal M4 pattern was observed at the target location.

Table 1: Summary of PFA techniques for rescuing damaged samples with cracks, scratches, or unevenness in delayering.

Types	Problems	Solutions
Cracks	Polishing difficulty, Edge effect	Balance polishing using dummy; Level the joined samples using diamond lapping films
Scratches	Over-RIE, Over-polishing	Deposit Pt on the scratched area to block RIE and resist polishing
Unevenness	Edge effect, Limited inspection area	Smear the partial RIE exposed metal, and then even out the IMD by another partial RIE

In this paper, advanced PFA techniques for rescuing damaged samples with cracks, scratches, or unevenness in delayering by finger polishing have been discussed, through three typical FA cases. The first case is on rescuing cracked samples where diamond lapping film and sacrificial dummy were used to cope with the polishing difficulty and the edge effect with the target defect area close to the gaps/crack lines. The second case is about using e-beam Pt deposition to repair the damage in ROI to the scratched samples and restore the sample to its even condition. The third case studied using controlled slurry polishing combined with partial RIE to remove sample unevenness and create ultra-large (> 50000 μm^2) inspection areas. These techniques are very useful in helping the FA engineers to tackle accidents, solve problems and deliver high-quality FA results, especially for handling "one of a kind" devices.

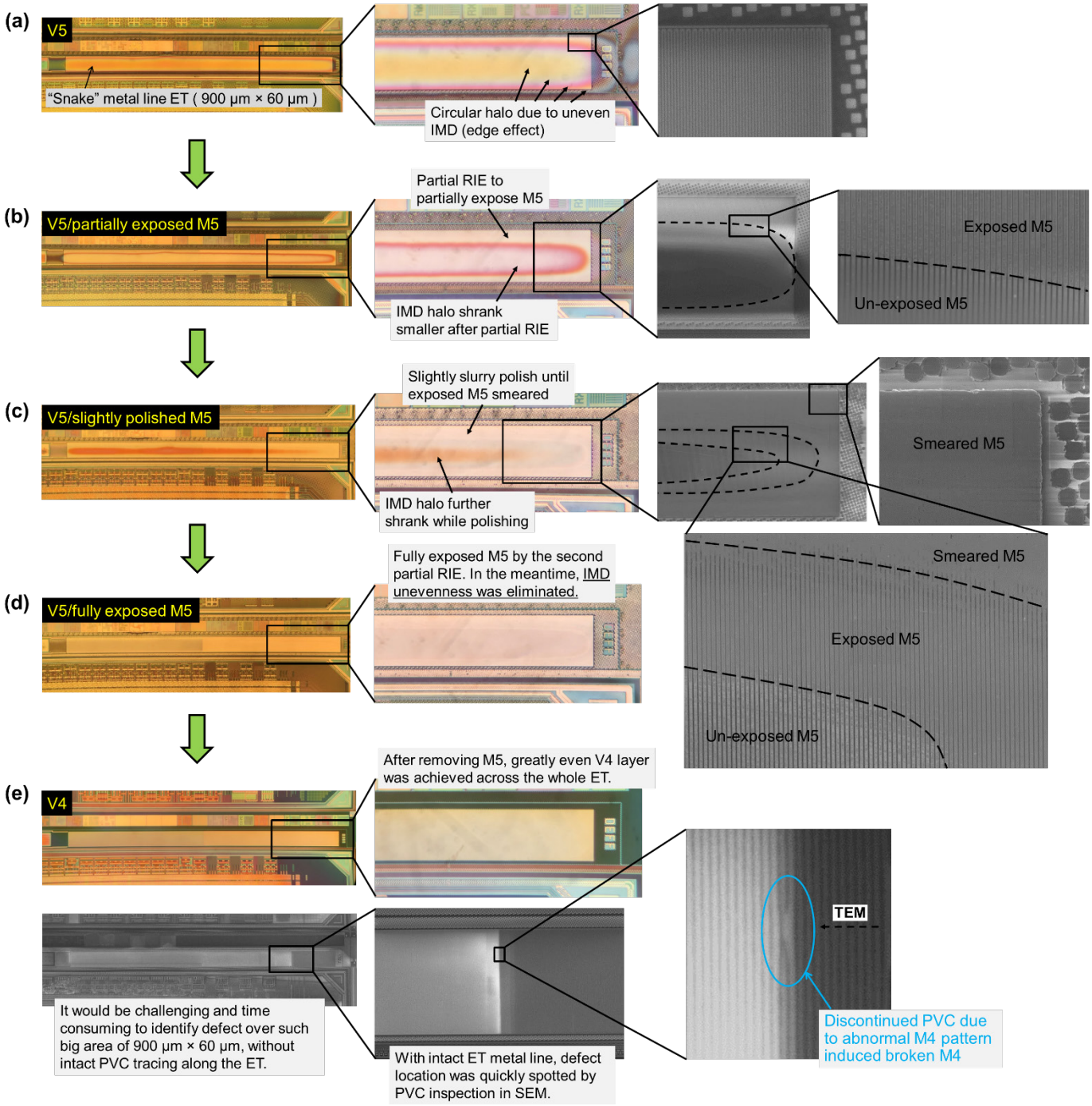


Figure 13: Successful case of sample rescue by using controlled slurry polishing combined with partial RIE to remove sample unevenness and create ultra-large (> 50000 µm^2) inspection area in the "snake" metal line ET. (a) A circular halo due to the uneven IMD at the ET edge was observed at V5. (b) M5 was partially exposed by the first partial RIE, and the sample was (b) slightly slurry polished until (c) the exposed M5 was smeared. (d) The second partial RIE was then performed to get rid of the IMD unevenness. (e) After M5 was removed by slurry polishing, the sample achieved great evenness globally through the whole ET structure. With an intact ET metal line, PVC inspection easily located the defect of broken M4.

Conflict of Interest

The authors declare no conflict of interest.

Abbreviations

DOF	Depth of Field
EFI	Electrical Fault Isolation
ET	Electrical Test
FA	Failure Analysis
FIB	Focused Ion Beam
IMD	Inter-Metal Dielectric
Mx	Metalx, x is layer number
NA	Numerical Aperture
OM	Optical Microscope
PFA	Physical Failure Analysis
RIE	Reactive Ion Etch

ROI	Region of Interest
SEM	Scanning Electron Microscope
SRAM	Static Random Access Memory
TEM	Transmission Electron Microscope
Vx	Viax, x is layer number

References

[1] J. Leo, H. Tan, Y. Z. Ma, S. M. Parab, Y. M. Huang, D. D. Wang, L. Zhu, J. Lam, Z. H. Mai, "Key Issues for Implementing Smart Polishing in Semiconductor Failure Analysis," Journal of Applied Mathematics and Physics, **5**, 1668-1677, 2017, doi:10.4236/jamp.2017.59139.

[2] M. E. Weldy, L. Serrano, "Increasing Planarity for Failure Analysis Using Blocked Reactive Ion Etching Combined With Planar Polish," in 2005 International Symposium for Testing and Failure Analysis (ISTFA), 206-208, 2005, doi:10.31399/asm.cp.istfa2005p0206.

[3] H. Feng, P. K. Tan, H. H. Yap, G. R. Low, R. He, Y. Z. Zhao, B. Liu, M. K. Dawood, J. Zhu, Y. M. Huang, D. D. Wang, H. Tan, J. Lam, Z. H. Mai, "A Sample Preparation Methodology to Reduce Sample Edge Unevenness and Improve Efficiency in Delayering the 20-nm Node IC Chip," in 2015 IEEE International Symposium on Physical and Failure Analysis of Integrated Circuits (IPFA), 2015, doi:10.1109/IPFA.2015.7224432.

[4] K. S. Wills, S. Perungulam, "Delayering Techniques: Dry Processes, Wet Chemical Processing, and Parallel Lapping," Microelectronics Failure Analysis Desk Reference Fifth Edition, 444-463, 2004.

[5] Y. M. Sub, B. T. H. Yap, F. I. Lee, A. B. Minhar, K. W. Tan, H. J. Looi, T. M. Foo, "A Case of Charging Induced Damage into the Common Metal Interconnect During Chemical Mechanical Polishing," in 2017 IEEE International Conference on Opto-Electronic Information Processing (ICOIP), 2017, doi:10.1109/OPTIP.2017.8030704.

[6] K. S. Wills, "Planar Deprocessing of Advanced VLSI Devices," in 2006 International Symposium for Testing and Failure Analysis (ISTFA), 393-397, 2006, doi:10.31399/asm.cp.istfa2006p0393.

[7] S. Bott, R. Rzehak, B. Vasilev, P. Kucher, J. W. Bartha, "A CMP Model Including Global Distribution of Pressure," in 2011 IEEE Transaction on Semiconductor Manufacturing, 304-314, 2011, doi:10.1109/TSM.2011.2107532.

[8] Y. Y. Meng, L. Zhang, Y. B. Li, W. Zhang, H. F. Zhou, J. X. Fang, "Impact of Bevel Condition on STI CMP Scratch," in 2020 China Semiconductor Technology International Conference (SCTIC), 2020, doi:10.1109/CSTIC49141.2020.9282450.

[9] X. Wu, Z. Huang, Y. J. Wan, H. T. Liu, X. Chen, "A Novel Force-Controlled Spherical Polishing Tool Combined With Self-Rotation and Co-Rotation Motion," in 2020 IEEE Access, **8**, 108191-108200, 2020, doi:10.1109/access.2020.2997968.

[10] A. Wieters, P. Thieme, "Wafer Edge / Bevel Treatment of Device Wafers by Means of CMP," in IEEE 2007 International Conference on Planarization / CMP Technology (ICPT), 2007.

[11] T. Moor, E. Malyanker, E. R. Moyal, "Single Die 'Hand-Free' Layer-by-Layer Mechanical Deprocessing for Failure Analysis or Reverse Engineering," in 2008 International Symposium for Testing and Failure Analysis (ISTFA), 2008, doi:10.31399/asm.cp.istfa2008p0363.

[12] H. H. Yap, P. K. Tan, G. R. Low, M. K. Dawood, H. Feng, Y. Z. Zhao, R. He, H. Tan, J. Zhu, B. H. Liu, Y. M. Huang, D. D. Wang, J. Lam, Z. H. Mai, "Top-down Delayering to Expose Large Inspection Area on Die Side-edge with Platinum (Pt) Deposition Technique," Microelectronics Reliability, 55, 1611-1616, 2006, doi:10.1016/j.microrel.2015.06.037.

[13] H. G. Ong, C. L. Gan, "Alternative Lapping Method to Reduce Edge Rounding Effect," in 2012 International Symposium for Testing and Failure Analysis (ISTFA), 462-464, 2012.

[14] K. P. Park, H. S. Kim, O. Chang, H. S. Jeong, "Effects of Pad Properties on Material Removal in Chemical Mechanical Polishing," Journal of Materials Processing Technology, **187-188**, 73-76, 2007, doi:10.1016/j.jmatprotec.2006.11.216.

[15] L. X. Wu, C. F. Yan, "Effects of Polishing Parameters on Evolution of Different Wafer Patterns During Cu CMP," in 2015 IEEE Transaction on Semiconductor Manufacturing, 106-116, 2015, doi:10.1109/TSM.2014.2387211.

[16] D. Lim, H. Kim, B. Jang, H. Cho, J. Kim, H. Hwang, "A Novel Pad Conditioner and Pad Roughness Effects on Tungsten CMP," in IEEE 2014 International Conference on Planarization / CMP Technology (ICPT), 352-355, 2014, doi:10.1109/ICPT.2014.7017318.

[17] C. Wang, P. Sherman, A. Chandra, D. Dornfeld, "Pad Surface Roughness and Slurry Particle Size Distribution Effects on Material Removal Rate in Chemical Mechanical Planarization," CIRP Annals-Manufacturing Technology, **54**, 309-312, 2005, doi:10.1016/S0007-8506(07)60110-3.

[18] T. F. Zeng, T. Sun, "Size Effect of Nanoparticles in Chemical Mechanical Polishing - A Transient Model," in 2005 IEEE Transaction on Semiconductor Manufacturing, 655-663, 2005, doi:10.1109/TSM.2005.858508.

[19] C. J. Evans, E. Paul, D. Dornfeld, D. A. Lucca, G. Byrne, M. Tricard, F. Klocke, O. Dambon, B. A. Mullany, "Material Removal Mechanisms in Lapping and Polishing," CIRP Annals, **52**, 611-633, 2003, doi:10.1016/S0007-8506(07)60207-8.

[20] L. Shan, C. H. Zhou, S. Danyluk, "Mechanical Interactions and Their Effects on Chemical Mechanical Polishing," in 2001 IEEE Transaction on Semiconductor Manufacturing, 207-213, 2001, doi:10.1109/66.939815.

[21] "Microscopy Basics: Depth of Field and Depth of Focus", Nikon, https://www.microscopyu.com/microscopy-basics/depth-of-field-and-depth-of-focus.

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 62-71 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Optimized Component based Selection using LSTM Model by Integrating Hybrid MVO-PSO Soft Computing Technique

Anjali Banga*, Pradeep Kumar Bhatia

Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology, Hisar-Haryana, 125001, India

ARTICLE INFO

Article history:
Received: 09 April, 2021
Accepted: 17 June, 2021
Online: 10 July, 2021

Keywords:
CBSE
LSTM
Deep learning
Neural Network

ABSTRACT

Research focused on training and testing of dataset after Optimizing Software Component with the help of deep neural network mechanism. Optimized components are selected for training and testing to improve the accuracy at the time of software selection. Selected components are required to be attuned and accommodating as per requirement. Soft computing mechanism such as PSO and MVO will be used for optimization. Deep Neural-Network mechanism is performing training and testing to get the confusion metrics of true positive/negative and false positive/negative. The accuracy, precision, recall value and f-score are computed to assure accuracy of proposed work. The proposed mechanism is making use of LSTM layer for more accurate output. Proposed research is exploring inadequacy of existing research and extent of incorporation of previous mechanism to soft computing mechanism in CBSE.

1. Introduction

Research is considering the dataset of the CBSE model [1] where the dataset presenting software component selection is trained and during testing of the trained network the confusion matrix is produced. According to the confusion matrix the accuracy, F-score, Recall, and precision values are found. On the other hand, data set of each grade would be passed to a hybrid MVO-PSO optimizer to find an optimized rating for each grade to filter the dataset. It could be said that the optimized value for each result is kept to find accuracy, F-score, Recall, and precision values accordingly. Finally, the comparison of accuracy, F-score, recall as well as precision values for non-optimized dataset to optimized Precision, F-score, and Recall for optimized dataset would be performed.

1.1. CBSE

The method of creating various software projects in different categories has been examined in Software engineering. Computer engineering applications have often been used to achieve this goal. However, the dependability of the software system is a difficult job to predict. CBSE may be regarded as a software reliability mechanism that is able to handle the issue. The broad method that supports the creation of various components depending on current software research was evaluated in terms of component-based software engineering. The newest software is not a simple task for beginners, but CBSE [2] allows developers to minimize efforts during the creation of software. Different influence variables are important in the case of CBSE, such as reusability, reliability, component dependence, and interactions between components. These characteristics promote the development of new software and reduce system complexity. There were many software computing methods that tried to predict software dependability. There are several observations. During software development, the selection of component steps has been discovered. Component-dependent design patterns of software are utilized for the recovery and assembly of components.

1.2. Reinforcement Learning

Reinforcement Learning is determined as Machine Learning. It is a branch of AI. Reinforcement Learning has been considered as a category of Machine Learning. It is also considered a branch of AI. Exactly permit a representative of hardware and computer program to mechanically identify perfect attitude in particular circumstances, for maximizing its efficiency. Exactly a type of Machine Learning method which permits representative of hardware and computer program to mechanically identified perfect attitude within a particular circumstance, for maximizing its efficiency. Reinforcement learning is utilized in various graphical games and uses a multi-agent mechanism to control the approach to environment exploration. It is also integrated with the company of abstraction methods which have the various levels to form powerful games dependent on artificial intelligence.

*Corresponding Author: Anjali Banga, Email: banga.anjali88@gmail.com

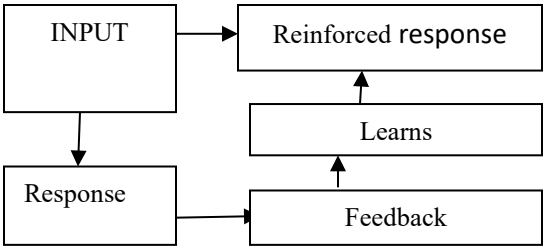


Figure 1: Reinforcement learning

Agent: The performer in the learning system is an agent. All actions are performed by the agent in the environment. The agent gets a reward as per action.

State: The state represents the current status of the environment that plays a significant role in assigning a reward to agents.

Environment: All action of agents is performed in the environment and rewards are providing to the agent as per the state of the environment.

Action: It is a method of representative due to which it communicates and exchanges its setting, in the middle of states. Whenever an action is executed by representatives it gives output inform of reward-dependent on the setting.

Reward: A reward in RL is part of the feedback from the environment. When an agent interacts with the environment, he could observe changes in state and reward signals through his actions.

1.3. Recurrent Neural Network

RNN network has been considered as a class of neural networks where interconnectivity among nodes is forming a directed graph. It is also creating a temporal sequence which is allowing it to show temporal dynamic behavior. RNN is capable to utilize their memory to compute sequences of inputs of different size as these are inherited from neural networks that are based on feed-forward. It is making them unable to implement operations like un segmented as well as interconnected consideration. RNN has been utilized to take into account two different broad categories of networks that are supporting the usual structure. Here one is having finite impulse. But another one is having impulse which is infinite. Such categories of networks have exhibited runtime actions that are not permanent.

1.3.1. Long Short-Term Memory

LSTM is a well-recognized artificial RNN. This is often used in the field of profound education. Feedback connectivity is provided by LSTM. It is not like a neural network feed. Not only are single data points like graphs processed. Sequences of information such as audio and video are also completed. In the case of LSTM networks, categorization is deemed appropriate. It does process and predicts based on information from time series. This is because there may be temporal delays not known throughout time series in important occurrences.

1.3.2. RNN and LSTM

An ongoing neural network is also known as an RNN. The category of ANN is examined. The node connections provide a network guided by a graph. You accomplished this with the sequence of time. Standard recurring neural networks have disappearances and explosions. LSTM networks were regarded as RNN types. Besides conventional units, LSTM is supported by special units.

1.3.3. Resolving Overfitting Problem by Dropout Layer

The dropout layer is playing a significant role in resolving the issue of over-fitting. The issue of over-fitting arises during the training of the neural network model. The dropout layer is used to handle such issues. If the training is continued, then the model adopts idiosyncrasies. Sometime training becomes less suitable for data that is new to it. This data could be different samples from the population. The model is considered to over-fit when it is too well-adapted to training as well as validating data.

Over-fitting is traced during plotting by checking the validation loss. The model is over-fitting when training loss is constant or it is decreasing. Techniques known as regularizes are used to minimize the influence of over-fitting. Dropout has been considered one out of them.

Dropout is working by eliminating or dropping out the inputs to layer. These could be input variables in a sample of data that is the output of the previous layer. In other words, the Dropout layer is attached to the model among previous layers. It applies to the results of the last layer that have been fed to the next layer. This is influenced by the simulation of a huge network with various network structures. Dropout rate could be considered to layer as chances of configuring every input to layer.

2. Literature Review

There are several types of research in the field of component-based software systems, optimization mechanisms, and neural networks. Researchers have used the SVM as a Classification Method for the Prediction of a defect in software with code metrics. Moreover, research for Performance Modeling of Interaction protocol for CBSD using OOPS based simulation came into existence. After some time, research related to software components selection optimization for CBSE development was made. Researchers applied particle swarm optimization for the performance prediction of the software components. Building models for optimized CBSE was built in several applications development. Some researchers did reliability estimation and performed prediction and measurement of CBSE. A Genetic Algorithm was proposed to manage SVM for predicting components that might be fault-prone. S. Di Martino proposed genetic algorithm. The objective of their research was to set the SVM in order to forecast components that are fault-prone. M. Palviainen did research to estimate reliability. They predicted and measured of component-based software.

In [3], the author applied PSO to software performance prediction.

In [4], the author proposed optimization model. This model has been developed for selection of software component. It has been used in development of several applications.

In [5], the author used SVM based classifier approach for reusability of software components.

In [6], the author performed multi-objective optimization. They did optimization of software architectures. Authors have used Ant Colony Optimization mechanism to accomplish their objective.

In [7], the author did estimation of software reusability in case of component-based system. They made used of several soft computing techniques in their research.

In [8], the author proposed adaptive Neuro fuzzy model. The objective of their research was to predict the reliability in case of component-based software systems.

In [9], the author introduced research on Test case prioritization. This research considered prioritization to perform regression testing. Research made use of ant colony optimization.

In [10], the author proposed research on Neuro-Fuzzy Model in order to find & optimize the Quality as well as Performance in case of CBSE.

In [11], the author presented dynamic mechanism in order to get software components with support of genetic algorithm.

In [12], the author proposed LSTM-based Deep Learning Models. The objective of research was to perform Non-factoid Answer Selection.

In [13], the author did research on multi-Verse Optimizer. They considered it as nature-inspired mechanism in case of global optimization.

In [14], the author proposed research to detect inconsistency in software component. Author made use of ACO and neural network mechanism.

In [15], the author presented deep learning mechanism in case of short-term traffic forecast. The research considered LSTM network.

In [16], the author proposed multiple target deep learning in case of LSTM.

In [17], the author proposed Preference-based component identification by making use of PSO.

In [18], the author proposed research on deep learning in order to perform solar power forecasting. Their approach made use of AutoEncoder along with LSTM Neural Networks.

In [19], the author presented quality assurance by soft computing mechanism. The research focused in field of component-based software.

In [20], the author did quality prediction by making use of ANN. Their research was based on Teaching-Learning Optimization.

In [21], the author proposed multi-objective model for optimization.

In [22], the author did Component selection considering attributes.

In [23], the author did software reliability prediction by making use of Bio Inspired approach.

In [24], the author proposed identification and selection of software component.

In [25], the author proposed model in order to predicting CBS reliability by making used of soft computing mechanism.

In [26], the author proposed enhanced Ant lion Optimizer along with Artificial Neural Network. This research focused on predicting Chinese Influenza.

In [27], the author Imoize considered software reuse and metrics in software engineering.

In [28], the author focused on improvement of reusability of component-based software. Research considered the advantages of software component by making use of data mining.

In [29], the author introduced hybrid Neuro-fuzzy as well as model for feature reduction in order to perform classification.

Table 1: Literature review

Author/ Year	Objective of research	Methodology	Limitation
U. Sharma/2012	Proposing reusability of software component	SVM	Slow mechanism has been proposed.
D. Gao/2015	Implementing test case prioritization in case of regression testing	ACO	Research has limited scope
G. Kumar/2015	Estimating and optimizing quality as well as performance in case of CBSE	Neuro fuzzy model	Need to improve the training efficiency
S. Vodithala/2015	Proposing dynamic mechanism to perform retrieval of software component	Genetic algorithm	Work is suffering from limitation of genetic algorithm
Ashu/ 2016	To build efficient IDS	LSTM algorithm	Research is not making using of optimizer to increase performance
O. Bhardwaj/2018	Assuring Quality by soft computing approach in component based software	CBSC	Research has not considered intelligent approach for accurate prediction.
P. Tomar/2018	To forecast prediction of quality by making use of ANN mechanism in case of Teaching-Learning Optimization for component-based software systems	ANN	The research need to do more work on performance and accuracy.
L. Mu/2018	Peforming the multi-objective optimization model of component selection	Multi-objective optimization	The optimization of multiple objectives is challenging and complex operation.
S. Gholamshahi/2019	Performing the preference based identification of component by making use of optimization technique.	PSO	There are several optimization mechanisms such as MVO that could perform better than PSO.
C. Diwaker/2019	Proposing prediction Model for CBSC for Reliability with support of Soft Computing	CBSC	Prediction model need to be more accurate and reliable.
HongpingHu/2019	Proposing improved ALO and ANN for Chinese Influenza Prediction	Ant lion optimization and artificial neural network	The performance of such system is slow there is need to filter the dataset
A. L. Imoize/2019	Reviewing the Software Reuse as well as Metrics in case of software Engineering	Software metrics	The research has limited scope due to lack of technical work

G. Maheswari/2019	To improve the reusability and Measuring Performance Merits in case of CBSE in case of Data Mining	Data mining	Research considered reusability and performance metric but there is need to consider optimization mechanism.
Himansu Das/2020	Proposing Hybrid Neuro-Fuzzy as well as model for feature reduction during classification	Hybrid Neuro-fuzzy	Need to introduce optimization mechanism to increase accuracy during feature reduction.

3. Problem Statement

Many studies shown the selection of software components, but the optimization mechanism is required for better performance. PSO was used to optimize results in previous studies. However, it is found that MVO offers better performance. In addition, an intelligent model should be introduced that may allow for a deep learning approach using RNN based on LSTM. However, it takes lot of time during training and testing. Then it finds accuracy, f-score according to confusion matrix. The optimization has been included to neural network model for better performance. Through the incorporation of the LSTM MVO-PSO hybrid method, the proposed study is needed to address the performance and accuracy problem.

The proposed work is answer for the accuracy and performance issues faced in previous researches.

4. Proposed Work

In the proposed work, the dataset of the CBSE model is considered. This dataset is trained using a neural network mechanism. During testing of the trained network, the confusion matrix is produced. On the basis of this confusion matrix the accuracy, F-score, Recall, and precision values are calculated. On other hand, the dataset is classified grade-wise in order to get optimized value for each grade. The data set of each grade would be passed to a hybrid MVO-PSO optimizer in order to get the optimized rating for each grade. The data of each grade would be filtered on basis of these optimized values. In another word, the data above the optimized value for each result would be kept. Then the accuracy, F-score, Recall, and precision values are calculated considering this filtered dataset. Then the comparison of accuracy, F-score, Recall, and precision values for non-optimized datasets are made to optimize, F-score, Recall, and precision values for the optimized dataset.

The figure illustrates the proposed system is required to train a component-based selection system that should be capable to predict with maximum accuracy. The trained model for component-based selection would be made with the support of LSTM and simulated in a Matlab environment. Existing researches in component-based selection have provided limited accuracy with limited precision, f-score, and recall value. The implementation of such a model is quite challenging but such research opens doors for innovations. There are several existing types of research that have contributed to the field of component-based selection. It has been observed that previous researches have made use of Fuzzy logic, Genetic algorithm, Machine learning mechanism, KNN classification, LSTM model. But these researches are suffering from accuracy issues.

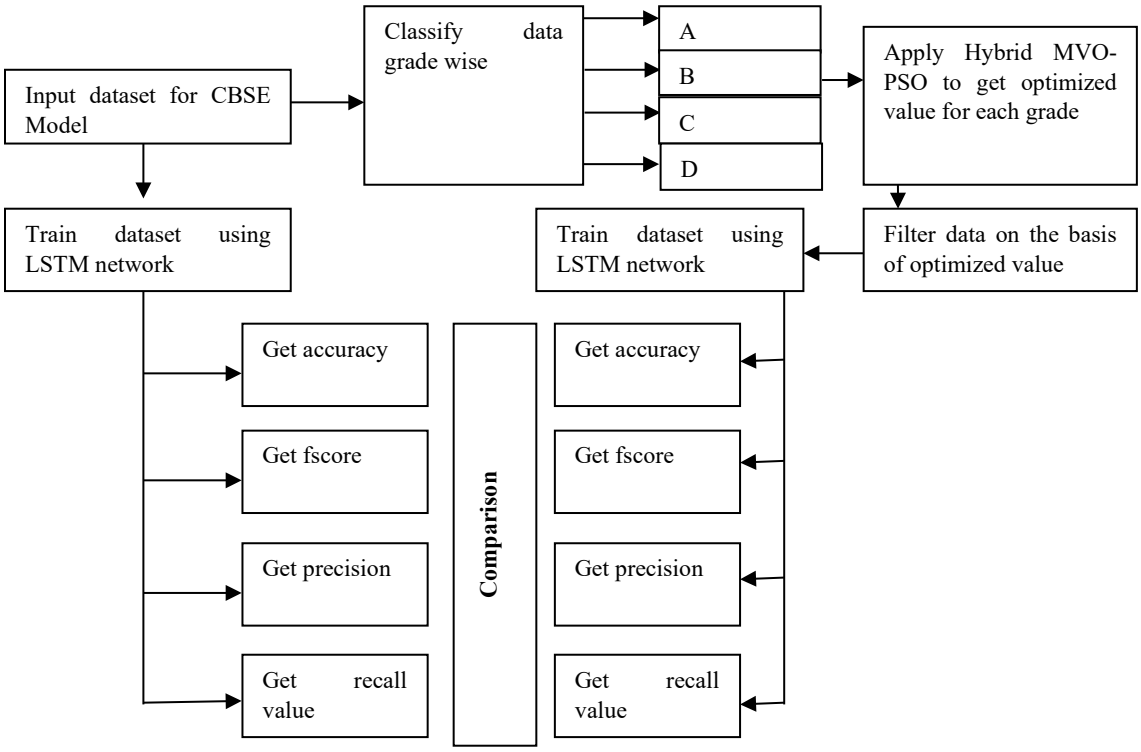


Figure 2: Architecture of Proposed model

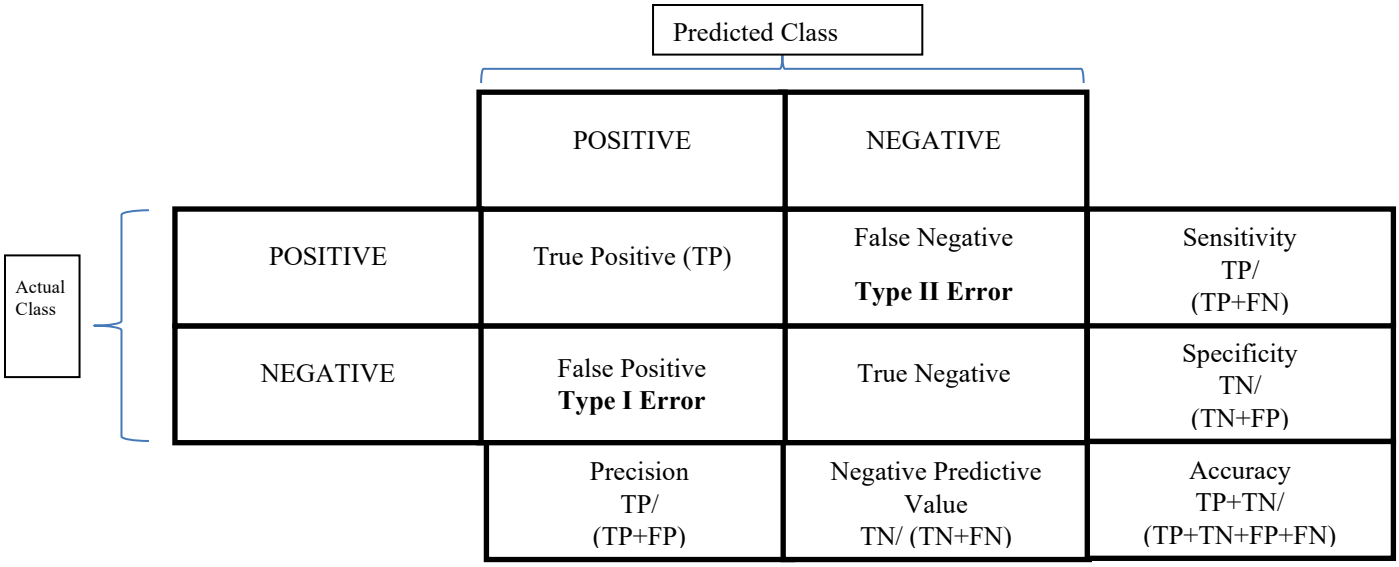


Figure 3: Confusion Matrix

Moreover, the time consumption during training of the network model is more. This research motivates to development of a model that should be trained fast as compared to previous models. Moreover, previous researches have also motivated to increase the accuracy using a two-layer LSTM model considering hidden layer. The proposed research is supposed to provide fast training to the dataset and more accurate prediction as compared to previous researches. The major objective of the research is to study existing literature on various component-based selections. The study and analysis of various component-based selections have been performed during research. Research would propose component-based selection with the support of LSTM layers. Then simulation would be made to perform result analysis. The comparison of existing and proposed work is made afterward.

Long Short-Term Memory networks have been considered as a category of recurrent neural networks. This is found capable to get taught order dependence in case of sequence prediction problems. This is a behavior needed in case of complicated issue domains like translation by machine. Long Short-Term Memory has been considered a complicated field of deep learning. This is difficult to understand Long Short-Term Memory. There has been little work in the field of Long Short-Term Memory. LSTM units are consisting of a 'memory cell'. These memory cells are capable to maintain data in memory for a large time. Users are moving from RNN to LSTM because it is introducing more controlling knobs. They are capable to manage the flow and mixing of Inputs according to trained Weights. So it provides flexibility during the management of outputs. Thus LSTM is providing the ability to manage and good results.

4.1. Performance Parameters

In this section, we will define the following parameters and confusion matrix is produced using true positive (TP), true negative (TN), false positive (FP), false negative (FN).

TP: True positives have been considered as correctly predicted positive values. In other words, the value of the real category is true and the value of the category that has been predicted is also true.

TN: True negative have been considered as correctly predicted negative values. In other words, the value of the real category is false and the value of the category that has been predicted is also false.

- FP: False positive is the case when the actual category is false but the predicted category is true.
- FN: False-negative is the case when the actual category is true but the predicted category is false.

Parameters utilized to confirm results have been f-score, recall, accuracy and precision which have been explained as follow:

1. Accuracy has been considered as intuitive performance measure. This is the ratio of correctly forecasted findings to total findings.

Accuracy = (True Positive + True Negative) / (True Positive + False Positive + False Negative + True Negative)

2. Precision has been considered as ratio of positive observations that have been correctly predicted to total predicted positive observations.

Precision = True Positive / (True Positive + False Positive)

3. Recall has been considered as ratio of positive observations that have been predicted in correct manner to overall findings in real class - yes.

Recall = (True positive) / (True Positive + False Negative)

4. F1 Score has been weighted average in case of Precision as well as Recall. Score is taking false positives as well as false negatives in consideration.

F1 Score = 2 x (Recall x Precision) / (Recall value + Precision)

5. Simulation

In this section, dataset of 629 packages have been considered for training purposed in research where the average rating is available considering factors such as number of reviews, total sentences, feature requests, feature requests in percentage, problem discoveries, problem discoveries in percentage, GUI Contents, Feature and Functionality, Improvement, Pricing, Resources, Security. A network model has been trained considering this dataset.

The grade is allotted according to the average rating

if Average Rating >4.5 then grade is A
if Average Rating lies between 4 and 4.5 then grade is B
if Average Rating lies between 3 and 4 then grade is C
if Average Rating <3 then grade is D

The classification of record counts according to grade have been discussed below

Table 2: Grade wise record count before optimization

GRADE	Record count
A	114
B	229
C	241
D	45
Total	629

In order to get 100% accuracy, there is need of following confusion matrix

Table 3: Confusion matrix required to get 100% accuracy

	A	B	C	D
A	114	0	0	0
B	0	229	0	0
C	0	0	241	0
D	0	0	0	45

But of the model is trained without optimization the confusion matrix is

Table 4: Confusion matrix before optimization

	A	B	C	D
A	110	1	1	0
B	1	225	2	0
C	2	2	237	1
D	1	1	1	44
Total	114	229	241	45

Considering above confusion matrix overall accuracy has been found

Results
TP: 616
Overall Accuracy: 97.93%

Table 5: Accuracy, precision, recall, f1 score in case of non-optimized dataset for 4 classes

Class	N(truth)	N(classified)	Accuracy	Precision	Recall	F1 Score
1	114	112	99.05 %	0.98	0.96	0.97
2	229	228	98.89 %	0.99	0.98	0.98
3	241	242	98.57 %	0.98	0.98	0.98
4	45	47	99.36 %	0.94	0.98	0.96

5.1. Optimization of GRADE A

Hybrid MVO-PSO is applied in order to get the optimized data set for training;

Hybrid MVO-PSO optimization of GRADE A results in

At iteration 50 the best universes fitness is 0.30119
At iteration 100 the best universes fitness is 0.30119
At iteration 150 the best universes fitness is 0.30119
At iteration 200 the best universes fitness is 0.30119
At iteration 250 the best universes fitness is 0.30119
At iteration 300 the best universes fitness is 0.30119
At iteration 350 the best universes fitness is 0.30119
At iteration 400 the best universes fitness is 0.30119
At iteration 450 the best universes fitness is 0.30118
At iteration 500 the best universes fitness is 0.30118

The best solution for class A obtained by MVO is: 4.6807. The best optimal value for class A of the objective function found by MVO is: 0.30118

Elapsed time is 6.498883 seconds.

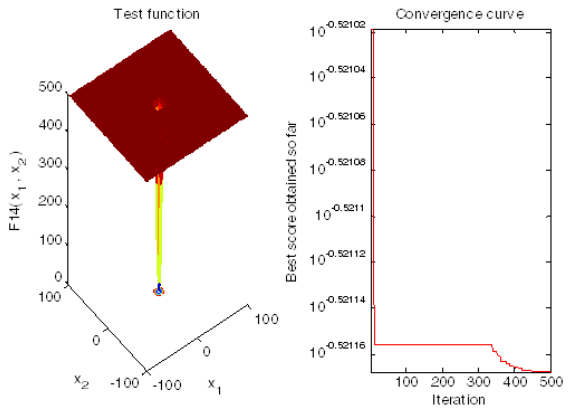


Figure 4: Hybrid optimization of Group A

After optimization 58 components have been selected where average rating is more than 4.6807

5.2. Optimization for Grade B

At iteration 50 the best universes fitness is 0.30119
At iteration 100 the best universes fitness is 0.30119
At iteration 150 the best universes fitness is 0.30119
At iteration 200 the best universes fitness is 0.30119
At iteration 250 the best universes fitness is 0.30119
At iteration 300 the best universes fitness is 0.30119
At iteration 350 the best universes fitness is 0.30119
At iteration 400 the best universes fitness is 0.30119
At iteration 450 the best universes fitness is 0.30119
At iteration 500 the best universes fitness is 0.30118

The best solution for class B obtained by Hybrid MVO-PSO is: 4.2356. The best optimal value for class B of the objective function found by Hybrid MVO-PSO is: 0.30118. Elapsed time is 3.650676 seconds.

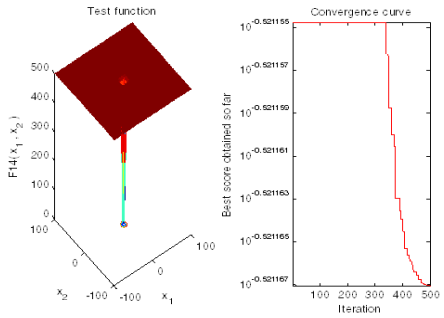


Figure 5: Hybrid MVO-PSO simulation to get optimized value for grade B

After optimization 120 components have been selected where average rating is more than 4.2356

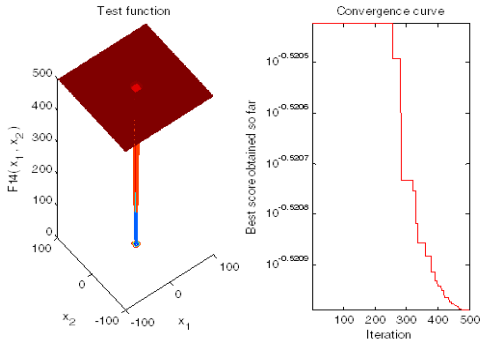


Figure 6: Hybrid MVO-PSO simulations to get optimized value for grade C

5.3. Optimization of Grade C

At iteration 50 the best universes fitness is 0.3017
At iteration 100 the best universes fitness is 0.3017
At iteration 150 the best universes fitness is 0.3017
At iteration 200 the best universes fitness is 0.3017
At iteration 250 the best universes fitness is 0.3017
At iteration 300 the best universes fitness is 0.30149
At iteration 350 the best universes fitness is 0.3014
At iteration 400 the best universes fitness is 0.30134
At iteration 450 the best universes fitness is 0.30132
At iteration 500 the best universes fitness is 0.30131

The best solution for class A obtained by MVO is: 3.6263. The best optimal value for class A of the objective function found by MVO is: 0.30131. Elapsed time is 3.911880 seconds.

After optimization 139 components have been selected where average rating is more than 3.6263

5.4. Optimization of Grade D

At iteration 50 the best universes fitness is 0.30156
At iteration 100 the best universes fitness is 0.30156
At iteration 150 the best universes fitness is 0.30156
At iteration 200 the best universes fitness is 0.30156
At iteration 250 the best universes fitness is 0.30156
At iteration 300 the best universes fitness is 0.30156
At iteration 350 the best universes fitness is 0.30153
At iteration 400 the best universes fitness is 0.30151
At iteration 450 the best universes fitness is 0.3015
At iteration 500 the best universes fitness is 0.30149

The best solution for class A obtained by MVO is: 2.4891. The best optimal value for class A of the objective function found by MVO is: 0.30149. Elapsed time is 2.961058 seconds.

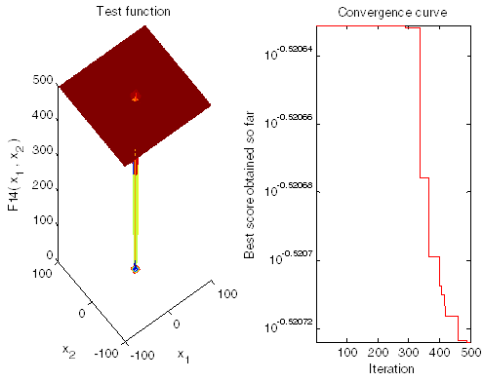


Figure 7: Hybrid MVO-PSO simulations to get optimized value for grade D

After optimization 31 components have been selected where average rating is more than 2.4891. After grade wise optimization the following components would be selected according to grade.

Table 6: Grade wise record count after optimization

GRADE	Component count
A	58
B	120
C	139
D	31
Total	348

But of the model is trained with optimization the confusion matrix is

Table 7: Confusion matrix produced after filter dataset considering optimization value

	A	B	C	D
A	58	0	0	0

B	0	119	1	0
C	0	1	137	0
D	0	0	1	31
Total	58	120	139	31

Considering above confusion matrix overall accuracy has been found

Results
TP: 345
Overall Accuracy: 99.14%

Table 8: Accuracy, precision, recall, f1 score in case of optimized dataset for 4 classes

Class	N(truth)	N(classified)	Accuracy	Precision	Recall	F1 Score
1	58	58	100%	1.0	1.0	1.0
2	120	120	99.43%	0.99	0.99	0.99
3	139	138	99.14%	0.99	0.99	0.99
4	31	32	99.17%	0.97	1.0	0.98

6. Comparative Analysis

This section is comparing the accuracy, precision, recall and F1 score before optimization and after optimization. Comparison of accuracy for previous and optimized data set has been shown in table 9.

Comparison of Accuracy for previous and optimized data set has been shown below

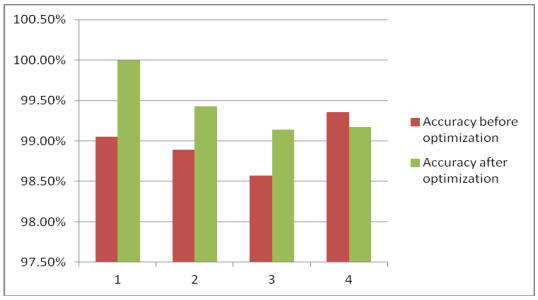


Figure 8: Comparison of accuracy

Comparison of precision for previous and optimized data set has been shown below

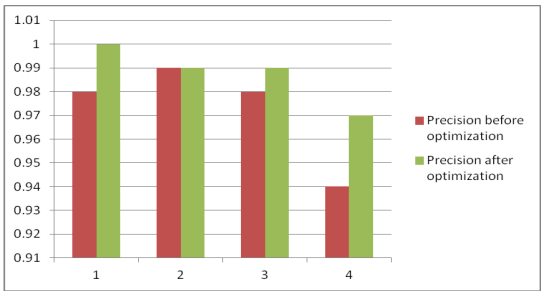


Figure 9: Comparison of precision

Comparison of recall for previous and optimized data set has been shown below

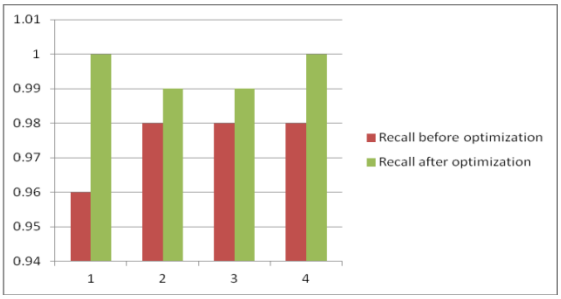


Figure 10: Comparison of recall

Comparison of fscore for previous and optimized data set has been shown below

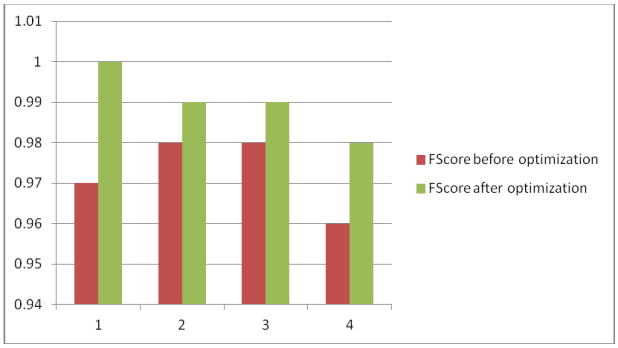


Figure 11: Comparison of Fscore

Table 9: Comparison of Accuracy, Precision, Recall, Fscore

Class	Accuracy before optimization	Accuracy after optimization	Precision before optimization	Precision after optimization	Recall before optimization	Recall after optimization	FScore before optimization	FScore after optimization
1	99.05%	100%	0.98	1.0	0.96	1.0	0.97	1.0
2	98.89%	99.43%	0.99	0.99	0.98	0.99	0.98	0.99
3	98.57%	99.14%	0.98	0.99	0.98	0.99	0.98	0.99
4	99.36%	99.17%	0.94	0.97	0.98	1.0	0.96	0.98

7. Conclusion

It has been concluded from simulation that the optimized dataset is capable to produce more accurate result as compared to non-optimized mechanism. Research has considered training and testing of dataset after Optimizing Software Component by deep neural network mechanism to improve the accuracy at the time of software selection. Deep Neural-Network mechanism has performed training and testing to get the confusion metrics of true

positive/negative and false positive/negative. Training has been performed using LSTM neural network mechanism to produce confusion matrix for getting accuracy, F-score, Recall and precision values. Simulation result concludes that the accuracy, F-score, Recall and precision values in case of optimized mechanism are better than non-optimized mechanism.

Table 10 shows that proposed work is providing high reliability and feasibility as compare to previous research models.

Table 10: Comparison of proposed work to existing researches

	Prediction of quality using ANN based on Teaching-Learning Optimization in component-based software systems [20]	PCI-PSO : Preference-Based Component Identification Using Particle Swarm Optimization [17]	A Hybrid Neuro-Fuzzy and Feature Reduction Model for Classification [29]	Proposed work
Optimization	✓	✓	✗	✓
Use of Neural network	✓	✗	✓	✓
CBSE	✓	✓	✗	✓
Performance	✓	✓	✗	✓
Accuracy	✗	✗	✓	✓
Reliability	✗	✗	✓	✓
Feasibility	✗	✗	✓	✓

8. Scope of Research

Such research could play a significant role in the field of software development, AI, big data processing, and many other fields where prediction is the major objective. Such a mechanism is suitable to provide an efficient and accurate approach to perform forecasting and decision-making in different areas. Moreover, further researches could use this research as a base in order to get more fruitful results.

References

[1] S. Di Martino, F. Ferrucci, C. Gravino, F. Sarro, "A genetic algorithm to configure support vector machines for predicting fault-prone components," Lecture Notes Computer Science (including Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **6759**, 247–261, 2011, doi: 10.1007/978-3-642-21843-9_20.

[2] M. Palviainen, A. Evesti, E. Ovaska, "The reliability estimation, prediction and measuring of component-based software," Journal of System and Software, **84**(6), 1054–1070, 2011, doi: 10.1016/j.jss.2011.01.048.

[3] A. A. A. Saed, W. M. N. W. Kadir, "Applying particle swarm optimization to software performance prediction an introduction to the approach," in 2011 5th Malaysian Conference in Software Engineering, MySEC 2011, 207–212, 2011, doi: 10.1109/MySEC.2011.6140670.

[4] J. F. Tang, L. F. Mu, C. K. Kwong, X. G. Luo, "An optimization model for software component selection under multiple applications development," Europian Journal of Operational Research, **212**(2), 301–311, 2011, doi: 10.1016/j.ejor.2011.01.045.

[5] U. Sharma, P. Singh, S. S. Kang, "Reusability of Software Components Using SVM Based Classifier Approach" , International Journal of Information Technology and Knowledge Management, **5**(1), 117–122, 2012.

[6] C. Mueller, "Multi-Objective Optimization of Software Architectures Using Ant Colony Optimization," Lecture Notes on Software Engineering, **2**(4), 371–374, 2014, doi: 10.7763/lnse.2014.v2.152.

[7] C. Singh, A. Pratap, A. Singhal, "Estimation of software reusability for component based system using soft computing techniques," in 2014 5th International Conference - Confluence The Next Generation Information Technology Summit, 788–794, 2014, doi: 10.1109/CONFLUENCE.2014.6949307.

[8] K. Tyagi, A. Sharma, "An adaptive neuro fuzzy model for estimating the reliability of component-based software systems," Applied Computing and Informatics, 10(1–2), 38–51, 2014, doi: 10.1016/j.aci.2014.04.002.

[9] M. Tan, C. dos Santos, B. Xiang, B. Zhou, "LSTM-based Deep Learning Models for Non-factoid Answer Selection," 1, 1–11, 2015, Available: http://arxiv.org/abs/1511.04108.

[10] D. Gao, X. Guo, L. Zhao, "Test case prioritization for regression testing based on ant colony optimization," in 2015 6th IEEE International Conference on Software Engineering and Service Science, 275–279, 2015, doi: 10.1109/ICSESS.2015.733905.

[11] G. Kumar, P. K. Bhatia, "Neuro-Fuzzy Model to Estimate & Optimize Quality and Performance of Component Based Software Engineering," ACM SIGSOFT Software Engineering Notes, **40**(2), 1–6, 2015, doi: 10.1145/2735399.2735410.

[12] S. Vodithala, S. Pabboju, "A dynamic approach for retrieval of software components using genetic algorithm," in 2015 6th IEEE International Conference on Software Engineering and Service Science, 406–410, 2015, doi: 10.1109/ICSESS.2015.7339085.

[13] S. Mirjalili, S. M. Mirjalili, A. Hatamlou, "Multi-Verse Optimizer: a nature-inspired algorithm for global optimization," Neural Computing and Application, 27(2), 495–513, 2016, doi: 10.1007/s00521-015-1870-7.

[14] A. Verma, S. Gupta, I. Kaur, "Inconsistency detection in software component source code using ant colony optimization and neural network algorithm," Indian Journal of Science and Technology, **9**(40), 2016, doi: 10.17485/ijst/2016/v9i40/101127.

[15] Z. Zhao, W. Chen, X. Wu, P. C. V. Chen, J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," IET Image Processing, **11**(1), 68–75, 2017, doi: 10.1049/iet-its.2016.0208.

[16] L. Sun, J. Du, L. Dai, C. Lee, "Multiple-target deep learning for LSTM-RNN based speech enhancement," in 2017 Hands-free Speech Communications Microphone Arrays, doi: 10.1109/HSCMA.2017.7895577.

[17] S. M. H. Hasheminejad, S. Gholamshahi, "PCI-PSO: Preference-based component identification using particle swarm optimization," Journal of Intelligence System, **28**(5), 733–748, 2021, doi: 10.1515/jisys-2017-0244.

[18] A. Gensler, J. Henze, B. Sick, N. Raabe, "Deep Learning for solar power forecasting - An approach using AutoEncoder and LSTM Neural Networks," in 2016 IEEE International Conference on Systems, Man, and Cybernetics, 2858–2865, 2017, doi: 10.1109/SMC.2016.7844673.

[19] O. Bhardwaj, S. Kumar Jha, "Quality assurance through soft computing techniques in component based software," Proceeding 2017 International Conference Smart Technology Smart Nation, SmartTechCon, 277–282, 2018, doi: 10.1109/SmartTechCon.2017.8358382.

[20] P. Tomar, R. Mishra, K. Sheoran, "Prediction of quality using ANN based on Teaching-Learning Optimization in component-based software systems," Software- Practice Experience., **48**(4), 896–910, 2018, doi: 10.1002/spe.2562.

[21] L. Mu and C. K. Kwong, "A multi-objective optimization model of component selection in enterprise information system integration," Computer Industrial Engineering, **115**, 278–289, 2018, doi: 10.1016/j.cie.2017.11.013.

[22] P. Chatzipetrou, E. Alégroth, E. Papatheocharous, M. Borg, T. Gorschek, and K. Wnuk, "Component selection in software engineering - Which attributes are the most important in the decision process?," in 44th Euromicro Conference on Software Engineering and Advanced Applications, 198–205, 2018, doi: 10.1109/SEAA.2018.00039.

[23] C. Diwaker, P. Tomar, R. C. Poonia, V. Singh, "Prediction of Software Reliability using Bio Inspired Soft Computing Techniques," Journal of Medical Systems, **42**(5), 2018, doi: 10.1007/s10916-018-0952-3.

[24] S. Gholamshahi, S. M. H. Hasheminejad, "Software component identification and selection: A research review," Software - Practice and Experience, **49**(1), 40–69, 2019, doi: 10.1002/spe.2656.

[25] C. Diwaker, "A New Model for Predicting Component-Based Software Reliability Using Soft Computing," IEEE Access, **7**, 147191–147203, 2019, doi: 10.1109/ACCESS.2019.2946862.

[26] H. Hu, Y. Li, Y. Bai, J. Zhang, M. Liu, "The Improved Antlion Optimizer and Artificial Neural Network for Chinese Influenza Prediction," Complexity, 2019, doi: 10.1155/2019/1480392.
[27] A. L. Imoize, D. Idowu, Imoize, Agbotiname Lucky, T. Bolaji, "A Brief Overview of Software Reuse and Metrics in Software Engineering," An International Scientific Journal, **122**(2), 56–70, 2019.
[28] G. Maheswari, K. Chitra, "Enhancing reusability and measuring performance merits of software component using data mining," International. Journal of Innovative Technology Exploring Engineering, **8**(6), 1577–1583, 2019, doi: 10.35940/ijitee.F1318.0486S419.
[29] H. Das, B. Naik, H. S. Behera, "A Hybrid Neuro-Fuzzy and Feature Reduction Model for Classification," Advances in Fuzzy System, 2020, doi: 10.1155/2020/4152049.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 72-79 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Industrial Engineers of the Future – A Concept for a Profession that is Evolving

Piwai Chikasha, Kemlall Ramdass, Ndivhuwo Ndou*, Rendani Maladzhi, Kgabo Mokgohloa

University of South Africa, Department of Mechanical and Industrial Engineering, Johannesburg, 2038, South Africa

ARTICLE INFO

Article history:
Received: 21 February, 2021
Accepted: 26 May, 2021
Online: 10 July, 2021

Keywords:
Industrial engineering
Innovation
Sustainability
Digital technologies

ABSTRACT

Just as industry is dynamic, constantly evolving according to the state of technology, economics, politics and so on, so must be, higher education. Studies have shown that higher education, for the past century, has constantly adapted to the dynamic skill and knowledge requirements of industry. This adaptation, however, is not always timeous and precise resulting in a widening gap between industry skill requirements and the skills that graduates receive during tertiary learning. This gap can be narrowed if higher education develops futuristic models that prepare students for not only the present day, but the future as well. Higher education in the fields of science, technology and engineering in particular, are in critical need of this future-prediction approach given the high levels of constant, and in some cases, even accelerating change or dynamics. This study develops a concept for industrial engineers of the future and demonstrates that is it possible to better prepare graduates for the uncertain future, by predicting some key skill requirements of industry ahead of time from information of yesterday and today.

1. Introduction

To enrol students for the Industrial Engineering qualification, and to provide the students with education for the set number of years, without taking in to account, the future of the profession, would be unfair practice. The same is in fact true for many other higher education qualifications, especially those in engineering. The reason for this is the dynamic nature of industry, the economy and even society as we know it. In the twenty-first century, it is therefore critical that higher education incorporates aspects of the futuristic profession, into the lessons of today, for the benefit of not only the students, but also the respective industry, and not only for today, but tomorrow as well.

This study discusses the concept of 'industrial engineers of the future'. It is well understood that the industrial engineering profession today is not as it was ten years ago. Similarly, the industrial engineering profession of 2030, ten years from now, may not be as that of the present day. If then, the higher education system is preparing tomorrow's engineers today, a tomorrow that presents different challenges demanding different solutions, substantial effort is required in ascertaining tomorrow's industry characteristics in today's education program.

In [1], the author elaborates that industrial engineering is multi-disciplinary, fully fledged and increasingly becoming more and more dynamic. The author points out that the growth of this profession was expedited in the twentieth century by the manufacturing sector, as well as government and service enterprises. The industrial engineering profession is widely understood to be dynamic, with a future dependent not only on the ability of the engineers to meet the respective economic and industrial operational demands, but also dependent on the ability of the engineers to innovate and actually drive the economic and industrial operational trends. To best answer the research questions, it is necessary to first approach the study from a more general point of view, considering the engineering discipline as a whole.

Technology is one certain part of engineering, that continues to evolve with time. This has been the case even over the past century. Curriculum therefore needs to prepare future engineers for a work environment that is in constant evolution, characterized by constant updates to the technologies used in engineering. In [2], this constant evolution is presented, demonstrating where the profession of industrial engineering is coming from.

This paper contributes to the question of where the profession is headed. Taking the software engineering profession for example, open-source technologies are topical and are highlighted as a basis for the future engineer. In [3], the author discusses artificial intelligence as having reshaped technology significantly over the past decades, and being posed for future dominance. Such

*Corresponding Author: Ndivhuwo Ndou, Email: mokgok@unisa.ac.za

are the factors to consider in ensuring that students today, are optimally oriented for tomorrow.

In [4], the author demonstrates how the future of engineering is one closely aligned to environmental sustainability as well as the well-being of humans. With growing demand for technologies, products and systems that are environmentally friendly and health friendly are also seeing growth in demand. This briefly demonstrates that engineering is not static but rather dynamic. If the trends of engineering can be anticipated or predicted, it is in the best interest of the students and overall economy and industry.

1.1. Problem statement

Conventionally, South African higher education is not adequately future oriented or future proof. This contributes to education lagging behind the dynamics of industry instead of overtaking industry in order to shape-out society and technological trends. The problem is that higher education lacks robust systems and mechanism that are capable of identifying and strategically accommodating elements of the future of the respective profession into today's education.

1.2. Research question

Given the dynamic nature of industry, can we develop a concept for industrial engineers of the future, in order to provide career-oriented future-proofing for industrial engineering graduates? What are the key factors to consider when developing the concept of industrial engineers of the future?

1.3. Concluding remarks to introduction

The paper is structured as (a) Introduction which provides some background information of the research, as well as the problem statement and the research question. (b) Literature review (c) Research methodology (d) Discussion and findings; and (e) Conclusions and recommendations. The literature review provides a view of some aspects of the industrial engineering profession that form a strong basis for the future of the profession. The research methodology therefore pivots on these aspects to create a complete concept.

2. Literature review

The literature survey of this work is structured to initially assess past trends of the industrial engineering profession. The idea is to appreciate the fact that the profession is indeed dynamic, shifting decade after decade in terms of skill areas. The next step then, is to establish a projection of the competences that will shape the industrial engineering profession of tomorrow.

2.1. Industrial engineering trends

The industrial engineering profession, as hinted by the title, is industry driven. In [5], the author illustrates how the future of industrial engineering is heavily influenced by customer demands and expectations. The demands and expectations of the customer are on a trajectory of rapid growth, giving rise to the need for constant improvement in production processes as well as operational systems, in order to minimize cost yet maximizing quality. Industrial engineering is key in realizing this goal.

In [6], authors show that Internet of Things (IoT) is a key part of tomorrow's industry. Industrial engineers of the future are therefore expected to embrace the concept of IoT. At curriculum level, it is presented that it is necessary to adjust curriculum to take more consideration of IoT in order to produce graduates who are more oriented for the work environment of today and also tomorrow.

In [7], the authors explore the value of interdisciplinary competences to the future of industrial engineering. It is shown that, due to the diversity of industrial engineering work, which continues to evolve, interdisciplinary skills are a critical aspect. A teaching method which emphasizes interdisciplinary scope and teamwork is proposed. In [8], the authors explicitly demonstrate how approaching engineering from an interdisciplinary perspective promotes technological development for national prosperity.

In [9], the authors take a look at how the circular characteristics of the economy affect the future of industrial engineers. It is demonstrated that the industrial engineer of the future is one to master concepts of circular economy strategies. It is proposed that circular economy concepts be realized in curriculum through collaborative design project works.

2.2. Projection

Analysing trends in the industrial engineering profession is only the first step in developing the concept of industrial engineers of the future. The next step is to project those areas of the profession that are likely to take the forefront, in the years to come. This involves prediction of those skills and competences that are likely to see increased job market demand. In [10], authors explore the idea of predicting future needs of the industrial communication field. In [11], a method to measure educational alignment to industry is proposed, further alluding to the relevance of the topic. This study predicts that the following competences are critical for the industrial engineer of the future:

1. Innovation and entrepreneurship
2. Sustainable development
3. Digital technologies

Collectively, these three competences are identified to shape the industrial engineer of the future.

2.2.1. Innovation and entrepreneurship

The national and even global call for an improved entrepreneurship environment means that the economy can anticipate growth in small to medium enterprise establishments as well as innovation and development, especially from the youth population. In [12], authors show that as for engineering education, this call has largely been answered, with entrepreneurship education being integrated into engineering degrees. In [12], the authors analyse the impact of entrepreneurial competence in the engineer of the future, to identify and appreciate the role of entrepreneurship education in economic development. Results of the study indicate that the future of the engineering profession strongly demands entrepreneurial competences, especially at graduate level.

In [13], the authors point out that over the past few decades, the engineering job market has generally become more competitive. The authors further show that trends indicate that the job market will in fact become even more competitive in the near and long-term future. It is therefore critical to invest in entrepreneurship competences for the future engineers. In [14], authors argue that entrepreneurial competences are so vital for the future engineer that there is need to globally standardize how entrepreneurial concepts are incorporated into engineering curriculum.

In [15], the authors actually argue that entrepreneurship education cannot be separated from engineering education. A successful career in engineering, as the world races Industry 4.0, would be heavily centred on entrepreneurship competences. In [16], authors in order to safeguard the career, propose the synchronization of the engineering and business courses. This is shown to potentially promote entrepreneurial competences towards the required level.

2.2.2. Sustainable development

The Sustainable Development Goal (SDG) initiative is a global movement and one that has massive influence on the economic future of South Africa, as with every other country. The SDG initiative calls for intensive transformation for every country, affecting the operations of the entire national value chain including government, private sector, industry, civil society and science [17] and has become the cornerstone of business across the world. Businesses are increasingly aligning more and more towards sustainable solutions, and massive investments are increasingly being channelled towards research into sustainable solutions for example in the field of electrical vehicles. Since the turn of the century, massive strides have already been made towards the SDG initiatives and it has never been more certain that sustainable development is the heart of the future global economy, as enterprises start to reflect sustainability concepts through internal objectives and visions. In [18], authors show that the concepts of environmental and social sustainability, are the most commonly discussed.

While the market is becoming more competitive, operational costs are generally increasing. This makes it important to be able to optimize business operations, systems or products. This includes improved loss control systems, price management, product development/improvement, resource management and so on. Sustainable development concepts are expected to be a significant part of the work of the industrial engineer of the future. In [19] authors reiterate that a cleaner/greener industry has huge economic implications and highlight that the growing advocacy for sustainable production will continue to grow into the future. In [20], the authors show that renewable energy projects have national-scale economic impact, and such projects deserve national support. The authors also highlight that the demand for renewable energy solutions will continue to rise into the future. In [21], authors conduct a survey on public statements of mining associations and the extent to which statements relating to sustainability are incorporated into policy, and the survey reveals that while the mining industry has allowed a sustainability shift, more needs to be done. The study shows that, out of 61 associations, 67 percent had public statements on sustainability.

In [22], the authors also highlight that sustainability concepts have been pivotal in the mining industry since the year 2010, and that it has become necessary for professionals in the mining industry to appreciate and accommodate sustainable mining practices. In [23], authors discuss how sustainable development is so key for the future that the concept of sustainable development should not be localized to the private sector only, or to engineering and science only. The authors show how, at national level economics, implementing strategic sustainable development objectives may contribute to improved trade deficit management.

In [24], the author discusses the steel-making process, revealing how the use of the by-products contributes to sustainable development. Literature suggests that slag from the iron and steel industries, when used to complement cement, improves the micro-structure of built concrete. In [25], authors study the current state of the hotel industry and show that more innovation is required to meet the sustainability demands of the industry, especially given the sophistication of customer requirements. Innovative controls are required to improve waste and power management, to make operations more sustainable.

2.2.3. Digital technologies

Industry 4.0 places automation at the centre of industrialisation, and digital technologies form the basis for automation systems hence the need for emphasis on digital technologies for the industrial engineer of the future. The twenty-first century is characterized by growing popularity of digital technologies, defined by electronic tools, machines, systems, and software that generate, process and utilize data. The manufacturing sector is particularly going through a fourth revolution and this calls for more strategic approaches to industrial engineering education, and engineering in general. In [26], authors discuss how Industry 4.0 calls for digital technologies, especially in the area of Internet of Things (IoT), big data, cloud computing as well as data analysis and processing.

In [27], the author highlights that in this digital era, there is more need to implement systems that can commercialize higher education research, that is to say, systems that can transfer academic work to industry. In [28], authors study the trends in industrial maintenance services and point out that the fourth industrial revolution will see the concept of 'big data' becoming key to efficient industrial maintenance management, in the near future. In the medical field, surgical processes have seen automation, patient profiles have been digitized for improved access to information, such as medical history [29]. In [30], authors, amidst the COVID-19 pandemic, discuss digital technologies that may ensure better patient isolation, by creating automated virtual centres to minimize physical crowding at hospitals and clinics.

Finance and economics have both been substantially transformed by digital technologies. In [31], authors propose the use of advanced digital mathematical models and algorithms to assess the feasibility of engineering projects in order to support capital investment decision making. The proposed digital system is proven to be robust and able to process large amounts of diverse data. In [32], the authors trace the age of digital marketing technologies in the sales of industrial services. The authors illustrate that smart systems have gained massive ground in

industry over the past decade and such technologies are still on the rise.

In [33], internet marketing is compared to traditional marketing and it is shown that internet marketing is a superior marketing tool in today's economy, and more so, tomorrow's. In agriculture, the present and future have never been more digital. Agriculture across the entire planet has been redefined by automation and robotics, information communication technologies, drones and sensors as well as digital surveys and advanced climate and environment modelling technologies. In [34], authors show how digital technologies have brought about solutions to the challenge of access to the market in farming, especially for the small-holder farmer. In [35], authors introduce the concept of 'Agriculture 4.0' with alignment to that of 'Industry 4.0', highlighting that the era of Agriculture 4.0 has begun and is characterized by advanced digital and biotechnological innovations in agriculture. As the world becomes more and more digital, the threat to information security however, worsens. Cyber-security is therefore a key area of technology today, and even more, in future, as new cyber-security threats continue to evolve [36].

We therefore identify a gap, that while it is evident that the industrial engineering profession is dynamic, no structures have been set up yet, to explicitly allow the concept of future industrial engineers to be incorporated into curriculum.

3. Methodology

The Fourth Industrial Revolution (Industry 4.0) is summarised by automation, smart technology solutions, internet of things and improved machine communication. Understanding the requirements of Industry 4.0 is the first step in determining and developing the concept of the industrial engineer of the future. The requirements may be clearly spelt out by investigating the needs of the different players in this Industry 4.0 dimension. A survey is therefore conducted to establish the skill and knowledge requirements of the industrial engineer of the future, based on current dynamics of the industry. The survey responses are a collection from fifty-eight South African and Zimbabwean engineering and technological companies, 65% being start-ups. The idea of including more start-ups in the survey is not only the fact that start-ups are easier to reach (the Harare Institute of Technology alone host a start-up hub with over fifteen enterprises) but also that such are the players expected to reshape the industry of tomorrow.

A quantitative approach is taken for the study, involving the analysis of the distribution of the future skill and knowledge needs of the engineering business processes of the respondent companies. A quantitative approach is adopted because real data from real enterprise specialising in various engineering works and projects will reflect a realistic and measurable index of the requirements of the future engineer.

The feasibility of this approach can be referenced to a study conducted in [10] which reviews technological trends of industry 4.0 and the impact thereof, on industrial communication.

The following assumptions are made:

- As current players in industry, engineering companies do provide reliable insight as to the futuristic trends in skill and knowledge requirements, based on the operations being carried out today, and the possible improvements.
- Start-up type companies represent industrial players in a drive to penetrate the industry with innovative solutions and hence amplify the reflection of the future industry.

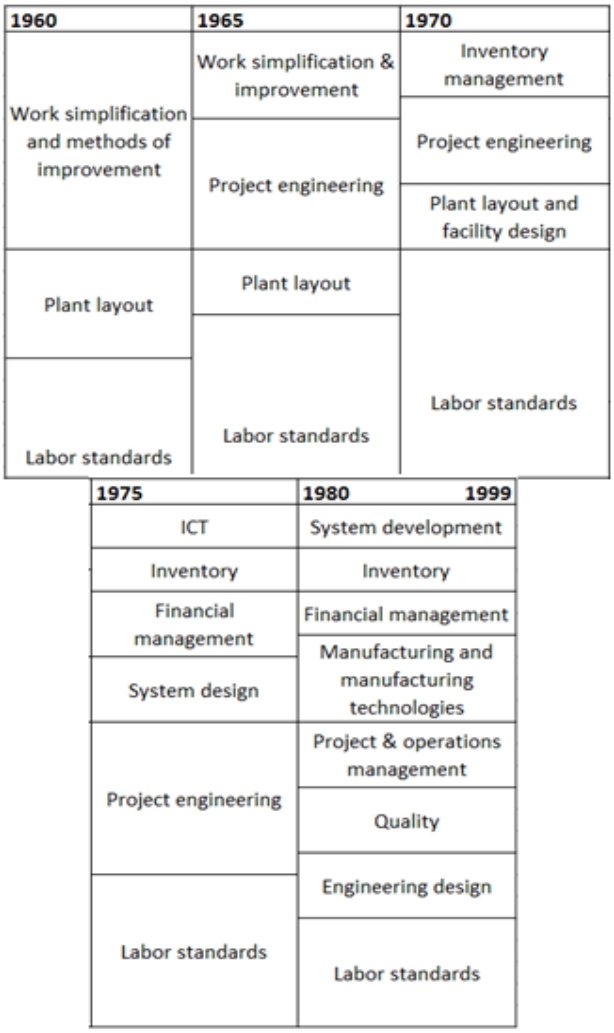


Figure 1: Twentieth century industrial engineering function trends - adopted from [2]

The illustration above proves that the industrial engineering profession is indeed under evolution. The next step is therefore to predict where this evolution is headed towards by means of survey data.

3.1. Data collection

Data is collected from an industrial survey with practising companies, to determine, from an operational perspective, the needs of the future industry.

3.2. Data evaluation

To evaluate the data, a comparison is made, against a recent study which investigated the technological trends of industry 4.0 and the impact of these trends on industrial communication. Reference is also given against the general directives surrounding industry 4.0. The evaluation shows acceptable levels of data and method reliability.

4. Discussion and findings

The general trend of the past century is that industrial engineering saw a shift from industry/factory functions to more commercial functions. The industrial survey conducted in this work, with enterprises practising engineering, aims to extend the trends given in [2] beyond the year 2025, and also to close the gap partially, from the year 2000 to-date. It is established that the period between the year 2000 to-date was generally characterised by growing diversity and flexibility within the industrial engineering profession. This period saw a rise in demand, particularly, for such competences as those aligned to:

- financial sustainability
- environmental friendly business processes
- process, system or product optimisation
- risk and loss control
- product and system development

This trend can actually be cross-referenced to the general global developmental priorities and goals. Risk management and loss control in particular, are two areas that both the private sector and government have prioritised over the past two decades. The idea of loss control has been globally received as a means to increase profits, without necessarily increasing product pricing, especially given the competitiveness of today's market. Impacts of the United Nations Sustainable Development Goals (SDGs) have significantly affected industry across the planet, especially in terms of business management and operations. The decade from the year 2020 has been set as the 'Decade of Action', with respect to the SDGs and this global campaign is expected, and is already affecting industry. Consequently, this goes to affect several professions, with industrial engineers included. This makes it vital for the education system to incorporate and prioritise sustainability concepts. It is in fact common at present day, for job functions to involve duties to do with operational sustainability, for example an industrial engineer at a mining company will typically be tasked to develop systems to manage power usage. With growing popularity of the UN SDGs, the majority of project support initiatives today, as well as project funding schemes, both locally and internationally, are typically set to prioritise projects with elements of sustainability. Figure 2 is an extension to Figure 1, illustrating a forecast beyond the year 2025 in terms of industrial engineering skill and knowledge needs, based on the survey conducted.

Figure 2 depicts the functions that are predicted to top demand as far as the industrial engineering profession is concerned. It is seen that the key function areas of innovation, optimisation techniques, sustainability, automation, digital systems and business development score high in terms of demand. Innovation relates to business development and collectively form the basis for the business growth, especially in the case of starting and early stage enterprises, such enterprises that today employ more people than formal established enterprises. Optimisation and automation relate to advanced industrial decision-making tools and digital technologies that empower machines to execute tasks in a manner that more effective, accurate or efficient than the human counterpart. This narrows down the key concepts to innovation, digital technologies and sustainable development.

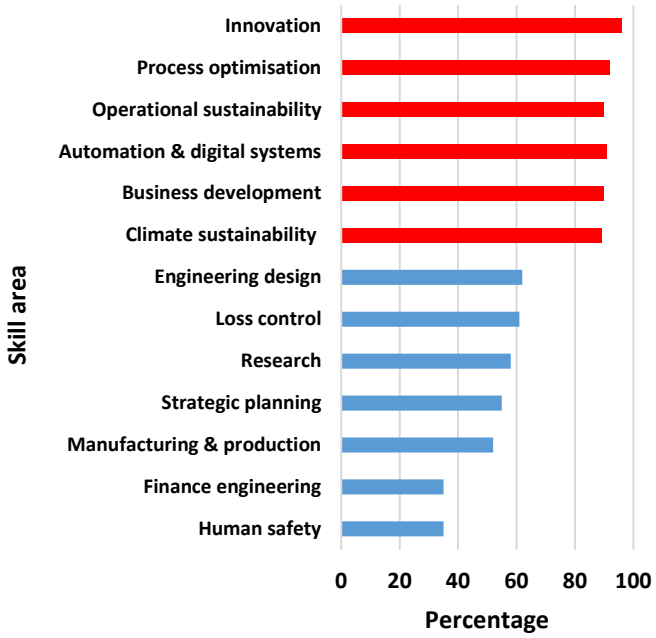


Figure 2: Extended industrial engineering function trends

The projections presented by Figure 2 call for the educational systems to incorporate the predicted elements into curriculum more explicitly. While curriculum is under constant management for the purpose of better addressing the evolving needs of the industry, this study proposes an improvement, characterised by prioritisation of some three aspects of industrial engineering education. The three aspects are:

1. Innovation and entrepreneurship
2. Sustainable development
3. Digital technologies

These economic aspects are projected to dominate the future of the industrial engineering profession. Academic prioritisation of these aspects is expected to better orient the industrial engineer for the future. The discussion presented in this work is summarised by Figure 3, which illustrates the three aspects above as pillars for the future of industrial engineering students.

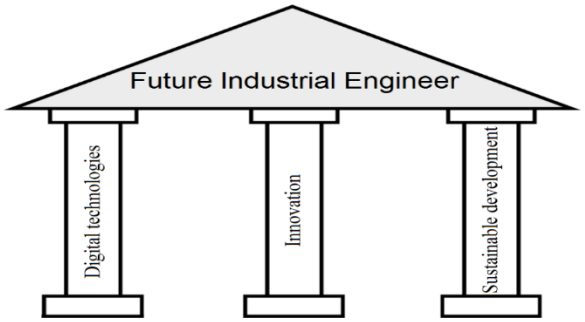


Figure 3: Illustration of the concept of the industrial engineer of the future

As with any structure, a foundation is necessary. This foundation is proposed from the perspective of graduate attributes. In [37], authors define graduate attributes as the qualities that a university or college community agrees to develop in the students during their study period at the respective institution. These attributes exceed academic competences and technical skills reaching over to qualities that prepare graduates for social good, in a global economy that is dynamic and with an unknown future. The most common and typical graduate attributes include:

- Effective communication
- Strong citizenship
- Leadership skills
- Problem solving skills

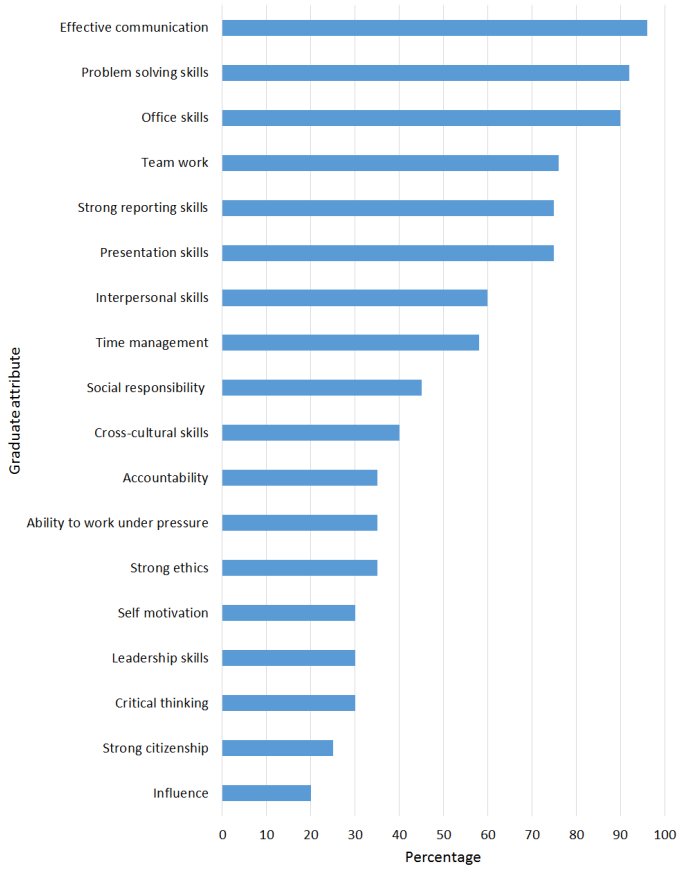


Figure 4: Graduate attribute distribution

Today, graduate attributes have become a core element of tertiary learning outcomes and integration of generic attributes into university or college curriculum is on a world-wide rise, especially towards promoting development of skills that better equip students for the work environment and also for self-employment. Higher learning institutions have therefore over the past few decades, placed 57 increasing value on developing graduate attributes and ensuring that the attributes are reflected within graduates.

In [38], authors show that one challenge is that some lecturers or educators may not see or perceive the value and essence of developing graduate attributes and may thus consequently be reluctant to realise graduate attribute obligations throughout their teaching experience. For graduate attributes to be effectively implemented, curriculum is one place to look at.

In this study, an investigation into the trends of distribution of demand for graduate attributes is made for the industrial engineering profession. The study outline is to collect data from industry through a survey to determine the attributes that employers seek when recruiting industrial engineers. Input data for this survey is collected from an on-line job advertisement platform. For each industrial engineering job post, recruiters typically highlight the following:

- job description
- minimum qualification
- desired personal attributes

Such on-line platforms therefore provide diverse information for survey. In [39], authors show that is it possible and effective, to collect education management control data from job advertisement platforms. One hundred job post samples are processed to produce the distribution depicted by Figure 4.

Figure 4 shows that, from the attribute study conducted, the communication attribute is most critical for the industrial engineer. In [40], authors conduct a study to determine the most demanded gradate attribute. By cross-referencing the survey results in Figure 4 above (industrial engineering profession), with those presented in [40] (general multi-profession study), it is possible to determine any reciprocity as well as any data disagreement too. The strategy is to therefore compare the top attributes sought by industry, as presented in [40], against findings from Figure 4. Table 1 outlines the comparison.

Table 1: Comparison of findings from graduate attribute study

Literature [40]	Figure 4
1. communication	1. communication
2. teamwork	2. problem-solving
3. citizenship	3. office skills
4. critical thinking	4. teamwork
5. problem-solving	5. reporting skills

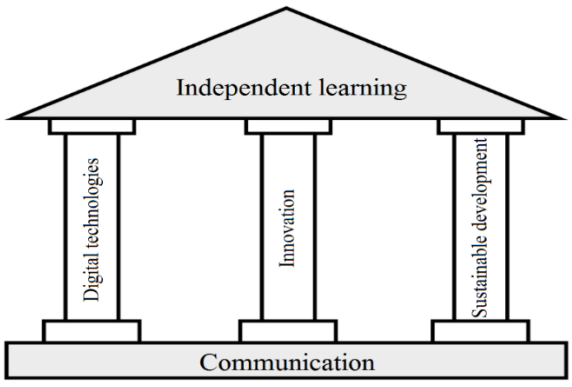


Figure 5: Final illustration of the concept of the industrial engineer of the future

It can be observed from this combined analysis, that the communication attribute is of highest priority, not only from the industrial engineering perspective, but also in a general sense. The problem solving and teamwork attributes are also of high priority. Analysing the attributes presented in Table 1 deeper, reveals that of the top attributes, the aspect of communication is in fact represented in 50 percent of the given top attributes.

The pillar diagram is therefore revised to present the industrial engineer of the future as depicted by Figure 5 below.

This figure shows that the industrial engineer of the future is one to master digital technologies, innovation and sustainable development, over and above, the attribute of effective communication. The advantage of the proposed concept is that the concept interprets the state of the real industry of today to produce a pathway for students to follow today, in order to fit into industry tomorrow and to make the necessary knowledge and skill contribution that will be required. This study considers real industry data (survey). The disadvantage is that industry, economics and society being dynamic, the requirements of today and those of 2030 may not be the same, hence the need for an evolutionary approach as an extension to the proposed solution.

4.1. Way forward

Today, there is worldwide concern for sustainable development, with emphasis on the 'three Ps' that is, the People, Planet and Profit, where businesses ensure not only the well-being of shareholders through profits, but also the well-being of the people and the planet. This study proposes curriculum to incorporate these three Ps. In the context of South Africa, the main economic, environmental, social and political factors (in the sense of sustainability) that need to be considered are proposed as:

- Women empowerment
- Youth empowerment
- Graduate unemployment
- Electricity shortage
- Environmental pollution
- Increasing fuel price
- Climate change
- Import substitution (indigenisation)

It is proposed that these topics become mark points for student work, especially projects. Design solutions for example, could be evaluated from an environmental perspective, towards determining potential impact on the planet.

In the case of innovation, over and above the inclusion of an innovation/entrepreneurship course within engineering curriculum, it is noted through this study, that there is need to improve the entrepreneurial experience of the engineering student during the study process. To achieve this goal of promoting entrepreneurship through curriculum, this study proposes that a unique case-study approach be taken, where learners are tasked with a target number of practical industry-based freelance work projects, as part of the study experience. These projects are set to be sourced from online freelance work advertising platforms, reflecting the true nature of the needs of the industry. Each industrial engineering student is required, after completing a project, to present the project for evaluation in class, by both lecturers and students. The projects are then evaluated according to:

1. Project scope and objective
2. Technological aspect
3. Deliverables
4. Complexity
5. Approach
6. Financial perspective

This proposal ensures that innovation and entrepreneurship are better accounted for by curriculum, especially given the prevailing trends where entrepreneurship is rapidly growing as a source of employment across various industries.

5. Conclusions

The skill and knowledge requirements of any profession evolve over time, according to various contributing factors such technology, economics, politics and so. In order to maintain relevance of higher education, it is important to adapt education to the prevailing industrial requirements. It is even more important to actually predict where the industry requirements' evolution is headed, and to then manipulate higher education accordingly. In this study, the concept of the industrial engineer of the future is discussed. Based on an industrial survey, this study shows that there are three key knowledge areas which will shape the industrial engineer of the future. These three are innovation and entrepreneurship, sustainable development and finally, the digital technologies.

It is recommended that the industrial engineering education lifecycle takes these three knowledge areas as knowledge areas of high priority for the benefit of graduate industrial engineers. Future research is recommended to allow scheduled updates and trend analysis tools to measure the precision of the model of the industrial engineer of the future as predicted today.

6. Conflict of Interest

The authors declare no conflict of interest.

References

[1] R. Kumar, Industrial engineering, Jyothis Publishers, 2020, ISBN: 9353962854.

[2] M. Zandin, Industrial engineering handbook, vol. 5 McGraw Hill Companies, 2004. ISBN 49780070411029.

[3] P. Debney, "The engineer of the future is a centaur", The Institution of Structural Engineer, **98**(1), 24–31, 2020.

[4] Y. Lobanova, "Basic guidelines, principles and psychological-pedagogical technologies of creation of the engineer of the future", Integrating Engineering Education and Humanities for Global Intercultural Perspectives. IEEHGIP 2020. Lecture Notes in Networks and Systems, vol 131. Springer, Cham, 2020, doi.org/10.1007/978-3-030-47415-7_66.

[5] R. Kumar, Industrial Engineering. Jyothis Publishers, 2020, ISBN: 978-93-5396-285-2.

[6] M. Wollschlaeger, T. Sauter, J. Jasperneite, "The future of industrial communication: Automation networks in the era of the internet of things and industry 4.0", IEEE industrial electronics magazine, **11**(1), 17–27, 2017, doi: 10.1109/MIE.2017.2649104.

[7] L. G. Veraldo, M. B. Silva, J. Lourenco, C. Herculano, E. Soares, H. Sampaio, J. L. Rosa, "Assessment method of the competencies of industrial engineer in an interdisciplinary project, in 12th International CDIO Conference", CDIO Project in Progress Contributions, Turku, Finland, 2016.
[8] C. Butts, H. Valentine, H, Building on our past to engineer the future, Peanut Science, **46**(1A), 82–90, 2019 doi: 10.3146/0095-3679-46.1A.82.
[9] J. Gonzalez-Dominguez, G. Sanchez-Barroso, F. Zamora-Polo, J. Garcia-Sanz-Calcedo, "Application of circular economy techniques for design and development of products through collaborative project-based learning for industrial engineer teaching", Sustainability, **12**(11), 2020, doi.org/10.3390/su12114368.
[10] M. Wollschlaeger, J. Thilo, "The future of industrial communication: Automation networks in the era of the internet of things and industry 4.0", IEEE, **11**(1), 17 - 27, 2017. DOI: 10.1109/MIE.2017.2649104.
[11] P. N. Chikasha, K. Ramdass, K. Mokgokloa, R. Maladzhi, "Aligning industrial engineering education with industry through atomic curriculum manipulation", The South African Journal of Industrial Engineering, **31**(4), 92 – 103, 2020.
[12] V. Barba-Sanchez, C. Atienza-Sahuquillo, "Entrepreneurial intention ´ among engineering students: The role of entrepreneurship education", European Research on Management and Business Economics, **24**(1), 53–61, 2018, doi.org/10.1016/j.iedeen.2017.04.001
[13] M. Martinez, X. Crusat, "The entrepreneurship journey: Fostering engineering students entrepreneurship by startup creation", 2019 IEEE Global Engineering Education Conference (EDUCON), Dubai, 8-15 (April) 2019, 120–123, doi: 10.1109/EDUCON.2019.8725115.
[14] C. L. Entika, M. K. Jabor, S. Mohammad, S. Osman, "Defining the meaning of entrepreneurship education for future engineering graduates", 2017 7th World Engineering Education Forum (WEEF), Kuala Lumpur, (November), 2017, 290–293, doi: 10.1109/WEEF.2017.8467166.
[15] C. Hixson, M. Paretti, "Unpacking why engineering faculty members believe entrepreneurship is valuable for engineering education", Advances in Engineering Education, **7**(1), 2018.
[16] J. E. Lugo, M. L. Zapata-Ramos, M. J. Perez-Vargas, "Promotion of innovation and entrepreneurship in engineering design by synchronizing engineering and business school courses", in International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 50138, American Society of Mechanical Engineers, 2016, doi: 10.1115/DETC2016-59701
[17] J. D. Sachs, G. Schmidt-Traub, M. Mazzucato, D. Messner, N. Nakicenovic, J. Rockstrom, "Six transformations to achieve the sustainable development goals", Nature Sustainability, **2**(9), 805–814, 2019.
[18] E. Aimagambetov, R. Bugubaeva, R. Bespayeva, N. Tashbaev, "Model of sustainable development of tourism industry in Kazakhstan (regional perspective)", Public Policy and Administration, **16**(2), 179-197, 2017, doi:10.13165/VPA-17-16-2-02
[19] G. C. Oliveira Neto, J. M. F. Correia, P. C. Silva, A. G. Oliveira Sanches, W. C. Lucato, "Cleaner production in the textile industry and its relationship to sustainable development goals", Journal of cleaner production, 228, 1514–1525, (August) 2019, doi.org/10.1016/j.jclepro.2019.04.334
[20] S. R. Paramati, N. Apergis, M. Ummalla, "Dynamics of renewable energy consumption and economic activities across the agriculture, industry, and service sectors: evidence in the perspective of sustainable development", Environmental Science and Pollution Research, **25**(2), 1375–1387, 2018, doi: 10.1007/s11356-017-0552-7.
[21] V. Vivoda, D. Kemp, "How do national mining industry associations compare on sustainable development?" The Extractive Industries and Society, **6**(1), 22–28, 2019, doi: 10.1016/j.exis.2018.06.002.
[22] J. Caron, S. Durand, H. Asselin, "Principles and criteria of sustainable development for the mineral exploration industry", Journal of Cleaner Production, 119, 215–222, (April) 2016, doi.org/10.1016/j.jclepro.2016.01.073.
[23] M. Alawin, M. Oqaily, "Current account balance, inflation, industry and sustainable development in Jordan", Revista Galega de Economia, **26**(3), 45–56, 2017.
[24] I. Yuksel, A review of steel slag usage in construction industry for sustainable development, Environment, Development and Sustainability, **19**(2), 369–384.
[25] F. Melissen, E. Cavagnaro, M. Damen, A. Duweke, "Is the hotel industry prepared to face the challenge of sustainable development?" Journal of Vacation Marketing, **22**(3), 227–238, 2016, doi: 10.1177/1356766715618997
[26] Y. Li, J. Dai, L. Cui, "The impact of digital technologies on economic and environmental performance in the context of industry 4.0: A moderated mediation model", International Journal of Production Economics, 2020, 107777.
[27] W. Sutopo, "The roles of industrial engineering education for promoting innovations and technology commercialization in the digital era", in IOP Conference Series: Materials Science and Engineering, 495, 012001, IOP Publishing, 2019.
[28] S. Marttonen-Arola, D. Baglee, "Adoption of information-based innovations in industrial maintenance", in ISPIM Conference Proceedings, 1–16, The International Society for Professional Innovation Management (ISPIM), 2019.
[29] S. Kalajdziski, N. Ackovska, ICT Innovations 2018. Engineering and Life Sciences: 10th International Conference, ICT Innovations, Ohrid, Macedonia, (September), 940, 2019, Springer, ISBN 978-3-030-00825-3
[30] M. Javaid, A. Haleem, R. Vaishya, S. Bahl, R. Suman, A. Vaish, "Industry 4.0 technologies and their applications in fighting covid-19 pandemic", Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 2020, doi: 10.1016/j.dsx.2020.04.032
[31] Y. Kirillov, E. Dragunova, A. Kravchenko, A. Dorofeeva, "Innovations in engineering: analysis of the increase effect in net present value," IOP Conference Series Materials Science and Engineering, 2020, 843:012018.
[32] M. Classen, T. Friedli, T., "Value-based marketing and sales of industrial services: A systematic literature review in the age of digital technologies", Procedia CIRP, 83, 1–7, 2019.
[33] K. O. Kayumovich, F. S. Annamuradovna, "The main convenience of internet marketing from traditional marketing", Marketing Academy, **1**(52). doi: 10.24411 / 2412-8236-2020-10101.
[34] U. Deichmann, A. Goyal, Y. Mishra, Y, Will digital technologies transform agriculture in developing countries? The World Bank, 2016, doi.org/10.1111/agec.12300.
[35] Y. Klerkx, D. Rose, D. Dealing with the game-changing technologies of agriculture 4.0: How do we manage diversity and responsibility in food system transition pathways? Global Food Security, 24, 2020, 100347. doi.org/10.1016/j.gfs.2019.100347.
[36] E. H. Peerzadah, I. Kaur, R. Banu, Multifaceted aspects of advanced innovations in engineering and technology, IJRECE, **7**(2), 2019, ISSN: 2348-2281 (online).
[37] D. Kember, C. Hong, V. W. Yau, S. A. Ho, "Mechanisms for promoting the development of cognitive, social and affective graduate attributes," Higher education, **74**(5), 799–814, 2017, doi:10.1007/s10734-016-0077-x.
[38] R. Moalosi, M. T. Oladiran, J. Uziak, "Students perspective on the attainment of graduate attributes through a design project," Global journal of engineering education, **14**(1), 40–46, 2012.
[39] E. Pitukhin, A. Varfolomeyev, A. Tulaeva, "Job advertisements analysis for curricula management: the competency approach," in 9th Annual International Conference of Education, Research and Innovation Proceedings, [Seville], 2026–2035, 2016, doi:10.21125/iceri.2016.1456.
[40] B. Oliver, T. Jorre de St Jorre, "Graduate attributes for 2020 and beyond: Recommendations for australian higher education providers," Higher Education Research & Development, **37**(4), 821–836, 2018, doi.org/10.1080/07294360.2018.1446415.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 80-90 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Multidisciplinary Systemic Methodology, for the Development of Middle-sized Cities. Case: Metropolitan Zone of Pachuca, Mexico

Montaño-Arango Oscar[1], Ortega-Reyes Antonio Oswaldo[*,1], Corona-Armenta José Ramón[1], Rivera-Gómez Héctor[1], Martínez-Muñoz Enrique[1], Robles-Acosta Carlos[2]

[1]*Autonomous University of the State of Hidalgo, Institute of Basic Sciences and Engineering, Academic Area of Engineering and Architecture, Hidalgo, 78556, Mexico*

[2]*Autonomous Mexico State University, Ecatepec University Center, Research and Postgraduate Coordination, Mexico*

ARTICLE INFO

Article history:
Received: 05 May, 2021
Accepted: 12 June, 2021
Online: 10 July, 2021

Keywords:
Systemic approach
Middle-sized city
Situational analysis
competitiveness
Model cities
Development

ABSTRACT

This paper analyzes the challenges a middle-sized city faces, particularly the Metropolitan Zone of Pachuca (MZP), which has not had the expected development given its current geographic, competitiveness, and crisis circumstances. The research proposes a systemic methodology for the analysis of the aforementioned Metropolitan Area. It emphasizes its internal and external behavior, the competitive environment of neighboring cities, conditions of middle-sized cities in Latin America, and the trends of model cities worldwide. The study identified the factors limiting their development and the competencies that can be used based on the opportunities, limitations, and perspectives. The result was a situational temporality matrix of the initiatives according to the impacts, which will specify the necessary characteristics to establishing strategies and the compatibility of the perspective of the Metropolitan Area of Pachuca for the transition towards the model of growth called New Urbanism.

1. Introduction

Nowadays, there are several questions about the capacity of middle-sized cities to face the phenomenon of globalization successfully, and the way they get inserted and participate in it. These peculiarities might reflect a difference in terms of competitiveness and the assessment of its meaning, to lay the basis for decision-making when public policy design and planning [1]; this enables to forge development and welfare in terms of the collective needs and the environment in which middle-sized cities interact [2]. Likewise, in [3], it is mentioned that these cities have advantages over large cities, which is why they should perform essential functions for territorial balance.

Cities are not only spaces for the production of goods and amenities, but places that generate knowledge, create new ideas that define new forms of social relationships. It is essential to understand their process of the conformation since it marks their daily life or, in some cases, subsistence, also the rate of growth, well-being, and progress. Their analysis, consequently, is not exclusive to a particular discipline. It is necessary to incorporate different approaches for their better understanding, because into them dwell more than 50% of the population in the world, and in the case of Mexico, more than three-quarters of the population [1], [4] and [5].

This research addresses an evaluation of the connotation of development in middle-sized cities through a systemic methodology. They demand urgent evolution, which considers their competitive environment and the formalization of instruments to articulate current elements of the territory. The challenge is even more significant when incorporating the best practices of the competitive mean due to the restrictions derived from weaknesses and threatens. Besides, there are local specifications which difficult the transition to modern cities. Therefore, it is important to plan their future according to the guidelines brought out by the study.

We present a temporary positional matrix with the highlights emanated from the SWOT and PESTEL analysis, composed of four quadrants where the initiatives point out priorities and

[*]Corresponding Author: Ortega-Reyes Antonio Oswaldo, Email: oswwaldoo@yahoo.com.mx

https://dx.doi.org/10.25046/aj060410

strategic lines to define an urban development project according to the temporary spatial horizons.

1.1. The concept of development as a guideline for the use of the territory

Faced with the new environment for the development and use of the territory of cities Sesmas [6], sets three main scenarios: contextual (integrated by the processes of globalization and decentralization), strategic (linked to a new organization and territorial management), and political (regarding a modern State, capable of territorial leadership, via the different policy instruments) (Figure 1).

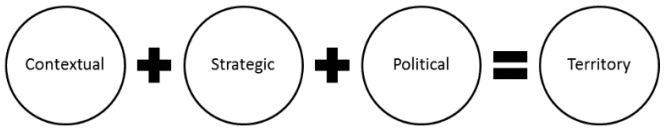


Figure 1: Scenarios that define the territory

Source: Author's elaboration based on Sesmas [6]

On the other hand, the concept of development is a frame of reference that gives the general guidelines that have an impact on the conformation of the territory and competitiveness. Economics was the first discipline to use this idea, almost as a synonym for economic growth. However, concerning Geography, it acquires social nuances that make it closer to the 'non-economic' needs of populations [7].

The word development comes from the Greek α νά πτυξη (anaptise) and means "unfold" or also "discover". Therefore, development is a set of potentialities that each social group possesses and must reveal. This etymological interpretation of the concept shows that progress and welfare of the population do not depend exclusively on external factors, but on latent endogenous potentialities which await to be "brazen" or "discovered" [7] and [8]. In this regard: "If ultimately we consider development as expanding the capacity of people to perform activities freely chosen and valued, it would be entirely inappropriate to exalt human beings as instruments of economic development" [9].

Therefore, in the search for development, the construction of utopias is essential since the transition might, at some point, go through one crisis [10]. In this way, a fundamental premise is to build a guideline that enables coexistence, socio-economic organization and competition of space, within a broad and global structural coherence.

1.2. Growth approaches for the development of cities worldwide

Since 1987, when the Brundtland Report established the concept of sustainability and sustainable as an adjective of development and, as part of the global lexicon [11], several tendencies arose about it and its focus on human settlements which allude to three approaches: smart growth, new urbanism, and ecological city.

Approaches to smart growth and new urbanism have become words of recognition incorporated in the United States (U.S.) into development planning goals and policies. Also, ecological cities have been less influential in the U.S. than the other two approaches [12]. However, in other parts of the world, this approach has received much attention to developing urban areas, particularly in Europe, Canada, Australia, and New Zealand [13] and, recently, Asia [14].

Smart growth represents an attempt to curb the expansion and its physical expression, ought to be integral and address issues such as protection of natural resources, diversity of homes - where the economic development depends on local capacity - and citizen participation [15].

Regarding New Urbanism, it is a design that, is oriented towards what represents a community architecture that is more humanized in scale and character [16] and [13] centering on tangible assets, urban landscapes, and design districts to improve the quality of life. It is composed of mixed uses, of a more compact configuration, a consistent and sensitive architecture to its place [17], abundant common open spaces (both: functional and natural) and, friendly as well as pedestrian-oriented inner circulation [18]. Multidisciplinarity is a relevant affair in New Urbanism in the U.S. [19]. Its processes include planners, developers, architects, engineers, government officials, investors, and community activists, as well as general stakeholders. Its goal is developing communities that do not exceed the limits of nature for livelihood, which is the load capacity. The gathering of these elements supports the concept of ecological cities. Thus, Ecocity Builders [20], - defined colloquially as Eco-city - in terms of land use policies, have the following objectives: to maximize urban density, reduce energy consumption, protect biodiversity, reduce travel distances and maximize options of transportation. Similarly, its principles are a way of giving shape and meaning to the concept of sustainability.

Based on the above, the following are crucial elements for the development of sustainable city planning since they allow carrying out result evaluations of the different policies implemented [21]:

- Level of urban competitiveness.
- Importance of services and industry with high degrees of innovation.
- Demographic changes
- Growth of middle-sized cities.
- Growth and, metropolitan concentration (sustainability).
- Decreasing urban-rural differentiation.
- Growth of the informal sector.
- Urban governance.
- Climate change and, city matters.

In [22], the author affirms the need for political will to create tracking systems based on precise indicators such as those in Dongtan, China or, Cambridge, England that have become icons of sustainable and competitive development; highlighting as well Stockholm, Sweden, and Hamburg, Germany, for being awarded in 2010 [23] and 2011 [24] respectively, for fulfilling the indicators of the European Green Capital Award.

1.3. Middle-sized cities

In recent decades, medium-sized towns have experienced a rapid sweep of spatial growth, changing their growth patterns in

terms of coverage, land use, and fragmentation of the urban landscape, where it is substantial to understand urban growth processes and their relationship with sustainability. By the year 2025, 13.6% of the population will live in megacities, and 42.4% in middle and small cities, which will require more resources for its operation, with the premise that currently, they have less technology and means to mitigate pollution and cope with urban dynamics, which attenuates their management effectiveness [25].

Medium-sized cities currently seem to constitute a suitable means that promote a development that adapts to new interpretations. On the other side, facing the risk of increasing gaps between large metropolitan and rural areas seem to be the ideal instrument for achieving more balanced development in the territory [26].

Thus, according to Ganau & Vilagrasa [27] and Brunet [28], intermediate cities present defining features: 1) they are non-metropolitan centers, but they have sufficient critical mass and the will to transform themselves and be well equipped; 2) they are nuclei that can act as intermediaries between the big city and rural spaces and 3) they might be capable of generating growth and development in their immediate surroundings and of balancing the territory against metropolitan macrocephaly. Bellet and Llop [29] point out that these can act as providers of specialized goods and services, as well as centers of social, economic, and cultural interaction for their environment. Local governments of middle-sized cities present a large number of interconnections with their surrounding territory and other cities. With the pressure for urbanization, which comes hand in hand with progress, affecting their peripheries and localized municipalities near the big cities [30], becoming the reason why the challenges of local governments in middle sized cities on urban planning are very complex [31].

1.4. Growth of Latin American medium cities

In the Latin American case, the growth of cities has been in cycles. Figure 2 shows how they changed from a very compact territorial body to a sectorial perimeter and from a polarized city to a fragmented one [32]. In the last phase, globalism had a vast influence on them, reflecting dramatic changes in their urban structure and development [33]. This circumstance made it necessary to expand the traditional model for urban development, establishing new phases [34].

In [35], the author refer that physically, growth in Latin American cities has been quite peculiar. In the middle of the '90s, its expansion was oil stain type. That is, in continuous extension. At present, most cities have adopted scattered growth patterns throughout the territory, generating an uncontrolled peri-urbanization [36] featuring this process -due to fast changes in the city- with lack of regulation.

New developments began in rural areas [37], where the value of the land was lower along with urbanization would generate more added value. But as it did not have any supervisory body other than the market itself, the quality concerning the new developments began to decline considerably, configuring new areas of cities with much more precarious standards than the previous ones. These new difficulties bring a profound process of crisis and transformation which, comes mainly from the necessity of adapting to new national economic and social conditions and also to the recent characteristics of urban development [38]. In this regard, [39], mention that regional development in Latin America has led to a vaster expansion and diversification of the system of cities because between 1950 and 2000, it went from 314 to 1851 towns with more than 20,000 populations. This more complex urban network forms a social and territorial base more prone to regional development, detecting that medium-sized cities (50,000 to 500,000 inhabitants) and small cities (20,000 to 50,000 inhabitants) are expanding rapidly in terms of nodal multiplication, which confirms the trend towards a more robust and complex urban system.

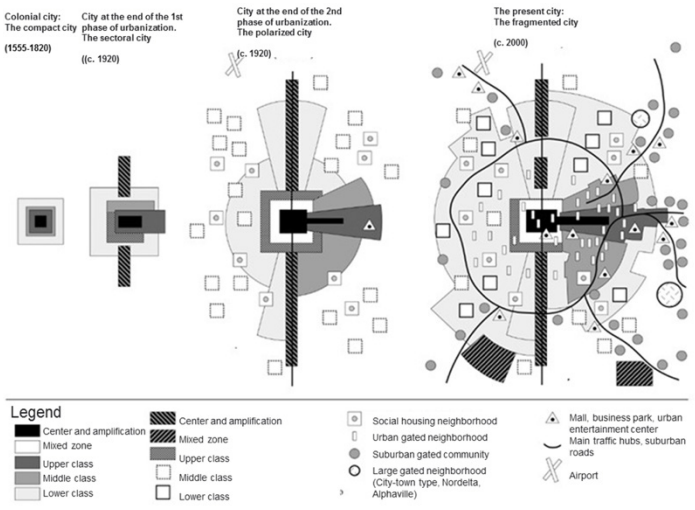


Figure 2: Transition of the development models of the Latin American City

Source: [33]

The development analyses reflect that the urbanization of cities allows greater prosperity in society (table 1).

Table 1: Level of urbanization and gross domestic product

Country	Level of urbanization 2010	GDP per cápita USD
Argentina	90,0	9.952
Bolivia	62,0	1.134
Brasil	87,0	4.375
Chile	89,0	6.248
Colombia	75,0	2.879
Ecuador	67,0	1.705
Guatemala	50,0	1.700
Haití	50,0	391
Jamaica (2007)	54,0	3.028
México	78,0	7.116
Perú	72,0	2.990
Venezuela	94,0	5.969

Source: [40]

In [41], the author mentions that middle-sized cities have the challenge of assuming a decisive role when it comes to redistributing better progress in the countries, which should have the function of articulating large cities and those of provincial rank, besides, to have a substantial impact on economic integration and territorial cohesion.

1.5. Growth of middle-sized Mexican cities

Mexican cities, as all the Latin American ones, have been characterized by having an accelerated growth of their urban areas, but not of urbanization. For many decades the new spaces incorporated into Mexican cities were occupied by their inhabitants with minimal services and infrastructure, and, in some cases, they were just nonexistent.

Each city solved this relevant gap of services and infrastructure over time according to their urbanization processes [42].

Thus, a centralist-productive structure propitiates changing migratory flows, which at the same time; delimit the short existence of middle-sized cities. Besides, policies to support middle-sized cities have been unsuccessful. The Mexican government created a priority stimulus for medium-sized and small towns to counteract the weightiness of Mexico City, Guadalajara, and Monterrey, but it did not provoke the expected territorial deconcentrating, demonstrating its failure. Although the number of medium-sized cities increased, it was not due to policies for a territorial reorganization of land, but growth logics [43].

In [44], the author points out that there is not enough information to analyze the development of medium-sized cities, so he proposes a scheme by growth zones through development time, identifying three "T" periods (figure 3):

(T1) Space developed until 1920, which is an urban space that represents the origin of the city. A regular layout of large properties, high construction density with predominantly commercial and service land uses. But, with some embedded sectors like housing.

(T2) 1950 to 1970. Space developed during periods of high rates of urban population growth in Mexico. At first, the area probably lacked infrastructure, but it incorporated it over time until achieving all the services and infrastructure.

(T3) 1990 to 2000. The features of this period are urban spaces with new subdivisions of low-income housing; warehouses, and industrial parks, large peripheral roads, and access to the city.

ET. Spaces in transition.

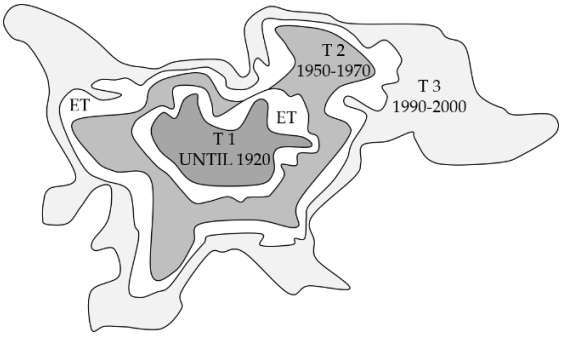


Figure 3. Characteristic periods of the urban space of Mexican middle-sized cities.

Source: [44]

In [45], the structural model of mid-sized cities in Mexico relies on the colonial urban layout (checkerboard shape), describing that the development phases occurred in the following way: Traditional city, Fragmented city, Regional Conurbation Metropolis. In this last phase, there is no spatial unit in the metropolitan area, where the urban elements were interwoven with each other and arranged on top of them.

Mexican cities must follow the policies established by three federal laws: The General Law of Human Settlements (LGAH), the Housing Law, and The General Law of Ecological Balance and Environmental Protection (LGEEPA). The LGAH delegates the responsibilities of design and urban planning to the municipalities, limiting itself to establishing procedures for development plans.

The outlook for the growth and development of cities for the coming years seems encouraging. That is due to the nascent policies and programs of the sector. However, it will depend on their proper implementation, coordination between the different political actors, federal and local institutions, as well as the continuity between the administrations, pointing to 1) the urban planning, 2) land use planning, 3) management, 4) implementation, 5) monitoring and control and, 6) improvement.

1.6. Metropolitan Zone of Pachuca (MZP)

The MZP is south of the state of Hidalgo. It is a middle-sized city in the range of 500,000 inhabitants and, it communicates with Mexico City through Federal Highway number 85. Table 2 shows the municipalities it comprises and figure 4 the delimitation of the Municipalities of the MZP.

Table 2. Characteristics of the municipalities that make up the MZP

Municipality	Population	Area (ha)	% Area
Epazoyucan	14,693	14,070	11.77
Mineral del Monte	14,640	5,339	4.47
Mineral de la Reforma	150,176	11,393	9.53
Pachuca de Soto	277,375	15,398	12.88
San Agustín Tlaxiaca	36,079	29,713	24.85
Zapotlán de Juárez	18,748	11,686	9.77
Zempoala	45382	31,962	26.73
Total	557,093	119,561	100.00

Source: Author's elaboration based on [46]

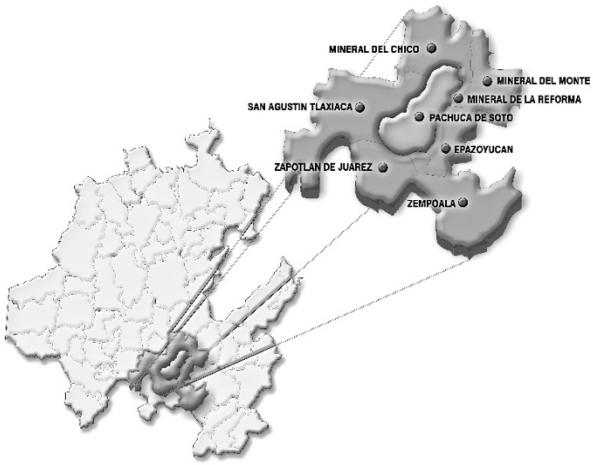


Figure 4: Delimitation of the Municipalities of the MZP [46]

According to the National Urban System [47], the level of urbanization of the state of Hidalgo is defined by the hierarchy, location, and degree of functional integration of each city, as shown in figure 5.

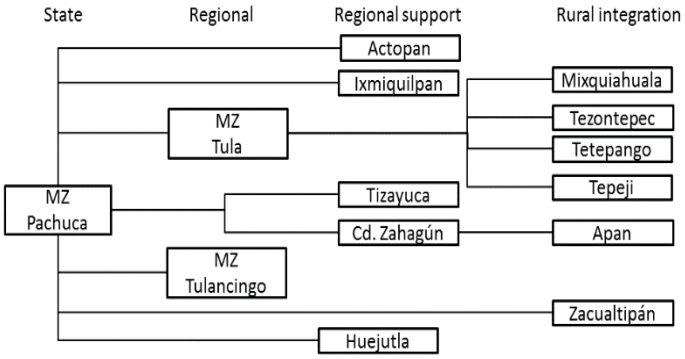


Figure 5: Urban System of the State of Hidalgo from the MZP

Source: Author's own elaboration based on [47]

1.6.1. Trends in urban expansion

The rapid growth of the MZP can be seen in the spread of its territory, indicating greater incorporation of hectares for urban purposes. In 2000, the conurbation area was 7,918 ha., which increased to 14,907 ha. in 2010. It is an increase of 6,989 ha. and represents an 88.2 % rise. The type of urban expansion manifests itself from two modalities:

A) Expansion by informal growth

Those made in irregular human settlements, which are in areas subject to natural hazards, for example: on steep slopes, in floodable areas, or vulnerable soils. On the other hand, the irregularity condition prevents them from having amenities, such as water, drainage, electricity, paving, garbage collection, lighting, surveillance, or adequate public spaces furthermore legal land property affairs. This situation contributes to the fact that their inhabitants present critical social lags, low school levels, poor health services, overcrowding, and deteriorated housing quality.

B) Expansion in disjointed housing complexes in the urban area continues.

Another type of recent urban expansion is the creation of housing facilities located in the urban periphery, characterized by the massive construction of housing, under one or two designs of predominant typology, whose promotion is through national institutional housing funds (Infonavit and Fovissste) or other organizations, to the eligible population. Although this housing estate has a better quality of construction and architectural settlements, many of them have problems in terms of the provision of public services such as regular water supply, public lighting, surveillance or -due to their remoteness-, public transport, which represents higher expenses for its inhabitants and a reduction in the quality of life. It does not imply that occasionally, some of these subdivisions are in areas subject to natural hazards. For example, in flood-prone zones, with consequent affectations for the inhabitants and the buildings they inhabit.

These growth patterns have different effects regarding municipal procurement:

a) They represent relevant increases in their population, with the consequent escalations in demand for public works, goods, and services.
b) They imply an increase in the maintenance and expansion of urban infrastructure and public facilities.
c) The lack of mechanisms for using the land and urban planning caused the habitation of areas at risk or productive zones. It translates into more work and more actions for their maintenance.
d) Irregular urban sprawl brings conflicts around the property, especially if there are no property titles or complete property regularization processes.
e) Municipal administrations face more limitations at getting resources when there are no up to date instruments for collecting local taxes related to property.

1.6.2. Land ownership regime

There are in the MZP a lot of communal agricultural centers that comprehend to 109.375 ha. several of these suburbs are near urban centers, where a big part of urban sprawl relies on vast extensions of minor lands known as "ejidos" which are a legal figure in the Mexican normativity that allows groups of small owners to use the land for no urban purposes:

A) The location of the "ejidos" where the proportion of parceled land is in the proximity of urban centers, represents possibilities of alienation for urban uses. Besides, there is some delay in the operation of public property records, which allows registering the "ejidos" as private property and not as collective goods, which generates irregular mechanisms for land buying and selling.
B) The territorial expansion of urban "ejido zones" and the lots destined for human settlements on communal lands may also constitute forms of conversion to urban land or, for the regularization of illegally established human communities [48].

1.6.3. Regulatory framework

The institutional action on urban planning derived from the publication in Mexico of the General Law on Human Settlements of the year 1976 and, the establishment and operation of the Ministry of Human Settlements and Public Works, promoted the development and publication of the State Program for Urban Development and Territorial Planning of the State of Hidalgo (PEDU and OTEH), as a pioneering initiative of the urban and regional planning system, undertaken in Mexico as part of the national policy on the territory. Its premises highlight the need for regeneration and exploitation of natural resources in the context of urbanization, the need for prevention and risk and vulnerability in cities (urban emergencies); to promote an adequate public administration for the urban development, the need for participation of the society on urban issues and, promoting financing for public works specifically in the municipal order.

Thus, the contributions of regulatory scope, stand on the adequacy of the public administration in the levels of municipalities and the state, regarding urban development, infrastructure to support the supply of energy and, the need to establish ways to promote the financing of municipal works. However, the dynamics and social and demographic structure offer a different picture, forcing to define specific policies to face enormous challenges such as the urbanization in the southern part of the state, where the MZP is.

The urbanization process of the Valley of Mexico in the State of Hidalgo is the most evident influence that obliges to rethink the territorial strategy. As a derivation of this reality, are critical situations such as connectivity, occupation of land without aptitude for urban development, and the consequent increase and sustained, long - term, costs of urbanization aside from considering the capacities and abilities that municipal governments must have to face their constitutional responsibilities in terms of development and freedom [49].

1.6.4. Cities within their competitive environment

The MZP competes for resources, investments, users' attraction, and infrastructure with the metropolis of Querétaro, Mexico, Puebla, and Tlaxcala. These cities are megalopolis - except for Tlaxcala-, which have grown better and have absorbed over time the small towns and surrounding municipalities. The expansion has been irregular because those cities have functioned as axes of development in the country, altogether with their spatial distribution of economic activities and population [50].

On the other hand, medium-sized cities have taken on the great importance, because they became escape valves for the growth of large cities, which, in the case of the MZP, has been forced to work as a bedroom city and has had to adapt to the policies and strategies of big cities. This brought the construction of large residential areas south of Pachuca City and the municipality of Mineral de la Reforma.

The MZP connects with the highway Arco Norte (North Arch) that communicates with the main cities of Central Mexico: Tlaxcala, Puebla, and Queretaro; and also allows to reach the port of Veracruz and, towards the north, with the most industrialized cities in the country.

The IMCO competitiveness index [51] ranks the villages where the MZP competes in the region as follows: 1. Mexico City; 3. City of Queretaro; 23. The towns of Puebla-Tlaxcala and, 47 City of Pachuca, which is considered a middle-sized city with a medium-low level of competitiveness. Best-rated cities stand out for their diversified economy for hosting large companies, having socially responsible companies; good quality urban services; good quality universities; use of financial services, airlines and, bus lines. Nevertheless, they face greater insecurity, high population density, higher costs for public services, and traffic congestions.

Therefore, this study's purpose was to characterize transitional strategies towards the structural basis for adopting the model known as New Urbanism. Its importance relies on identifying the main elements that allow doing a benchmarking with neighboring cities, which we did through the systemic approach.

2. Methods

According to [52], a systemic approach is necessary to study cities since they are complex adaptive systems. Meanwhile in [53], the author indicate that urban environments have socio-environmental variables that should be studied systemically. The Systems Approach considers the whole system in the search of means to achieve a goal and its choice [54] and [55]; it represents optimization and efficiency in a network of complex interactions within a dynamic environment, likewise to study cities. It relies on interdisciplinary and transdisciplinary studies. It also considers the Cybernetic approach, which is a resource of transdisciplinarity which serves to distinguish the two main subsystems at any system: the management or driver and, productive or conducted, comprising the fundamental relationships of these: of information and those of execution [56].

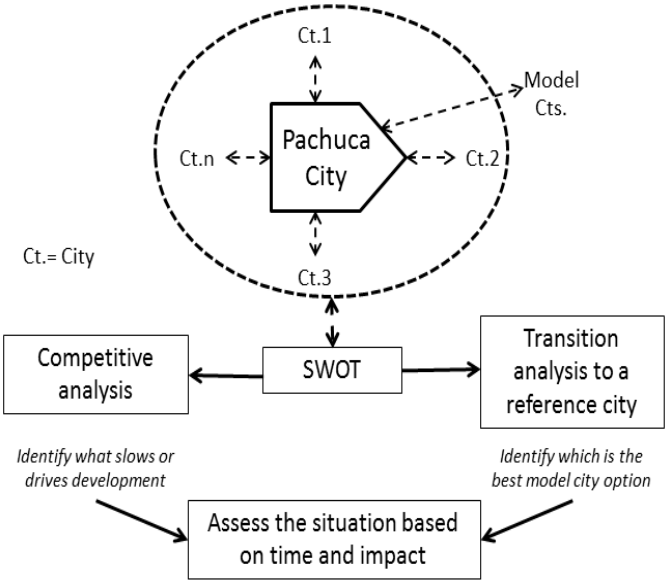


Figure 6. Multidisciplinary Systemic Methodology for the study of Middle-sized Cities
Source: Author's own elaboration (2021)

Thus, [57] describe that, in the dawn of the XXI Century, people, regions, and cities face new challenges that require interventions based on multidisciplinary methodologies which enable to project and try to control facts, because the challenges are too many and too powerful to let them happen randomly. The conceptualization of the context, the diagnosis, the planning, and the operation, result in this way, fundamental tools to achieve the sustainability of the development of the current and future regions and cities asides to ensure equity and participation of the regional society. Figure 6 shows the systemic methodological design of the study.

Table 3: Impact of MZP with its environment

Strengths (S)	Weaknesses (W)
1. Growing and articulation of the primary communication routes. 2. Equipment availability at different service levels. 3. Development of industrial parks and the promotion of innovation. 4. Stable political system. 5. Availability of natural, human, and cultural resources. 6. Sufficient supply at different educational levels. 7. Rating of the MZP as a safe zone.	1. Obsolete regulations to use the territory (territorial ordering) at the state, regional, and municipal levels. 2. Regulatory framework out of date of urban growth at different levels (national, regional, metropolitan, suburban, population, and areas of interest). 3. Growth over "ejido" zones and conflicts over land ownership and use. 4. Regional inequity in the provision of basic services. 5. Low quality of life. 6. Poor economic diversity. 7. Stable but reduced work market. 8. Poor management of political actors. 9. Insufficient inputs and suppliers.
Opportunities (O)	**Threats (T)**
10. Location near the Valley of Mexico. 11. Proximity to highways, to industrial and energy infrastructure corridors nationwide. 12. Location within the state of equipment and infrastructure of regional and national importance. 13. Application of national and international standards and best practices regarding development.	14. Territorial expansion of the metropolitan area of the Valley of Mexico. 15. Increased competitiveness of near metropolitan areas (State of Mexico, Queretaro, Tlaxcala, Puebla). 16. Better management capacity in other entities to expedite the location of large companies.

Source: Author's own elaboration (2021)

We identified and analyzed relevant facts to establish grounds for a general plan of development and growth. We took as a basis the framework of multifactorial analysis known as PESTEL to categorize the Political, Economic, Socio-cultural, Technological, Ecological, and Legal fields. We also used the SWOT competitiveness analysis to establish the Strengths (S), Weaknesses (W), Opportunities (O), and Threats (T) of the MZP, interrelating Strenghts-Opportunities (SO), Strenghts-Threatens (ST), Weaknesses (W) and Weaknesses-Opportunities (WO). Based on both, a temporal scheme by quadrants is proposed, to identify the impacts and times necessary in acting on competitive priorities, which facilitates their study to identify the possibilities and restrictions inherent to development from the perspective of New Urbanism.

We applied the systemic approach and multidisciplinary analysis for the methodological design. We validated instruments and collected information through the technique of transdisciplinary consultation with experts and random selection. We also used brainstorming and comparative analysis techniques to evaluate the competitiveness of the MZP and determine the transition path to the reference model city.

3. Results

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

Table 3 shows the cross-impact of MZP with its environment. To set the relevance of the impacts, the following values were used: 3 = High impact, 2 = Medium impact, 1 = Low impact and 0 = No impact.

Table 4 describes the results concerning the strategic direction pro-posed as a result of the analyzes carried out.

Table 4: Crossing of impacts SO, ST, WO and WT

	O1	**O2**	**O3**	**O4**	**T1**	**T2**	**T3**
S1	3	3	3	3	1	3	0
S2	2	3	3	1	1	2	0
S3	3	3	2	2	2	2	0
S4	3	2	2	2	3	2	2
S5	3	2	0	2	1	2	0
S6	1	1	0	2	1	2	0
S7	3	2	2	3	3	3	0
Sum	18	16	13	15	12	16	2
W1	3	0	2	2	3	3	3
W2	3	1	2	3	3	3	3
W3	1	1	1	2	3	1	0
W4	1	1	0	3	2	1	0
W5	0	0	0	2	2	1	0
W6	0	2	0	2	2	2	0
W7	1	0	0	2	2	1	0
W8	3	2	3	3	3	3	3
W9	2	2	1	2	1	3	2
Sum	14	9	9	21	21	19	11

Table 5: Impact analysis

	O	T
	SO 1 Possibilities of improving regional articulation around the Valley of Mexico and the North	ST 1 The development of industrial parks, political stability, and security of the area enables comp

S	Central and Gulf of Mexico regions. SO2 Possibilities of attracting large companies due to their proximity to communication routes, qualified workforce, and availability of infrastructure. SO4 Development of conditions to apply national and international benchmarks and best practices in the different fields of competitiveness, depending on the type of city under view.	anies to consider the MZP as an option for its establishment, which can attenuate the growth of the metropolitan area of Mexico City. ST2 The expansion of the network of communication routes in an environment of political stability and security makes the MZP more attractive compared to the metropolitan areas considered.
	WO 1 The adaptation of the regulatory framework and more effective government management will facilitate taking advantage of the proximity to the Metropolitan Zone of the Valley of Mexico (MZVM). WO4 More effective government management, the existing communication network, and the strategic location of the MZP concerning energy infrastructure will allow its promotion as a supplier of the industrial	WT 1 Through effective government management, it is necessary to adapt the regulatory framework that contemplates systemically, all factors of the territorial organization to expedite the permits for the establishment of companies. WT2 Through effective government management, it is necessary to adapt the regulatory framework that considers all the factors of territorial ordering and urban expansion; that

W	corridors of the central region of the country.	makes the MZP attractive for the establishment of companies compared to other areas. WT3 The region has not been favored by companies that have sought a place to establish themselves and have selected other alternatives. That is the reason why it is relevant to review the management, the policies adopted, the land regulation, adaptation of the regulatory framework, and the leadership of political actors.

Source: Author's own elaboration (2021)

From the analysis of tables 4 and 5, we highlight that the SO quadrant is the one with the highest impact (62 points); which indicates the ability to develop offensive strategies with the limits in ST (30 points) since there are no elements enough to counter threats from the competitive environment. Therefore, government management, the periods of government (which truncate the learning curve), the long-term planning (support targets), quality infrastructure, and the regulatory framework are the points that need a reorientation to detonate development and growth in the region.

For the PESTEL analysis we identified each of the perspectives and the level of impacts based on threats and opportunities, as shown in Table 6.

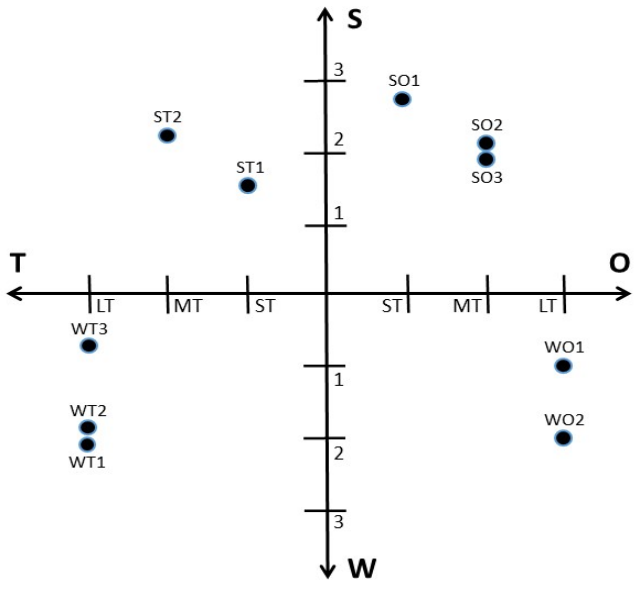
In this analysis we point out that the highest weaknesses occur in the legal and ecological aspects, with better positive effects on the economic and socio-cultural areas. However, the resulting graph shows a stronger trend in positive aspects, which favors the development possibilities of the MZP, based on its socio-cultural, technological, and economic advantages. Figure 8 shows the initiatives arising from these analyses, which consist of four quadrants, according to impacts and the degree of priority we identified.

Table 6: PESTEL

PERSPECTIVE	THREATS VN	THREATS N	B	OPPORTUNITIES P	OPPORTUNITIES VP
POLITICS					
– Stable political system					X
– Qualification of the MZP as a safe area.					X
– Mismanagement of political actors	X				
– Greater management capacity in other entities to speed up the location of large companies	X				
ECONOMIC					
– Growth and articulation of the primary communication pathways.				X	
– Location near the Valley of Mexico			X		
– Proximity to highway, industrial and energy infrastructure corridors nationwide				X	
– Application of national and international benchmarks and best practices for the conception of development					X
– Regional imbalance in the provision of basic services	X				
– Low quality of life	X				
– Little diversified economy		X			
– Stable but small working markets		X			
– Insufficient suppliers and inputs		X			
– Greater competitiveness of proximal metropolitan areas (State of Mexico, Querétaro, Tlaxcala, Puebla)	X				
SOCIO-CULTURAL					
– Availability of natural, human and cultural resources					X
– Sufficient supply at the different educational levels					X
TECHNOLOGICAL					
– Equipment availability at different service levels.				X	
– Development of industrial parks and promotion of innovation.					X
– Location within the state of equipment and infrastructure of regional and national importance					X
ECOLOGICAL					
– Growth over ejidal zones and conflicts over land ownership and use	X				
– Territorial expansion of the metropolitan area of the Valley of Mexico		X			
LEGAL					
– Obsolete regulations for the use of the territory (territorial ordering) at the state, regional, and municipal levels	X				
– Outdated regulatory framework for urban growth at different levels (state, regional, metropolitan, conurbation, population and areas of interest)	X				

Source: Own elaboration based on Prieto Piña, Damirya Danesa Analysis of environmental factors under the PESTEL and SWOT approach for the "FEQUIMA" business project: Web portal of machinery, equipment and tools in Brazil Sapienza Organizacional, vol. 2, no. 3, January-June, 2015, pp. 129-152 University of the Andes Mérida, Venezuela	VN	Very Negative
	N	Negative
	B	Balanced
	P	Positive
	VP	Very Positive

Source: Author's own elaboration (2021)



ST= Short term, MT= Medium term LT= Long term
Figure 8: Situational timing for SWOT analysis initiatives

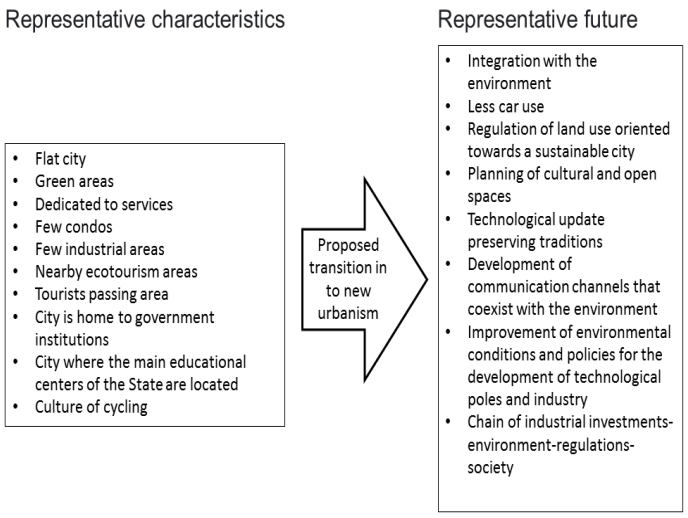


Figure 9: Proposed transition for the Metropolitan Zone of Pachuca (MZP)

Figure 9 shows the transitional projected strategic lines of action identified for purposes of valuation and relevancy to establish the guidelines and spatiotemporal horizons for the set up with the proposed urban development of the MZP. We should note that, in the analyzes, the SO quadrant is limited (3 opportunities to work) concerning the other three quadrants (7 distractions). It represents the requirement to direct and adequately manage the necessary resources to strategically counteract the threats and weaknesses we exposed in the study.

The connotation of development urgently requires evolution, considering the formalization of instruments capable of articulating current glimpse in the territory. However, the challenge is tougher when attempting to incorporate the best practices of the sphere because of restrictions on the weaknesses and threats. In addition to that, some peculiarities forecast urgent attention, for example, land tenure, provision of services, conceptualization, and design of the society of the MZP along with equipment and transportation, which are only part of the current joint.

The MZP is currently moving into modernization, so it is important to plan its future according to this study. A primary approach consists of moving into the Model of New Urbanism since its innocuous characteristics are necessary for this option, as exposed in figure 9.

4. Conclusions

The lack of tools to identify and set the trend in urban development and land use show the urgency to define ways to conceptualize and implement the target image of a middle-sized city as required in the MZP, considering international references in an adequate dimension, for their implementation at different levels of analysis, planning, and execution.

The international benchmarks provide alternatives to the conditions in the conformation of urban clusters that could be useful. However, each territory-region has its dynamics, which is in the function of the environment in which it competes. Therefore, it is necessary to establish the magnitude of the

implied factors, including their elements and environmental aspects that interact to reach the development.

The intense relationship and dependence that the MZP has with Mexico City, which is the largest in the country, represents a major challenge, where the management of the territory must obey a planning around urban development and land use, under a context of competitiveness, with a perspective towards sustainable development.

The planning of the MZP, as has been done so far, has led to a time lag in the face of a constantly changing territorial and urban reality, it is necessary that there be proper management to adapt to the dynamics of modern times.

The systemic treatment, -based on a methodology that involves the study area, referents of model cities, competitiveness of surrounding cities, treatment of feedback through the PESTEL and SWOT matrices-, gives a reasonable approach to the status and support for decision-making, which is linked to the situational temporality of the actions based on the analysis of the impacts, which is vital to project the growth and development of a middle-sized city.

Thus, the preceding, applied to the Metropolitan Area of Pachuca, allowed defining the characteristics necessary to transition to the New Urbanism Model.

As limits for future studies remain the changing environment of the MZP due to political trends in the different government levels and the inner dynamics of the mean, which depends on the features of the territory where it competes, implying its resources in the search for development.

Author Contributions

All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest

References

[1] E. Cabrero, Retos a la Competitividad Urbana en México. México: Centro de Investigación y Docencia Económicas (CIDE), 2013.

[2] IMCO, Índice de Competitividad Urbana 2010: Acciones Urgentes para las Ciudades del Futuro, Instituto Mexicano para la Competitividad A. C., 2010.

[3] M. Garrido, J. Rodríguez, E. López, "El papel de las ciudades medias de interior en el desarrollo regional. El caso de Andalucía", Boletín de la Asociación de Geógrafos Españoles, **71**, 375-395, 2016. doi: 10.21138/bage.2287

[4] G. Garza, M. Schteingart, Desarrollo urbano y regional. El Colegio de México, 2010.

[5] F. Carrión, La ciudad construida, urbanismo en América Latina. Facultad Latinoamericana de Ciencias Sociales (FLACSO), 2001.

[6] R. Sesmas, "Reseña-Crecimiento económico y desarrollo social: una contribución al estudio del territorio", Economía, Sociedad y Territorio, **11**(35), 265-271, 2011. doi: 10.22136/est002011127

[7] G. Berton, "Apreciaciones conceptuales del término desarrollo", Huellas, 13, 192-203, 2009.

[8] C. Torres, "Planeación y Desarrollo Territorial, Metodología para su diseño", Austral de Ciencias Sociales, **3**, 141-158, 1999. doi: 10.4206/rev.austral.cienc.soc.1999.n3-10

[9] A. Sen, Ética y desarrollo: la relación, BID, 1998.

[10] J. González et al, "La territorialización de la política pública en el proceso de gestión territorial como praxis para el desarrollo", Cuadernos de Desarrollo Rural, **10**(72), 243-265, 2013. doi: 10.11144/Javeriana.cdr10-72.tppp

[11] E. Minea, "Territorial Attractiveness – A Long-Term Issue for Public Policies", Juridical Current Journal, **58**(3), 101-110, 2014.

[12] T. Saunders, "Ecology and community design: lessons from Northern European ecological communities" Alternatives, **22**(2), 24-29, 1996.

[13] J. Edwards, M. Edwards, "How Possible is Sustainable Urban Development? An Analysis of Planners' Perceptions about New Urbanism, Smart Growth and the Ecological City", Planning Practice and Research, **25**(4), 417-437, 2010. doi: 10.1080/02697459.2010.511016

[14] ARUP, Cities. Shaping a better world by shaping cities, 2014.

[15] J. Gavinha, D. Sui, "Crecimiento inteligente-breve historia de un concepto de moda en Norteamérica", Scripta Nova, **7**(46), 39, 2003.

[16] D. Godschalk, "Land use planning challenges. Coping with Conflicts in Visions of Sustainable Development and Livable Communities", Journal of the American Planning Association, **70**(1), 5-13, 2004. doi.org/10.1080/01944360408976334

[17] P. Katz, The New Urbanism: Toward an Architecture of Community. McGraw-Hill, 1994.

[18] S. Wheeler, Planning for Sustainability, Routledge, 2013.

[19] Congress for the new urbanism, "Canons of Sustainable Architecture and Urbanism", 2011.

[20] Ecocity Builders, "Why eco-cities", 2014.

[21] E. Moreno, "Indicadores para el estudio de la sustentabilidad urbana en Chimalhuacán, Estado de México", Estudios Sociales, **22**(43), 161-186, 2013. doi.org/10.24836/es.v22i43.51

[22] L. Farias, El transporte público urbano bajo en carbono en América Latina. Innovación ambiental de servicios urbanos y de infraestructura: hacia una economía baja en carbono. CEPAL-Naciones Unidas, 2012.

[23] Eco-inteligencia, Estocolmo, referente de sostenibilidad, 2011.

[24] Comisión Europea, Ciudades del mañana-Retos, visiones y caminos a seguir. Unión Europea, 2011.

[25] C. Henríquez, Modelando el crecimiento de ciudades medias. Hacia un desarrollo urbano sustentable, Ediciones UC, 2014.

[26] J. Michelini, C. Davies, "Ciudades intermedias y desarrollo territorial; un análisis exploratorio del caso argentino", Documentos de Trabajo GEDEUR, **5**, 1-26, 2009.

[27] J. Ganau, J. Vilagrasa, "Ciudades medias en España: posición en la red urbana y procesos urbanos recientes", Mediterráneo Económico, **3**, 37-73, 2003.

[28] R. Brunet, "Des villes comme Lleida. Place et perspectives des villes moyennes en Europe", Bellet, C. y Llop, J. M. (comp), Ciudades Intermedias. Urbanización y sostenibilidad. Lleida, Milenio. 108-124, 2000.

[29] C. Bellet, J. Llop, "Miradas a otros espacios urbanos", Scripta Nova, **8**(165), 1- 30, 2004.

[30] C. Bellet, E. Olazabal, "Formas de crecimiento urbano de las ciudades medias españolas en las últimas décadas", Terr@ Plural, Ponta Groosa, **14**, 1-19, 2020. doi: 10.5212/TerraPlural.v.14.2013229.013

[31] G. Salazar, F. Irarrázaval, M. Fonck, "Ciudades intermedias y gobiernos locales: desfases escalares en la Región de La Araucanía, Chile", EURE, **43**(130), 161-184, 2017. doi.org/10.4067/s0250-71612017000300161

[32] A. Borsdorf, "Cómo modelar el desarrollo y la dinámica de la ciudad latinoamericana", EURE, **29**(86), 37-49, 2003. doi.org/10.4067/S0250-71612003008600002

[33] A. Borsdorf, R. Hidalgo, "Städtebauliche Megaprojekte im Umland lateinamerikanischer", Geographische Rundschau, 57(10), 30-39, 2005.

[34] J. Bähr, A. Borsdorf, "La ciudad latinoamericana. La construcción de un modelo Vigencia y perspectivas", Ciudad, urbanismo y paisaje, **2**(2), 207-222, 2005.

[35] F. Sabatini, G. Cáceres, J. Cerda, "Segregación residencial en las principales ciudades chilenas: tendencias de las tres últimas décadas y posibles cursos de acción", Eure, **27**(82), 21-42, 2001. doi.org/10.4067/S0250-71612001008200002

[36] C. De Mattos, "Globalización, negocios inmobiliarios y transformación urbana", Nueva Sociedad, **212**, 82-96, 2007.

[37] J. Hernández, B. Martínez, J. Méndez, "Reconfiguración territorial y estrategias de reproducción social en el periurbano poblano", Cuadernos de Desarrollo Rural, **2**(74), 13-34, 2014. doi:10.11144/javeriana.CRD11-74.rter

[38] O. Figueroa, "Transporte urbano y globalización. Políticas y efectos en América Latina", Eure, **31**(94), 41-53, 2005. http://dx.doi.org/10.4067/S0250-71612005009400003

[39] J. Da Cunha, J. Rodríguez, "Crecimiento urbano y movilidad en América Latina", Latinoamericana de Población, **3**(4-5), 27-64, 2009. doi: 10.31406/relap2009.v3.i1.n4-5.1
[40] ONU, Estado de las ciudades de América Latina y el Caribe. ONU-Habitat, 2010.
[41] B. Iglesias, "Las ciudades intermedias en la integración terriotiral del Sur Global", Revista CIDOB d´Afers Internacionals, **114**, 9-132, 2016. doi.org/10.24241/rcai.2016.114.3.109
[42] Fundación Idea, SIMO Consulting y Cámara de Senadores, México compacto: Las condiciones para la densificación urbana inteligente en México, 2014.
[43] J. Castillo, E. Patiño, "Ciudades medias", Elementos, Ciencia y Cultura, **6**(34), 29-33, 1999.
[44] G. Álvarez, "El crecimiento urbano y estructura urbana en las ciudades medias mexicanas" Quivera, **12**(2), 94-114, 2010.
[45] C. Göbel, "Una visión alemana de los modelos de ciudad. El caso de Querétaro", Gremium, **2**(4), 47-60, 2015.
[46] INEGI, Panorama sociodemográfico de Hidalgo 2015, Encuesta Intercensal. INEGI, 2015.
[47] SEDESOL, La expansión de las ciudades 1980-2010. México 135 ciudades. SEDESOL, 2012.
[48] A. Aguilar, I. Escamilla, Peri-urbanización y sustentabilidad en grandes ciudades, Porrúa 2011.
[49] E. Villegas, "Las Unidades de Planificación y Gestión Territorial como Directriz para la Zonificación Urbana", El Ágora-U.S.B., **14**(2), 551-581, 2014. doi.org/10.21500/16578031.67
[50] C. Pérez, "Expansión de la ciudad en la zona metropolitana de Pachuca: procesos desiguales y sujetos migrantes e inmobiliarios", Territorios, **38**, 41-65, 2018. doi.org/10.12804/revistas.urosario.edu.co/territorios/a.5577
[51] Índice de competitividad Urbana (IMCO), ¿Quién manda aquí?, IMCO, 2014.
[52] J. M. Fernández-Güell, Planificación estratégica de ciudades: Nuevos instrumentos y procesos 10, Reverté Segunda edición, 2019.
[53] A. Posada-Arrubla, Á. D. Paredes-Buitrago, G. E. Ortiz-Romero, "Enfoque sistémico aplicado al manejo de parques metropolitanos, una posición desde Bogotá DC-Colombia", Revista UDCA Actualidad & Divulgación Científica, **19**(1), 207-217, 2016.
[54] L. Von Bertalanffy, Teoría general de los sistemas. Limusa, 1996.
[55] J. P. Van Gigch, "General systems theory", Systems Research and Behavioral Science, **14**(2), 149-150, 1997.
[56] O. Gelman, G. Negroe, "Papel de la planeación en el proceso de conducción" Boletín IMPOS, Instituto Mexicano de Planeación y Operación de Sistemas, **11**(61), 1-17, 1981.
[57] A. Miguel, J. Torres, P. Maldonado, Fundamentos de la planeación urbano-regional. México; Instituto Municipal de Investigación y Planeación de Ensenada (IMIP), 2011.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 91-99 (2021)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

New Neural Networks for the Affinity Functions of Binary Images with Binary and Bipolar Components Determining

Valerii Dmitrienko, Serhii Leonov*, Aleksandr Zakovorotniy

Modeling, System Control and Artificial Intelligence Team, Computer Engineering and Programming Department, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, 61002, Ukraine

ARTICLE INFO

Article history:
Received: 01 December, 2020
Accepted: 22 June, 2021
Online: 10 July, 2021

Keywords:
Hamming neural network
Generalized architecture of the Hamming neural network
Maxnet network
Recognition and classification problems
Information technologies
Computer systems

ABSTRACT

The Hamming neural network is an effective tool for solving problems of recognition and classification of objects, the components of which are encoded using a binary bipolar alphabet, and as a measure of the objects' proximity the difference between the number of identical bipolar components which compared include objects and the Hamming distance between them are used. However, the Hamming neural network cannot be used to solve these problems if the input network object (image or vector) is at the same minimum distance from two or more reference objects, which are stored in the weights of the connections of the Hamming network neurons, and if the components of the compared vectors are encoded using a binary alphabet. It also cannot be used to assess the affinity (proximity) binary vectors using the functions of Jaccard, Sokal and Michener, Kulchitsky, etc. These source network Hamming disadvantages are overcome by improving the architecture and its operation algorithms. One of the disadvantages of discrete neural networks is that binary neural networks perceive the income data only when it's coded in binary or bipolar way. Thereby there is a specific apartness between computer systems based on the neural networks with different information coding. Therefore, developed neural network that is equally effective for any function of two kinds of coding information. This allows to eliminate the indicated disadvantage of the Hamming neural network and expand the scope of discrete neural networks application for solving problems of recognition and classification using proximity functions for discrete objects with binary coding of their components.

1. Introduction

Hamming distance R_X between arbitrary binary vectors $J_p = (j_{p1}, j_{p2}, \ldots, j_{pn})$, $J_q = (j_{q1}, j_{q2}, \ldots, j_{qn})$ is determined by the number of binary digits in which the compared vectors do not match [1]. This distance is often used to assess the proximity of discrete objects, which are described using binary alphabets with binary {0, 1} or bipolar {–1, 1} components. On the basis of bipolar neurons, a discrete Hamming neural network has been developed [1], which is successfully used to solve problems of recognition and the proximity measure of binary objects, the components of which are encoded using alphabet elements {–1, 1}, estimation. However, the evaluation of the measure of proximity of the compared objects (binary bipolar vectors) is carried out only by the most noticeable signs - different binary digits. At the same time, more "subtle" features for compared objects are ignored, which are often used to assess the similarity of binary vectors using the affinity (similarity) functions of Russell and Rao, Dice, Kulchitsky, Yule, etc.

In particular, in works [2–5] a brief description of these functions and their application is provided to compare different objects. In studies [2] they are used in the development of artificial immune systems, in works [3] and [4] they are used in the development of recognition systems for various objects. In this case, the characteristics of such objects are encoded by binary features. In work [5] given function is used to classify microorganisms.

In works [2–5], it is described and used not more than 15 commonly used functions. Although their total number is noticeably higher. So, in research in mathematical biology [6], 20 functions are used, and with hierarchical clustering, the set of random binary data (Hierarchical clustering Result of Random Binary Data Set) 76 different functions are indicated [7]. Once

*Corresponding Author: Serhii Leonov, Email: serleomail@gmail.com

again it indicates the relevance of the application of this approach for the classification and assessment of the proximity of various objects, which is described by a plurality of binary features.

2. Problem Statement

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

Analysis of recent publications on the theory and application of neural network Hamming [1 – 10] shows that the Hamming network has distinct disadvantages. In the architecture of the Hemming neural network, highly specialized layers of neurons are used: neuron layer calculating only the dot product of binary bipolar vectors; a neural network that allows you to select only one reference image from the network memory. This image is at the minimum distance from the input image in connection with this network

– It cannot work with the binary input data, when the measure of proximity of objects is estimated using finer characteristics than the Hamming distance. In particular, using the distances [2 – 5, 7];

– It cannot recognize objects, if they are at the same minimum distance from the two or more reference objects [1];

– cannot switch from processing information encoded using a bipolar alphabet to processing information presented using a binary alphabet and vice versa.

All this requires the development of a new Hemming neural network with a more advanced architecture.

When comparing objects with qualitative features encoded using a binary alphabet, for each pair of objects (binary vectors) $J_p = (j_{p1}, j_{p2}, \ldots, j_{pn})$, $J_q = (j_{q1}, j_{q2}, \ldots, j_{qn})$ in works [2 –5] four variable: a, b, f, g are used (table 1).

The variable a is used to count the number of bits with coinciding unit components of the vectors J_p and J_q. The variable b is needed to determine the number of binary digits of vectors J_p and J_q., in which both vectors have zero components, at the same time that means these objects do not have encoded with zeros in the same bits features.

The number of unit components that an object (vector) J_p has, but vector J_q. does not have, is determined using the variable f. The variable g, on the contrary, counts the number of unit components that the vector J_q. has, but are absent in the second vector J_p.

From the analysis of the Table. 1 it follows that an increase in the variable a unambiguously indicates an increase in the similarity (affinity) of the compared vectors. There is no such unambiguity with an increase in the variable b, since an increase in this variable can indicate both an increase in the similarity of the compared objects, and their belonging to different classes (if there is no similarity in the variable a). As for the measure of proximity of vectors J_p and J_q with respect to the variables f and g, it is symmetric with respect to these variables. An increase of these variables indicates the differences between the compared vectors increase.

Table 1: Variables for comparing binary vectors with binary components

	J_p	
J_q	1	0
1	$a = \sum_{k=1}^{n} j_{qk} j_{pk}$	$g = \sum_{k=1}^{n} (1 - j_{pk}) j_{qk}$
0	$f = \sum_{k=1}^{n} (1 - j_{qk}) j_{pk}$	$b = \sum_{k=1}^{n} (1 - j_{qk})(1 - j_{pk})$

Variables a, b, f and g from Table 1 are used in a number of well-known binary object similarity functions, in which the presence or absence of features is determined by the components of the binary alphabet. As an example, we can note the affinity (similarity) functions of Sokal and Michener, Jaccard and Needham, and so on. Using the variables f and g, one can easily determine the Hamming distance between binary vectors J_p and J_q

$$R_X(J_q, J_p) = f + g. \quad (1)$$

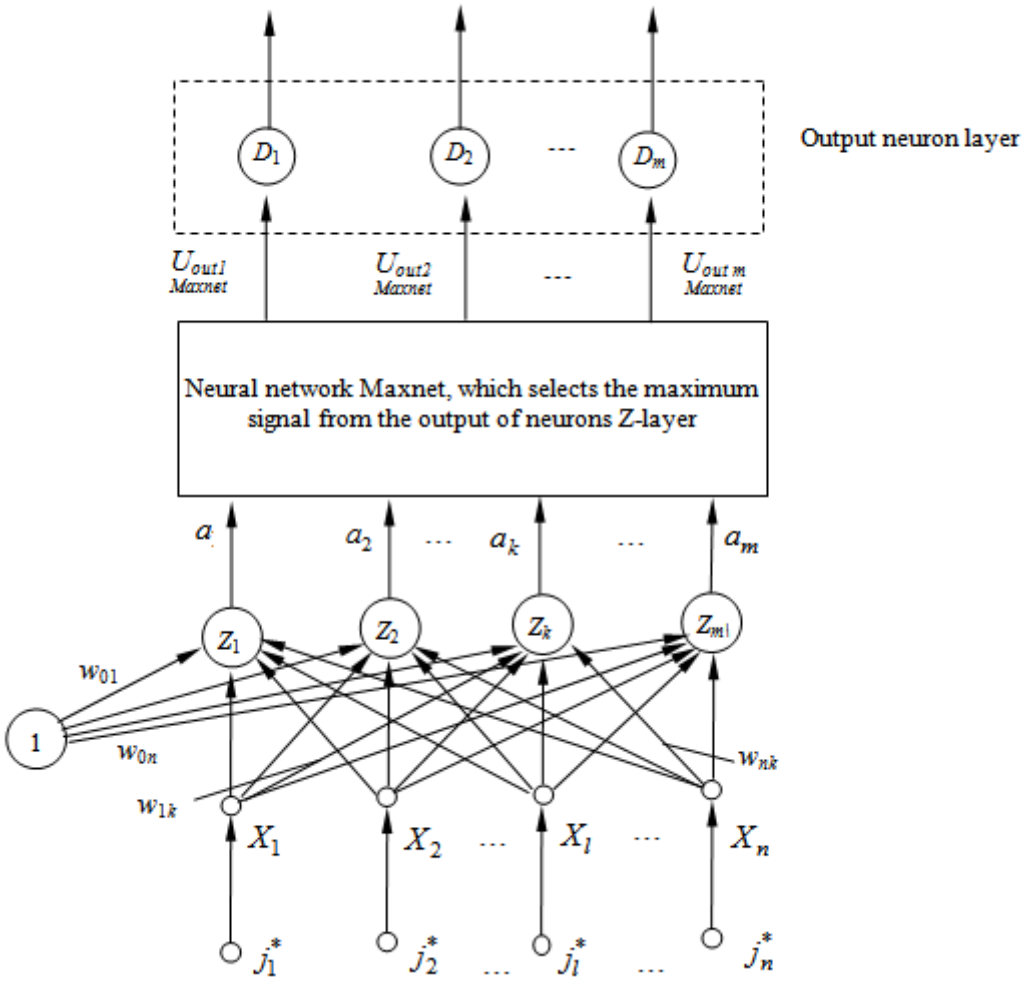
Variables a, b, f and g, as well as the affinity functions of Sokal, Michener, Jaccard, Needman, Hamming distance can be calculated using a new neural network based on binary neurons. The disadvantage of this neural network is the impossibility of processing information presented using the bipolar alphabet, as well as the inability to calculate the Hamming distance using the above affinity functions of Sokal and Michener, Jaccard, etc.

3. Generalized Block Diagram of a Hamming Neural Network

Since the theory, architecture and algorithms for the functioning of the Hamming network are described in detail in modern literature [1, 6 – 10], we will give only a generalized block diagram of the neural network (Figure 1) and a brief description of its functioning.

In the Hamming neural network, the input image $J^* = (j_1^*, j_2^*, \ldots, j_n^*)$ is fed to the inputs of X-neurons, which are multipliers of the bipolar image components J^*. Each Z-neuron stores one of the m reference images in the scales of its relations $J_k = (j_{k1}, j_{k2}, \ldots, j_{kn})$, $k = \overline{1, m}$ and compares this image with the input $J^* = (j_1^*, j_2^*, \ldots, j_n^*)$. As a measure of the similarity of the compared images (bipolar vectors), their scalar product is used [1, 8]

$$J_k J^* = \sum_{i=1}^{n} j_{ki} j_i^* = a_k - b_k, \quad (2)$$



where a_k and b_k are, respectively, the number of identical and different components of the compared vectors (images) J_k and J^*.

Since $a_k + b_k = n$, the scalar product (2) can be written in the form

$$J_k J^* = 2a_k - n\,. \tag{3}$$

From relations (2) and (3) it is easy to obtain

$$a_k = \frac{n}{2} + \frac{1}{2} J_k J^* = \frac{n}{2} + \frac{1}{2} \sum_{i=1}^{n} j_{ki} j_i^*\,. \tag{4}$$

Relation (4) can be regarded as a description of the input signal of a neuron Z_k (Figure 1), which has a displacement of $n/2$ and n inputs, that perceive the components of the input image $J^* = (j_1^*, j_2^*, \ldots, j_n^*)$ through neurons X_l ($l = \overline{1, n}$). In this case, the output signals of the X-layer elements repeat their input signals:

$$U_{outXl} = U_{inpXl} = j_l^*,\ l = \overline{1, n}, \tag{5}$$

where

$$U_{outXl} = \begin{cases} +1, & if \quad j_l^* = 1, \\ -1, & if \quad j_l^* = -1. \end{cases} \tag{6}$$

Each element of the X-layer is associated with the input of each neuron of the Z-layer. The scales of these relations contain information about m reference images $J_1, J_2, \ldots, J_m$, stored in the neural network. The scales of relations $w_{1k}, w_{2k}, \ldots, w_{nk}$ contain information about the k-th reference image J_k.

Z-layer elements compute their output signals

$$U_{outZk} = g_Z(U_{inp\ Z_k}) = a_k,\ k = \overline{1, m},$$

where g_Z – function of activation of neurons of the *Z*-layer; a_k – output element's Z_k signal, that arrives at the input of the neuron of the Maxnet network [1, 8]. From the output signals of the *Z*-layer elements, the Maxnet neural network, using an iterative process, extracts a single maximum signal greater than zero. Since the output signals of *D*-layer neurons are calculated by the ratio

$$U_{outDi} = \underset{\text{Maxnet}}{g_z(U_{out\,A_i})} = \begin{cases} 1, if \;\; U_{inpDi} > 0, \\ 0, if \;\; U_{inpDi} \le 0, \end{cases}, \; i = \overline{1, m}, \quad (7)$$

then at the output of the *Z*-layer of *D*-elements there will be only one single signal, which indicates which reference image is closest to the input vector J^*.

Analysis of the architecture of the discrete neural network shown in Figure 1 shows, that the layers of the network elements function relatively independently of each other. In particular, the layer of *Z*-neurons that determine the similarity functions of the input and reference images using the dot product can be replaced by blocks or neurons that calculate the similarity functions using other relations. In particular, the input bipolar signals and neurons of the *X*- and *Z*-layers can be replaced with binary.

The relative independence of the functioning of the main layers of the Hamming neural network from each other makes it possible to propose the use of binary neurons in the first layer of the network to work with binary input images (vectors). And also, to propose more "subtle" comparisons of input and stored in memory images using a variety of similarity (affinity) functions, modified using the Hamming distance (1) R_X.

The purpose of the article is to develop a generalized architecture of the Hemming neural network, which allows you to process input information that can be specified using both bipolar and binary alphabets. Also, based on the new architecture, the development of methods for solving problems of recognition and classification of binary objects is shown. In this case, the Hamming distance is used, as well as the functions of Jaccard, Sokal and Misher, Kulchinsky, etc.

4. Analytical Relationships Connecting Similarity Functions for Binary Objects and Hamming Distance

Since similarity functions for binary vectors (objects) and for objects described using the bipolar alphabet are used to assess the similarity of binary objects, then there should be dependencies connecting the similarity functions for binary objects (vectors) and the Hamming distance. Table 2 shows the aforementioned classical affinity functions (first column of Table 2) and similarity functions, modified using the Hamming distance (second column of Table 2).

Simple analytical relationships for determining the Hamming distance through the classic similarity functions Russel and Rao,

Table 2: Relationship between Hamming Distance and Similarity Functions for Binary Objects

Classic similarity functions	Similarity functions using Hamming distance	Hamming distance expressed in terms of binary similarity functions
Russell and Rao similarity function $S_1 = a/(a+b+f+g)$	$S_1 = \frac{a}{a+b+R_X}$	$R_X = \frac{a}{S_1} - (a+b)$
Sokal and Michener similarity function $S_2 = \frac{a+b}{a+b+f+g}$	$S_2 = \frac{a+b}{a+b+R_X}$	$R_X = \frac{a+b}{S_2} - (a+b)$
Jaccard and Needham similarity function $S_3 = a/(a+f+g)$	$S_3 = \frac{a}{a+R_X}$	$R_X = \frac{a}{S_3} - a$
Kulchitsky similarity function $S_4 = \frac{a}{f+g}$	$S_4 = \frac{a}{R_X}$	$R_X = \frac{a}{S_4}$
Dice similarity function $S_5 = a/(2a+f+g)$	$S_5 = \frac{a}{2a+R_X}$	$R_X = \frac{a}{S_5} - 2a$
Yule affinity function $S_6 = (ab-fg)/(ab+fg)$	Bulky expression	$R_X = f+g$
Correlation $S_7 = \frac{ab+fg}{[(a+f)(a+g)(b+g)(b+f)]^{1/2}}$	Bulky expression	$R_X = f+g$

Table 3: The Values of the Similarity Functions for the Compared Binary Vectors

	a	b	f	G	S_1	S_2	S_3	S_4	S_5	S_6	S_7	R_X
Values of similarity functions for vectors J_1 and J_k	5	5	2	0	5/12	10/12	5/7	5/2	5/12	1	5/√35	2
Values of similarity functions for equal vectors $J_1 = J_k$	7	5	0	0	7/12	1	1	∞	5/12	1	1	0
Numerical values of similarity functions for vectors J_1 and J_k with opposite components	0	0	7	5	0	0	0	0	0	– 1	1	12

Sokal and Michener, Jaccard and Needham, Kulchitsky, Dice were shown in the third column of the Table 2 for the first time ever. It was not possible to obtain simple analytical dependences connecting the Yule similarity function S_6 and the correlation coefficient S_7. But it is possible to determine the Hamming distance by functions S_6 and S_7 can using the relation (1).

Example. Let's perform the comparison using functions a, b, f and g, $S_1 - S_7$ some vector's pairs:

1) $J_1 = (111001110001)$, $J_k = (101001100001)$;

$J_1 = J_k = (111001110001)$;

$J_1 = (111001110001)$; $J_k = (000110001110)$.

The calculation results are shown in Table 3.

Analysis of the calculation results shown in Table 3 shows that the similarity functions $S_1 - S_5$ depend on the compared vectors. The values of these functions for vectors with opposite components take the minimum values, and for vectors with the same components, on the contrary, the maximum values.

If the compared vectors J_1 and J_k do not all coincide and not all are opposite in their components, then the similarity functions $S_1 - S_5$ take their values between the data of the second and third rows of Table 3, that is, between its maximum and minimum values. The function S6 practically obeys this rule, and the calculated values of which, given in the first and second lines, coincide.

As follows from the data Table 3 and the third column of the Table 2, the Hamming distance in comparison with the similarity functions $S_1 - S_5$ behaves in a certain sense opposite: using the functions $S_1 - S_6$ the similarity of the compared objects (vectors) is calculated, and using the Hamming distance, on the contrary, the difference of objects (vectors), therefore the Hamming distance R_X takes the maximum value for vectors that do not have any of the same components and the minimum distance (equal to zero) for vectors in which all components are the same.

The affinity functions are not limited to the ratios given in Table 2 [2 – 5]. A significant number of these similarity functions indicate the absence of one universal function suitable for comparing any binary objects with binary information encoding. The choice of one similarity function for solving a specific problem is determined, as a rule, by repeated modeling on the initial data in the space of features encoded using a binary alphabet.

5. New Neural Networks that Recognize Binary Images with Binary Components

The generalized block diagram of the Hamming neural network in Figure 1 clearly shows that the three main layers of a neural network can change and function almost independently of each other. In particular, instead of the Maxnet network, another neural network can be used that selects one or several identical maximum signals (if any). This allows you to find reference images that can be at the same minimum distance from the input. It also follows from the network architecture (Figure 1) that the layer calculating the measure of proximity of the input and reference images can calculate various measures of proximity. The relative independence of the main layers of the Hamming neural network makes it possible to propose neural networks for working with binary images (vectors) and using various affinity functions given in Table. 2, between the input and reference binary images. The Hamming distance can also be used (1). The classical architecture of the Hamming neural network [1, 6] for calculating the closeness measure of the input and reference vectors involves the use of the scalar product of two bipolar vectors [1, 8], one of which is a vector of connection scales, and the other is an input vector. In this case, the sum of the products of bipolar components does not contain any zero terms. If the scalar product of two binary vectors J_q and J_p is calculated in the classical Hamming neural network, then it is calculated as follows

$$J_q J_p = \sum_{i=1}^{n} j_{qi} j_{pi} = (a+b) - (f+g),$$

where variables (a, b, f, g) are defined by relations from Table 1.

Thus, if only one neuron is required to calculate the scalar product of two bipolar vectors using a classical neural network, then to calculate the scalar product of two binary vectors, you must first calculate the variables a, b, f and g using the relations given in Table 1.

In Figure 2 a block for calculating the variable b from Table 1 via neuron Z is shown:

$$b = \sum_{k=1}^{n} (1 - j_{qk})(1 - j_{pk}), \quad (8)$$

where $(1-j_{qk})$, $(k=\overline{1,n})$ – scales of neuron relations; $(1-j_{pk})$, $(k=\overline{1,n})$ – neuron Z input vector components; n – number of neuron relations scales Z.

With the help of neurons ∑1, ∑2, …, ∑n the components $(1-j_{p1}),, (1-j_{p2}),, \ldots, (1-j_{pn})$ of the input vector for neuron Z are computed. These components are accordingly communicated with the scales $(1-j_{q1})$, $(1-j_{q2})$, …, $(1-j_{qn})$, which allows us to calculate the variable b.

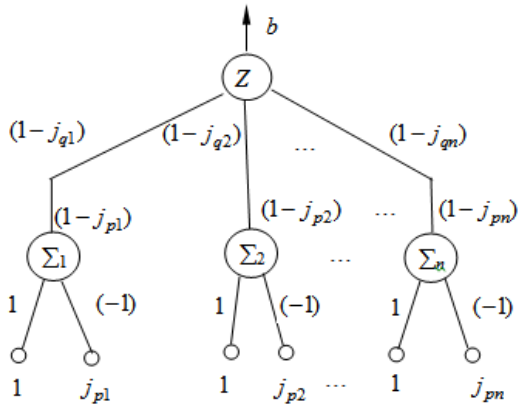


Figure 2: Neuron to calculate a variable b

If in Figure 2 to the inputs of neuron Z to apply input signals j_{p1}, j_{p2}, …, j_{pn} instead of signals calculated by summing elements $\sum_1$, $\sum_2$, …, $\sum_n$ of signals $(1-j_{p1})$, $(1-j_{p2})$, …, $(1-j_{pn})$, then neuron Z will calculate the variable f. To calculate the variable g (see Table 1) in Figure 2 it is necessary to replace the scale coefficients $(1-j_{p1})$, $(1-j_{p2})$, …, $(1-j_{pn})$ with the corresponding scale coefficients j_{p1}, j_{p2}, …, j_{pn}.

To obtain from the neural network shown in Figure 2, the neural network calculating the variable a needs to leave only one neuron Z with relation scales j_{q1}, j_{q2}, …, j_{qn}, to the inputs of which these j_{q1}, j_{q2}, …, j_{qn} signals, respectively have to be sent.

Having neurons or neural components for calculating the variables a, b, f and g, using table. 2 easy to get neural network architectures to compute any affinity function like Dice function S_5 (Figure 3).

The neural network for calculating the affinity function S_5 of the input vector and the reference vector is shown in Figure 3. The network has three layers of neurons. On the first layer of the network, the functions a, f and g are calculated. On the second layer, based on the calculated functions a, f and g, the sum $2a + f + g$ is calculated, which is used by the division unit to calculate the similarity function S_5.

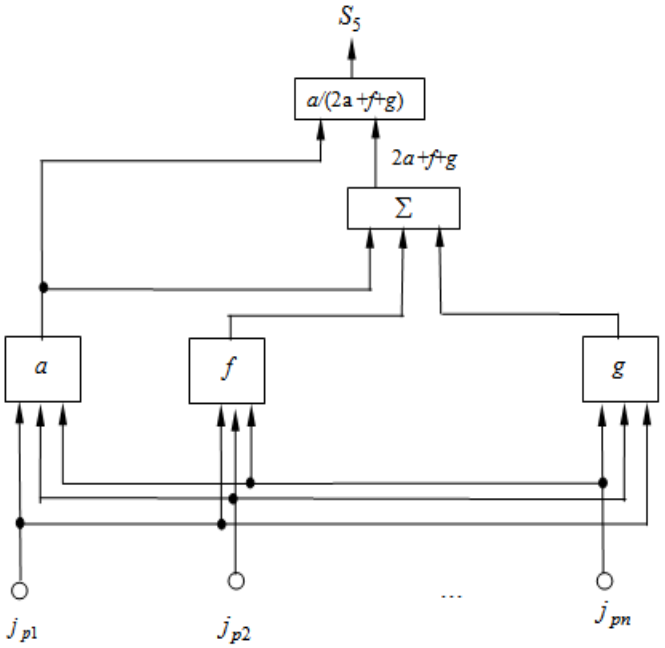


Figure 3: A neural network calculating the Dice function S_5

Similarly, neural network architecture can be obtained for calculating any other affinity functions. Figure 4 shows the architecture of a neural network that calculates the correlation coefficient S_7, the formula for calculating which is given in the very bottom line of the table. 2. With the help of this coefficient the similarity of the input vector and the reference vector, stored in the memory of the neural network, is estimated.

The neural network (Figure 4) for calculating the correlation coefficient has six layers of neurons. On the first layer, using blocks of neurons 1 – 4, functions a, b, f and g are calculated, which on the second layer are used to calculate intermediate variables ab, $(a + f)$ $(a + g)$, $(b + f)$, $(b + g)$, fg using blocks 5 – 10. On the third network layer (fig. 4) for calculating intermediate variables $(a + f)$ $(a + g)$ and $(b + f)$ $(b + g)$ the results of calculations, respectively, of blocks 6, 7, 8 and 9 are used. On the fourth network layer, the intermediate variables of the third network layer are used to calculate the radical expression in the denominator of the correlation S_7 coefficient formula (Table 2). The fifth layer of the neural network is required to determine the numerator and denominator of the correlation coefficient S_7, which is calculated on the last layer of the neural network.

5.1. Binary Neural Networks with the Transformation of Input InformationRrepresented Using the Bipolar Alphabet.

The architecture of neural networks for determining the Dice S_5 function and the correlation coefficient S_7, as well as other similarity functions, given in Table 2, in the general case can be represented in the following form (Figure 5).

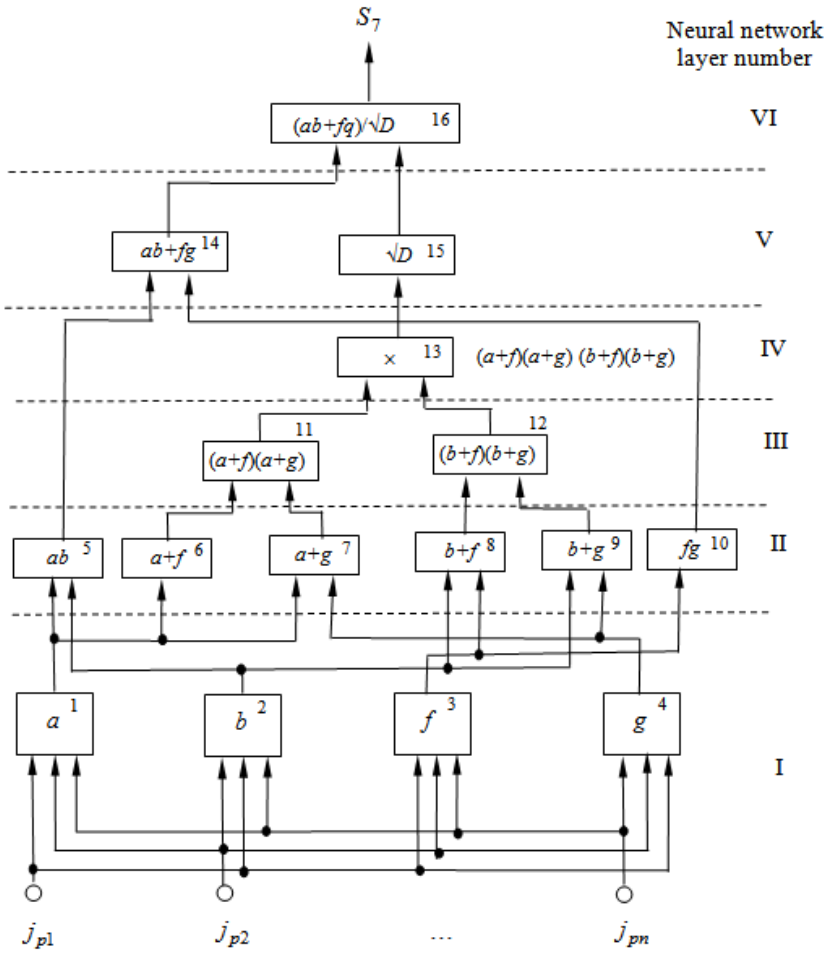


Figure 4: Neural network for calculating the correlation coefficient S_7 (the formula for calculating the coefficient S_7 is given in the bottom line of the Table. 2)

The neural network shown in Figure 5 has four layers of neurons. Both binary and bipolar signal vectors can be applied to the network input (*B*-neuron layer). If binary signals $j_{p1}^{(0,1)}$, $j_{p2}^{(0,1)}$, …, are fed to the input of the neural network, then all bipolar components of the vector ($j_{p1}^{(-1,1)}$, $j_{p2}^{(-1,1)}$, …, $j_{pn}^{(-1,1)}$) are equal to zero. If nonzero bipolar components are fed to the network input, then all components of the input vector ($j_{p1}^{(0,1)}$, $j_{p2}^{(0,1)}$, …, $j_{pn}^{(0,1)}$) are equal to zero. When specifying input information using a binary alphabet, the first three layers of neurons are used in the function. In this case, the input vector of binary signals ($j_{p1}^{(0,1)}$, $j_{p2}^{(0,1)}$, …, $j_{pn}^{(0,1)}$) arrives at the first inputs of neurons B_1, B_2, …, B_n. Second inputs of *B*-neuron layer are connected to the outputs of the neuron layer, which converts the input bipolar signals $j_{pk}^{(-1,1)}$, $k=\overline{1,n}$ to binary $j_{hk}^{(0,1)}$, $k=\overline{1,n}$.

Since the input information to the network defining the similarity function comes in binary form, the input vector ($j_{p1}^{(-1,1)}$, $j_{p2}^{(-1,1)}$, …, $j_{pn}^{(-1,1)}$) has to have all zero components. In this regard, zero signals $j_{h1}^{(0,1)} = j_{h2}^{(0,1)} = \ldots = j_{hn}^{(0,1)} = 0$ also appear at the output of layer 4 of the neural network. Therefore, the second inputs of neurons B_1, B_2, …, B_n will receive zero signals from the outputs of the elements of layer 4. *B*-layer neurons have activation functions of the form

$$U_{outBi} = \varphi(U_{inpBi}) = U_{inpBi} = j_{jhi}^{(0,1)},\ i=\overline{1,n}, \qquad (9)$$

where U_{outBi}, U_{inpBi} – respectively, the output and input signals of the neurons of the *B*-layer. In this regard, the output signals of the neurons of layer *B* repeat the input signals of the neural network $j_{p1}^{(0,1)}$, $j_{p2}^{(0,1)}$, …, $j_{pn}^{(0,1)}$ and are used to calculate the variables *a, b, f* and *g* on the second layer of the neural network, which are necessary to determine a specific similarity function S_d.

If the input information at the inputs of the neural network shown in Fig. 5, comes in the form of a bipolar vector $j_{p1}^{(-1,1)}$, $j_{p2}^{(-1,1)}$, …, $j_{pn}^{(-1,1)}$, then the neurons of layer 4 convert bipolar signals into binary according to the ratio

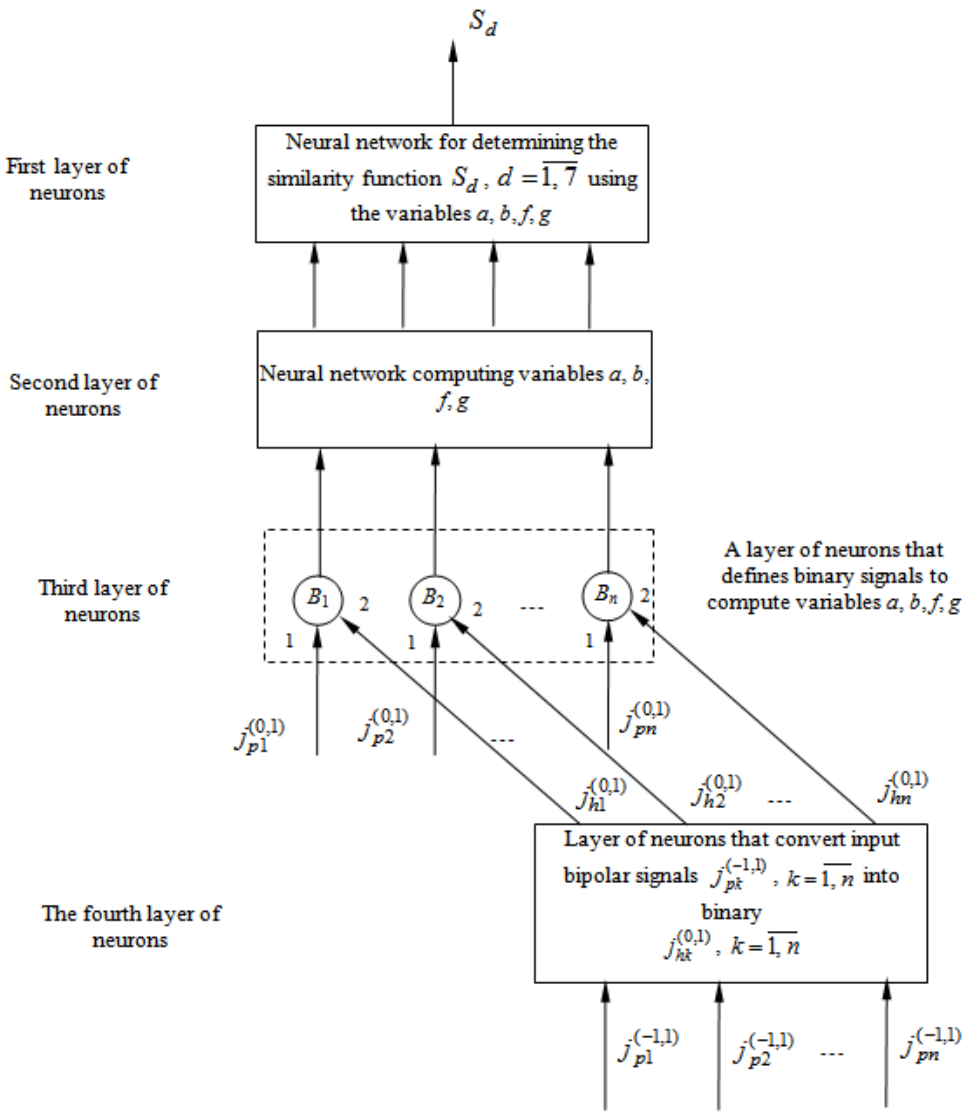


Figure 5: Architecture of neural networks for determining similarity functions S_1 – S_7 when the input information using binary or bipolar alphabets is specified

$$j_{hi}^{(0,1)} = \begin{cases} 1, \ if \ \ j_{pi}^{(-1,1)} = 1, \\ 0, \ if \ \ j_{pi}^{(-1,1)} = -1, \end{cases} \quad i = \overline{1,n}. \qquad (10)$$

The signals j_{h1}, j_{h2}, …, j_{hn} from the outputs of the neurons of layer 4 arrive at the second inputs of neurons $B_1, B_2, \ldots, B_n$, the first inputs of which receive zero signals $j_{p1}^{(0,1)}$, $j_{p2}^{(0,1)}$, …, $j_{pn}^{(0,1)}$, since the information necessary to determine the similarity function, in this case, is set in bipolar vector ($j_{p1}^{(-1,1)}$, $j_{p2}^{(-1,1)}$, …, $j_{pn}^{(-1,1)}$).

Thus, the architecture of the neural network shown in Figure 5, provides the computation of similarity functions for both binary and bipolar input signals. Obviously, a block of neurons can be synthesized in a similar way, which transforms input information from a binary alphabet into a bipolar one.

Similar blocks for transforming input information from one alphabet to another can be used at the inputs of any neural networks.

6. Conclusions

Binary and bipolar alphabets are used in various information technologies and methods for solving problems of recognition, classification, and estimation of the proximity of various binary objects. However, there is certain isolation in the application of methods based on different alphabets. The estimation of the measure of proximity of the compared bipolar vectors (objects, images) is mainly carried out using the dot product of vectors or Hamming distance. On their basis, the Hamming neural network has been developed and widely used. However, the Hamming neural network cannot be used when the recognition problem can have several solutions, to solve problems in cases where the components of binary vectors or images are encoded using a binary

alphabet. It cannot be used to estimate the affinity of binary vectors using the functions of Jaccard, Sokal and Michener, Kulchitsky, etc. In this regard, based on the generalization of the architecture of the Hamming network, new neural networks have been developed using binary coding of input information and the above similarity functions for binary objects. This expands the area of application of neural networks for solving problems of recognition and classification of input information using proximity functions using more "subtle" signs of proximity of discrete objects with binary coding. At the same time, when describing the general architecture of the generalized Hemming neural network, it is noted that the Maxnet network can be replaced by a neural network that allows you to define several solutions (if they exist). This allows, if necessary, to synthesize neural networks that allow obtaining a predetermined number of solutions in advance.

The idea of developing blocks for converting input information from one binary alphabet to another made it possible to propose not only a generalized Hamming network working with input information encoded using both binary and bipolar alphabets, but also modifications of any other binary neural networks that have worked so far with input information encoded with just one binary alphabet.

Thus, the scientific novelty of the results obtained is as follows.

For the first time, on the basis of the Hamming neural network, is proposed the concept of constructing new neural networks for solving problems of classification and recognition of binary objects using different distances and proximity functions. In doing so, the following principles are used:

1) adaptation of the first layer of neurons that determine the similarity functions of the input and reference images depending on the applied distances. This has allowed to develop a unified approach to the synthesis of neural networks that use different similarity function;

2) replacement Maxnet neural network in a neural network that can allocate one or more identical maximum signal. This allows you to select one or more reference images that are at the same minimum distance from the input image.

Also for the first time obtained the architecture and algorithms of functioning the neural network that can process the input information provided by both the bipolar and in binary form.

In real conditions some uncertainties while specific features comparison can be arisen with the binary objects' juxtaposition. For example, while the fruits or plants juxtaposition the color can be changed in different time intervals. In this case there could be uncertainties, that will be described with the third truth state.

Wherein, the third truth value could be interpretated in different ways [11 – 14]:

– like the intermediate value between truth and false;

– like the lack of information;

– like a specific paradoxical or even meaningless value, that gives such a meaningless value too;

– like values of three-valued paraconsistent logics, in which the third truth value can be interpreted in some statements as both true and false at the same time;

– like uncertainty of the three-valued logics, that describes quantum mechanics interpretations, etc.

The theory of making solutions' development in conditions of binary objects is a promising direction in our opinion.

Conflict of Interest

There is no conflict of interest reported between the authors.

References

[1] V.D. Dmitrienko et al., Neural networks: architecture, algorithms and usage. Tutorial, Kharkiv, NTU "Khpi", 222, 2020.

[2] S.A. Babichev, Theoretical and practical principles of information technology for processing gene expression profiles for gene network reconstruction. Dissertation for the degree of Doctor of Technical Sciences in specialty 05.13.06 – Information Technology, Kherson, Kherson National Technical University, 382, 2018.

[3] V.D. Dmitrienko et al., Methods and algorithms of artificial intelligence systems, Kiev, Kafedra, 282, 2014.

[4] A. For, Perception and pattern recognition, Moscow, Engineering, 272, 1989.

[5] R.S. Michalski A recent advance in data analysis: clustering jbjects into classes characterized by conjunctive concepts, Invited chapter in the book Progress in Pattern Recognition. **1**, North-Holland Publishing Company, Amsterdam-NewYork-Oxford, 33-49, 1981.

[6] Z. Chay et al., "Russel and Rao Coefficient is a Suitable Substitute for Dice Coefficient in Studying Restriction Mapped Genetic Distances of Escherichia colis". – iConcept Pocket Journal Series: Computational and Mathematical Biology, January 2010.

[7] S. Choi et al., "A Survey of Binary Similarity and Distance Measures", Systemics, Cybernetics and Informatics, **8**(1), 43-48, 2010.

[8] V.D. Dmitrienko et al., "Neural networks for determining affinity functions", 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 1-5, 2020, doi: 10.1109/HORA49412.2020.9152830.

[9] L.S. Yampolsky, Neurotechnology and neurosystems, Kiev, Monograph, "Dorado-Printing", 508, 2015.

[10] L. Fausett, Fundamentals of Neural Networks: Architectures, Algorithms and Applications, New Jersey: Prientice Hall. Int., 483, 2006.

[11] N. Tomova, "The emergence of three-valued logics: logical and philosophical analysis", Bulletin of Moscow University. Series 7. Philosophy, Moskow, MSU, 68-74, 2009.

[12] A.S. Karpenko, Development of multi-valued logic, Ed. 3-ht, Moskow, Publishing house LCI, 448, 2010.

[13] A.S. Karpenko, "Bochvar's three-valued logic and literal paralogics", RAS, Institute of Philosophy, Moskow, IF RAS, 110, 2016.

[14] N.E. Tomova, "Natural three-valued logocs and classical logic", Logical Investigations, **19**, 344-352, 2013.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 100-109 (2021)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Web-based Remote Lab System for Instrumentation and Electronic Learning

Jose María Sierra-Fernández*, Agustin Agüera-Pérez, Jose Carlos Palomares-Salas, Manuel Jesús Espinosa-Gavira, Olivia Florancias-Oliveros, Juan José González de la Rosa

Research group PAIDI-TIC-168, Electronics, University of Cádiz, Algeciras, 11202, Spain

ARTICLE INFO

Article history:
Received: 02 December, 2020
Accepted: 15 April, 2021
Online: 10 July, 2021

Keywords:
Remote lab
Web Interface
Flex design
Minimal network privileges
Scalable

ABSTRACT

Lab sessions in Engineering Education are designed to reinforce theoretical concepts. However, there is usually not enough time to reinforce all of them. Remote and virtual labs give students more time to reinforce those concepts. In particular, with remote labs, this can be done interacting with real lab instruments and specific configurations. This work proposes a flexible configuration for Remote Lab Sessions, based on some of 2019 most popular programming languages (Python and JavaScript). This configuration needs minimal network privileges, it is easy to scale and reconfigure. Its structure is based on a unique Reception-Server (which hosts students database, and Time Shift Manager, it is accessible from the internet, and connects students with Instruments-Servers) and some Instrument-Servers (which manage hardware connection and host experiences). students always connect to the Reception-Server, and book a time slot for an experience. During this time slot, User is internally forwarded to Instrument-Server associated with the selected experience, so User is still connected to the Reception-Serer. In this way, Reception-Server acts as a firewall, protecting Instrument-Servers, which never are open to the internet. A triple evaluation system is implemented, user session logging with auto-evaluation (objectives accomplished), a knowledge test and an interaction survey. An example experience is implemented, controlling a DC source using Standard Commands for Programmable Instruments. This is an example regarding how systems enable students to interact with hardware, giving the opportunity of understand real behaviour.

1. Introduction

In Engineering learning, Lab sessions act as a reinforcement of theoretical concepts, taking a really important role in student education, e.g. manipulate diodes in lab, examine threshold and only positive conduction, gives an additional level of knowledge, increasing student interest and making putting in order some of theory studied. However, there is usually not enough time to put in practice all theoretical concepts, due to time limitations to access to physical lab sessions. In this line, more experiences can be offered, expanding lab access time to 24/7, this is Remote Labs.

Sometimes, Remote Lab concept is put together with Virtual Lab (a simulated experience). When these two options want to be compared, few concepts should be considered, as stated in [1]. There, cost and effect over learning are studied. Conclusions of that work postulate that the increase of the cost of Remote Lab is compensated by the improvement in the learning process, due to students are more involved when they feel that they are interacting with real equipment.

Remote labs allow students to access expensive lab Instruments, and specific configuration, any time, in order to put in practice concepts studied in theoretical lessons. Remote access to labs instruments is gaining attention in order to grant access to students, which cannot be there (medical problems, case of force majeure, or other excused absence). Nowadays, at 2020 globally outbreak of COVID-19 is one of those cases of force majeure, which has turn into remote as more activities as possible.

First option for Remote Labs is to use a commercial equipment, as VISIR, as stated in [2], [3], which is a matrix of connections and components, where student can change connections. This same philosophy can be applied to matrix connections to servers, for networking lab [4]. In other situations, Remote labs are mixed with virtual labs, an example is Easy Java Simulations (EJS) as stated in [5]. This system, designed for simulated experiences

*Corresponding Author: Jose María Sierra-Fernández, Email: josemaria.sierra@gm.uca.es

https://dx.doi.org/10.25046/aj060412

(virtual labs), can be connected with hardware (oscilloscopes, engines, sensors, etc.) to perform an interactive experience.

There are too non-commercial systems in Remote Labs, e.g., as stated in [6], [7], where Labicom, a complete new system was designed. Server, client, solutions, all ad-hoc solutions are designed.

Even a compact solution has been developed, all integrated in a Raspberry Pi, as stated in [8], hosting a webserver, with an Arduino as sensor interface.

Other implementations, as the stated in [9], proposed a complete system for flexible Remote Lab testing, based in Lab Server and Web Server. It is a good approach for distributed systems, due to all communications are done throw the internet. However, it requires for each Lab Server, external access, if it is in a corporate network, it is a complex, or impossible configuration.

Even there are different approaches for Remote Labs, centred in collaborative work, as stated in [10]. This work is centred in a system where few students can work together in the same experience.

With a general study of the work, as stated in [11], it can be observed that these solutions are only a part of the problem, and it is demonstrated that a system structure stable, flexible and scalable is needed.

However, all reviewed Remote Labs systems are not enough flexible, due to hardware compatibility are limited, or web interface are not enough intuitive. The aim of this work is to design a structure and function of a Remote Lab system for engineering learning, flexible, in order to be able to connect any Instrument (with computer interface), scalable (in order to add nodes) and with the required data security level.

With that objective, servers are designed using the programming language Python, one of the most popular programming languages with several libraries, which allows interacting with almost any device. In addition, it can be set up as a web server, providing sufficient level of security and simplifying and ensuring data storage. The user interface is designed in HTML and JavaScript, introducing a fluid and asynchronous experience.

This work is structured as follows: in section II general lines and objectives of this project are given; then in section III the system structure is explained, followed in IV, where the relation with learning and examples are explained, the evaluation methods for learning outcomes are explained in section V, finally conclusions are given in section VI.

2. System Objective

The aim of this work is to present a complete Remote Lab solution which can implement experiences related with engineering learning. This system must manage user login, and user data, guaranteeing their privacy, and must have a time shift manager, in order to create time slots for remote experiences, allowing users to book them. In addition, networks requirements needed for the implementation in a complex network must be minimal, due to University networks are usually really complex.

This proposal must allow students the interaction with Lab Equipment from anywhere (from the internet). For that, only one Student can use each lab site at time. With this aim, time-slots are set, in order to coordinate students access. With this, Lab Instruments can be used beyond the Lab sessions hours.

2.1. Basic System Design

Experiences offered in remote experinces should be addequated to lessons. So it should be configurable, with the posibility to have some experiences in the same lab site, allowing to offer multiple remote experiences with the same equipments.

In reation with network requirements, learning labs use to be in complex networks, wich any special requiremnt (as direct access to a computer from the internet) is a risk for all the network. For that, only one system would have direct acces from the internet. This single system would manage user data, login and book system, at the time proxy conection to lab sites. Systems connected to lab equipments are in internall network, so it does no suposse a risk to the network, and only requiere a static IP.

User (student) connects to this single computer, for book a experince, in a specific time slot, for take the experince, or for take the evaluation excercises. All user interaction is done via this computer (connections, forwardings, data transmissions, or other procedures to connect user with remote lab station), so user seem to interact always with the same server, in the same web-site.

Aim of this remote lab system is to reinforce of theoretical knowledge, so test learning result is one fo the most important things in the process. The most dificult thing is to know if the experience has been done as designed (without any interaction problem). Interaction problems cause difficulties in perform laboratory practice and difficulty in acquiring knowledge, and aspect as camera quality is essential for a propper interaction. In order to know if some kind of problem in experince design (interaction site, software conectors, conection speed, camera quality, camera ilumination, etc.) could cause problems at taking the experience, a fast survey must be done for every user after take an experience.

Moreover, knowledge is evaluated in a multiple way. Examinig the experince itself (steps accoplished, time taken, total time in finish experince, tries for step, etc.) and an auto-evaluation is geenrated. And after the experince, a knowledge test must be take by user. This test include the knowledges concepts included in the experince, in order to confirm that experience has helped in reinforce that knowledge.

For a fair use of the systen, a limit at booking time slots must be considered. Deppendig on the number of users and time slots availables for take each experince, more or less conditions should be applied. It is recommended a limit of only one time slot booked by user, but sometimes, it should be considered additional limts as one try for each experince by user (at least until all users have done that experince). Time slots are created, and number of remote lab stations are limmited. So this limits are really importat in some situations.

Time slot sould not be the same for all experinces, and even some experinces can be done simultaneously, if both use different equipments connected to the same remote lab station. Roles for

that situations are implemented in remote lab stations. Time slots are calculated according to expriences time length, taking in consideration that some time is needed among time slots for intialize instruments.

Remote lab stations have instruments grouped by experinces compatibility, this is a set of instruments that can be used for many experiences. They are designed as much flexible as possible, but with just the needed instruments, in order to create as much as remote lab stations as possible. This is, a remote lab station can be done with a controlable DC source, a Fuction generator, a Multimeter and a Oscilloscope. These equipments can be used for a huge set of experiences in the areas of electronics (analog and digital). However, many experiences can be done only with a DC controlable source and mutimeter, and many others just with a function generatorn an oscilloscope and a fix DC source. So with almost same Lab Instruments, two lab stations can be done (for a lower set of experiences) or one (for a higher set of experinces, including more complex ones). Deppending on the kind of experiences wich want to be done, and the number of users, Instruments are gruped in one way or another.

With that, a clear idea of the system has been given, now a description of the system implementation is going to be done.

3. System Structure

Proposed remote lab system structure is organized around two types of servers, as explained before. One single server, which manage the whole system, and one server per remote lab station. All of then must have implemented a web-server, for enable the user web interaction, in addition with the application programming interface (API) interaction. This is a machine-machine interaction, via web-request. For this propose, servers need a back-end programming, and web interface needs a front-end design and programming.

For the back-end (all procedures and services done by the server, hidden by the user, as web server itself, camera stream, user login, data management, etc.), Python language have been selected. This is a really extended and supported language, in continuous development, with many packages, which gives many extra functionalities to it. In particular, a framework for web-server creation have been selected, Flask, which manage user login, connections, interactions, and more. This will be explained in more detail in both server types description.

For front-end (user interaction experience, including visualization and data interchange with backend), in web sites HTML and CSS are used for design appearance of websites, and JavaScript is used for introduce “programming” in web sites. While HTML and CSS design the content and structure, this design is static, and only can change at web site refresh. JavaScript allow elements resize and reorganize at window size change, update data without refresh all the website, even process data inside the browser, in the user computer.

Moreover, Python and JavaScript are the most popular languages in 2019 as stated in [12], [13], which grant many support and community. A schema about languages used are shown in Figure 1.

3.1. Instrument-Server

First, remote lab station will be examined. These stations are composed of a set of instruments, components and connections, and a server, which manage that hardware. That server is named Instrument-Server, and host remote lab experiences and manage the communication with instruments, using a module named Instrument-connector.

Instrument-Server is a web-server with many functions, with the back-end designed in Python, relation with hardware is done in an easy way with calls to Instrument-Connector. This module involves all procedures to communicate with hardware. Python has libraries for exchange data via GPIB, USB, and LAN, using VISA protocols, RS-232, and many others, with them, and Instrument documentation, a set of functions are designed for each Instrument, for access to all needed functions. Sometimes, procedures would be a single line, others a set of command exchange. Finally, from Instrument-Server, Lab Instrument interaction is done as a single call to Instrument-Connector.

The specific structure of this module depends of the specific hardware connected, but if we have a GPIB interaction with a Multimeter, from Instrument-Server there would be an instruction “read_voltage()”. Inside Instrument-connector, in the address selected for the instrument, some order are sent (mode voltmeter, range-auto, read) and then return is taken from GPIB bus from the address of the Multimeter. All this sequence, with time among them, watchdogs, and other specifications of GPIB, and error handle, are managed by the method read_voltage()”, which is simply called from the Instrument-Server and returns the voltage.

In relation with data, Flask gives a package, called, Flask-SQLAlchemy, which allows the use of Object-Relational mapping (ORM) of databases. This is defining a data set as an object (a programming class), linked with a database. Schema for tables, interaction instructions, or other database operations are hidden by the programmer, and a programming object is manipulated. Moreover, instructions are the same for different database types, as SQLite, PostgreSQL, MySQL, Oracle, MS-SQL, Firebird, Sybase and others. That gives freedom to not only change database server, even change database technology, without changing code.

In order to provide beautiful, flexible and configurable front-end, a HTML design, structured with CSS is done for the basic structure, with JavaScript for content reaction. In order to allow in-frame video streaming, and variable exchange with the back-end (with instrument-connector and data storage) without the needed of refresh the site, "Asynchronous JavaScript and XML" (AJAX) is used.

JavaScript and, in particular AJAX, allow to implement a complete program in the user we browser. When this is join with API interaction with the back-end, experience is half front-end half back-end. This allows front-end access data, but only with the filter considered by API, so non-illegal (non-allowed by design) actions can be done with back-end data or over Instrument-control.

User interaction and functions related are implemented in the front-end (in JavaScript) and steps (values, validation, sequences, steps done, etc.) are implemented in back-end. In this way, user

cannot modify variables related with the progress or grade of the experience.

Figure 1: Basic structure of technologies in the Remote Lab system

A lab experience is a set of data in back-end and front-end. Back-end has all connections needed with hardware, steps, goals for evaluation, etc. All of them are functions, which can be called via API, in order front-end can use them. Front-end has the user interaction web site, with all its functionalities, and use back-end API to get the data related with the experience and for interact with the hardware. It is important to understand that some experiences proposed are similar, (e.g. frequency response (components, amplifier, filters, etc.) involve function generators and oscilloscope) so same web interface can be used, only must be changed the component connected. For these situations, part of the interface can be parametrized and loaded used parameters associated with the remote lab experience.

Sometimes, a lab experience only uses part of the instrument connected to the Instrument-Server. This let free a set of instruments, and if there exists a remote lab experience which use only those instruments, there could be interesting give to students the opportunity of take both experiences at the same time. With this option, number of user, which can take remote lab sessions, would be increased. This can be done with an "instrument/hardware requirement" in any remote lab experience. At any lab experience booking, instruments associated to that experience are booked, and rest of experiences are checked. If any of them are possible, their time slots are let free for booking, in other case, time slots overlapped with the one booked, are booked, as "Instrument-Server booked, experience not possible".

Some remote lab experiences would need test-boards, which are connected with the Instrument-connector and are controlled as another instrument, controlling sources, switches, and detecting its presence or not. Experiences associated to test boards are only available when test-boards are connected to the system.

As indicated before, requirements for Instrument-Server is a computation system, which can run Python and can interact with the Instruments present in that Lab Site. Depending on those instruments and its requirements, in relation with the communication, Instrument-Server can be implemented in a Single-Board Computer as Raspberry pi (Rpi), or ever, using alternatives to Flask, in a microcontroller as ESP32 with microPython, but this option is more complex, proposed system recommends hardware which support Flask. For a usual set of instruments, taking in consideration that an Instrument-Server has one client at time (in addition with communication with main server) a Raspberry-pi V3b has enough computation power and interfaces to interact via LAN, and USB Instruments, even with the camera streaming. In relation with computers, there are not needed for Instrument-Connector a high computational power system, with at least 2-4GB RAM (depends on the OS), 1 core with 2 GHz, Windows 10, Linux or OSX works fine.

3.2. Reception-Server

In relation with the single server, which manage all the system, it is named Reception-Server. This server is point where the user interacts, and all its experience is managed inside this system. This server is this link among users, out of the University network and Instrument-Servers, in University Local Network. This server need special network configuration, due to it must be reachable from the internet, so it must have a domain or sub-domain associated, at the same time to it can access any system in the University Network, in order to reach Instrument-Servers. Actually, only are exposed common web ports (80, 443) to the internet.

Reception-Server host the user database and manage the user login system using the capacities of Flask-User. This extension manages user authentication, sing up, user validation via email, and even gives a basic web-interface (customizable) for all steps. This solves programmer all user related tasks. As all user-logging systems, passwords are not stored. It is stored in database the result of a calculation done over the combination of password and username, called hash.

Reception-Server host the Shift Manager. This is a compound of databases entries and interactions. Lab technician must create time slots duration per lab experience, set available experiences, and Instrument-Servers available time, in order to create available time slots. When time slots are created, they can be booked, with the restrictions sets by the fair use rules. Fair use rule ensure that all users can access all experiences. As indicated before, a basic fair use rule is a limit of simultaneous booking (one per user). Depending on the number of users and the number of time slots available per experiences, a maximum number of tries per user per experience can be set. That rules can be one or two, or a complex rule as one until everyone has done the experience, or one in this time range, giving in a time period the chance to everyone, and then, giving the chance to any user to repeat an experience. These rules may seem complex, but they are simple, having in the database a register of all experiences taken, with date, user, and other data of interest. Only reading the database with the proper filters, all conditions can be tested, and every condition check can be implemented with the condition as a zero limit avoid its application.

Reception-Server is a manage server, it does not host any instrument or experience. However, it need to know the Instrument-Servers, which are present in the network. This implies the IP address, API commands and related ports (common for all Instrument-Servers), Instruments and experiences available, time range in which they can be used, etc. Periodically test connection is done in order to detect sudden disconnections. As Reception-Server is designed for manage the whole system, some configuration of Instrument-Servers can be change from Reception-Server, as Instruments available, time range in which system can be used, experiences available, fair use rules, etc.

Reception-Server is the endpoint for the user experience in any moment. User registration, booking time slots, and take tests, actually is done in Reception-Server. However, remote Lab experiences are host in Instrument-Servers, whose are not reachable by users. This is solved with a forward done by Reception-Server. Instrument-Server associated with the lab experience shows its web-interface, and Reception-server create a SSH tunnel, creating a forward of that web-interface to an Iframe (a web-site object) inside the main interface of the Reception-Server. With this, user still in Reception-Server, and can interact with Instrument-Server. SSH port forwarding can be done with openSSH, and Python has a module to manage it, so connections can be created easily, and stopped when needed by the back-end. As no ports want to be open, a redirection to a URL, under the Reception-Server domain is done.

Users must register in the Reception-Server, and include its identifier in the Learning Management Systems (LMS), as Moodle, is it is used in lessons. With this, data can be packed and prepared to be uploaded to the LMS automatically.

When a user books an experience, data is exchanged with the related Instrument-Server, and an instrument compatibility check is done. In that moment, non-compatible experiences are detected, and all time slots overlapped with the booked one, are booked as "Instrument-Server booked, experience not possible", as indicated before. User booking (time slot, experience, and user identifier) is registered in Instrument-Server, in order to prepare the experience, at the booked time.

If a booked time slot is un-booked, a search for "Instrument-Server booked experience not possible" books in overlapped time slots, for the same Instrument-Server. For each one, an instrument compatibility test is done with time slots overlapped with them. Slots booked as "Instrument-Server booked, experience not possible" (it could be possible that next or previous time slot of other experiences are booked, overlaps with tested time slot, and are not compatible), if results non-compatible, time slot still booked as "Instrument-Server booked, experience not possible", if not, time slot is change to free for be booked.

3.3. General Concepts

One of the most important part of the remote lab experience is the camera. This implies a live streaming during the experience. It is important to understand that, depending on the camera quality, and the frame per second (fps) sent, data flow involved could be too high. Therefore, it is important to control these two parameters, and adjust them, according to the experiences. If only slow changes are expected, 5-10 fps could be enough, but for watch an Oscilloscope, at least 20 fps are needed. Quality should be revised for each experience, depending the surface the camera is recording; in order to information streamed can be read.

Even when SQLAlchemy disconnect partially programming of database with ORM, database type used could include more or less options in those ORM. In particular, classical SQL databases as SQLite, MySQL or Oracle, are pure relational databases. This implies a strong data relation and order, and a very fix data structure. All data are organized in tables; each row of each table has the same number of columns (cells). Each cell is one data, of one data type. Each one of them has its particularities. However, there is another database type, PostgreSQL which works as same as they are, and include a special data type JSON and JSONB, which are data structures, in a cell, where any number of data (non-limited to the number of columns), each one labelled with a key. Moreover, JSONB included search functions, as seen in SQL instructions for search data in column. This kind of variable is useful for store interaction logs, tries, goals, steps of experiences, and other data in a flexible way. Taking one table of experiences, instead one table per experience. For that reason, PostgreSQL is the database selected. Actually, this procedure, which is done by the database engine, can be done in a more basic way, with blob (binary large object) or text variables, and load there JSON variables. Then, recover and decrypt the data into a python variable. Anyway, if a usual relational database wants to be used, and only relational data want to be stored, data, which is stored in each step, is fixed and one data model (table) per experience is introduced.

Instrument-Server usually have direct connections with Lab Instruments, as USB, or GPIB. This kind of Instrument-Servers are not hot swap. However, if connections with all Instruments are done via LAN, Instrument-Server could have a redundant copy, and, in case of fail, change the destination for forwarding from the main system to the backup one. Most Usual situation is the first one, so a periodical complete backup of all systems for all Instruments Servers is highly recommended.

In relation with the Reception-Server, this server has not any specific hardware connected, so a redundant server can be connected, with a periodical copy of the database. If the main Reception-Server disconnect from the network, the redundant one takes its place. For Instrument-Servers, this rule could be implemented in Reception-Server, but in this case, we need a previous server, which redirect the data, or implement routing rules.

With all presented, a stable and scalable system is obtained, it is easy to configure (almost all configuration in Reception-Server, even of Instrument-Servers) and programming, with a base of required programming languages. General system structure and data flow is shown in Figure 2.

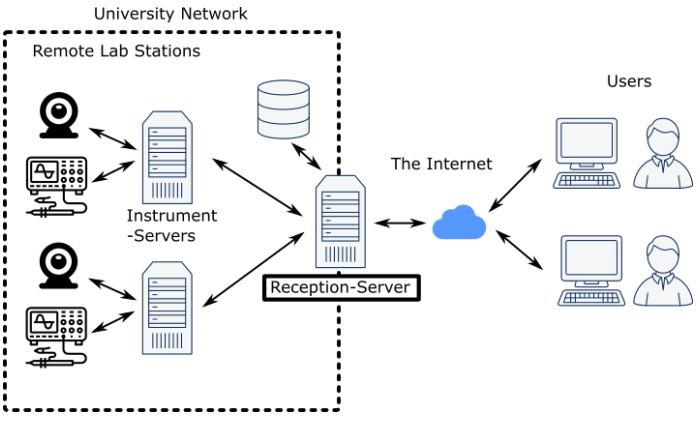


Figure 2: System global structure and data flow.

As indicated before, user traffic ends at Reception-Server and is Reception-Server that reaches Instrument-Servers. Reception-Server only has opened ports 80/443, so it is safe against attacks.

System cannot be properly defined as free of use or non-proprietary, even when all languages and packages in the core are free of use and distribute, due to Instrument-Connector may require privative drivers. Designed system core has not copyright lock or charge, all modules are free of use (some of them “”as is”, without modifications), including author references or copyright.

4. Integration in Engineering Sessions

In engineering teaching, there are too much knowledge to communicate to students, but as technical teaching, many of that knowledge is related with real world aspects, interactions, cause and consequence. In engineering, it explains how things work, including how the physical laws on which the function of those things are based work. Even in most situations complex mathematics support those explanations, real examples can be done in Labs, and students can see a real world example of those theoretical concepts that they are been studding, only supporting by numbers.

This is the reason of Lab sessions in university education, give students real examples of theoretical concepts studied. However, an engineering student must study many concepts, and this implies more theoretical hour than lab session hours. In addition, in a two hours’ theoretical session, many concepts can be explained, and in a two hours’ Lab session, only few concepts can be put in practise, due to they must be examined careful with Lab experiences.

For all explained before, there is not enough time to put in practise all concepts explained in lessons, and that is not the best for students. Searching a learning improvement, more lab time is needed, but there is not available lab time or professors. This can be solved with remote lab experiences. Even when students are not touching hardware, it has been proved in [1] that remote lab is a better solution than simulated lab, when no access to lab is possible.

Moreover, in special situations, as seen with the globally outbreak of COVID-19, even programmed physical lab session may not be possible, or may be reduced. For those situations, remote lab experiences are a great option.

The point is to configure a lab station, for work all time as remote lab station, and even lab stations used for lab sessions, could be set to remote lab stations out of Lab session hours. This requires a configuration, which must be done for the Lab technician. With these options, students can take remote lab experiences any time, with a station exclusively remote, or only out of lab hours, with flexible lab stations. As same as the fair use rules, this depends on the number of students, and instruments and stations available.

As indicated before, remote labs stations are designed for a set of experiences, due to in most areas of engineering, same set of instruments are needed for take all experiences in a subject, or in a part of a subject. E.g. in electronics, if DC bias point is studied, a controllable DC power source, Voltmeters and Amp meters are needed for almost any test. This kind of experience could be behaviour of a capacitor or inductor in DC, in bias point, resistor V-I response, amplification of a transistor without polarization (beta calculation), polarization of a transistor, amplification of an Operational Amplifier in DC, etc. With a Function generator, constant level DC power sources (5V, 12V), and an Oscilloscope, AC response can be studied, with experiences as amplification in AC of an Operational Amplifier, amplification of a polarized transistor, frequency responses of filters, frequency response of components. In addition, mixing, VCO (voltage controller oscillator), limits of Operational Amplifier in relation with feed voltage, effect in transistor amplification of polarization voltage, etc.

As seen, in electronics, there are few equipment, very common in many experiences, and there are others, as pulse generators, clocks, counters, which can give other experiences. The difference among experiences are the components connected to them and the configuration (connections among instruments) needed. A common design can be done for all experiences and Instruments. Creating experience boards, where only instruments must be connected, and a connection with Instrument-Server, in order Instrument-Connector can detects it and make available experiences.

4.1. Experience Board

An example of that common design is done, throw the schematic design of a board for an experience with an Operational Amplifier (LM358), as inverter with variable gain, and switchable to open loop.

First, common elements must be described, starting with the computer communication capacity of the board are shown in Figure 3.

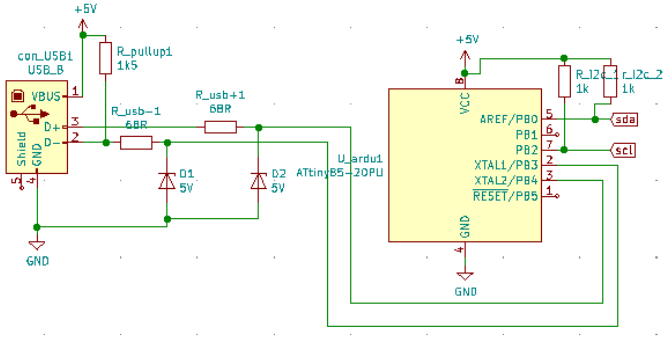


Figure 3: Communication system of test board.

Chip ATtimy85 is a low cost DIP-8 chip, which can be configured with a special bootloader for load Arduino code from USB, without the needed of a USB-RS232 conversion chip, even a serial communication. This chip works with 5V, so even with the USB power can be feed, so only with the elements required for the safe connection for the data transfer (1 pull-up resistor, zener protection diodes, and current limit resistors) chip can be connected with a USB connector.

This chip has few terminals, most interesting ones, I^2C communication. This kind of communication allows connection of some devices to a single port (two wires, with pull-up). Nowadays there are digital elements, as ADC, digital potentiometers, DAC, current meters, and some wind of sensors, with I^2C interface; so many elements can be integrated. Use of external ADC and DAC instead internal of ATtiny85 make easy the design, in order some of them have an analogue level and a digital level, so they can communicate with chip in 5V, and

operate in 20V. This is the case of the digital potentiometer selected for this board, which will be explained later.

In relation with lab instruments connections, as seen in Figure 4, most commons instruments in electronics are DC controlled power source, Function Generator and Oscilloscope. DC source has two outputs, each out with two connections type banana, Function Generator has an output type BNC connector, and Oscilloscope has two inputs type BNC.

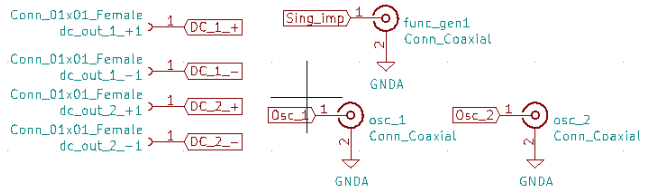


Figure 4: Connections to Instruments of test board.

This connections, join with the ATtiny85 are the base of most test boards (unless other instruments are required, as counters or pulse generators, but those are special experiences, special boards are created, with more connections).

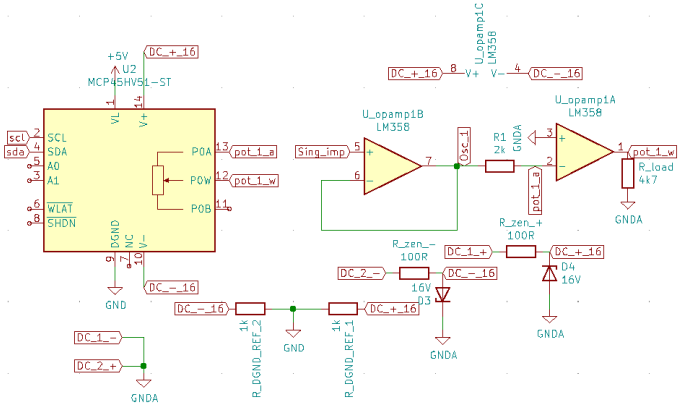


Figure 5: Inverting amplifier with Operational Amplifier experience.

Figure 5 shows the components related to the specific experience, a digital potentiometer, with I2C communication, which can work up to ±18 V, and a LM358 Operational Amplifier, with a maxima power feed of ±16 V. As positive and negative voltage want to be controlled, both DC inputs are used, one for positive and the other for negative voltage. With that objective, negative output of DC1 and positive output of DC2 are connected together to GNDA (analogue ground). Now DC1 controls positive voltage and DC2, negative voltage.

As DC power source can generate more than ±16 V, voltage is limited with zener diodes, to that level, protecting the board. Digital potentiometer requires that digital ground be in the range of analogue voltage, so a resistor divider is done, from limited voltage, and the middle point is connected to the digital ground.

As LM358 has two Operational Amplifiers, one is used as input buffer, and then, as input resistor a constant resistor is used. Digital potentiometer has a limit of ±12.5mA, and input resistance (R1) is the one, which set the current. Fixing the current to a level under the limit, for the extreme input, system is stable. Digital potentiometer is used as feedback resistor (R2), and its impedance could be from 0 to 10kOhms, taking negative amplifications (inverting) lower and upper to one. Digital potentiometer can disconnect terminals, so feedback resistor can be removed, examining Operation amplifier in open loop. For ensure a proper measure a load resistor is included.

As seen this board structure is flexible in its design, and can be change to many electronics experiences, even integrating some in the same board. A price evaluation is done, showing in Table 1 common elements for some experiences, and in Table 2 specific elements for this experience.

Table 1: Price related to common elements

Element	€/unit	Units	€
ATtiny85	1	1	1
USB connector	0.5	1	0.5
Diodes	0.16	2	0.32
Resistors	0.03	5	0.15
Banana connectors	2	4	8
BNC connectors	2	3	6
Total			15.97

Table 2: Price related to specific experience elements

Element	€/unit	Units	€
LM358	0.6	1	0.6
MCP45H51	1.31	1	1.31
Diodes	0.16	2	0.32
Resistors	0.03	5	0.15
Total			2.38

As seen in tables, whole experience, including common and specific elements does not require expensive elements, being all needed components 18,35€. However, as common elements can be used for some different experiences, edge connectors can be included between common and specific part, creating a test system, with a common part, and a changeable experience. In this situation, experience must identify itself for the computer, and it is done including an EEPROM memory, with I2C communication, in the test board. Edge connector has a cost of 3€ (only one is needed, male edge connector is part of the board and EEPROM chip has a cost of 0.1€. With this increase, and minimal design changes, same interface board to PC and instruments (the most expensive part) can be used to some experiences board. At the same time, this change allows to replace interface board or experience board, if any of them fails, and is not needed to replace the whole board.

PCB design for higher frequencies must be done carefully, in order not to introduce coupled effects. Some experiences, as the one proposed in this board, are designed for frequencies up to 1MHz. This make them sensitive to track routes, even to track shapes. This point is the actual develop point, ensure no

Figure 6: Main web interface GPIB SCPI experience

interaction is done, in interface board (for all Oscilloscope or Function Generator frequency range) and in experience board (for all frequency required in the experience).

This is an iteration process, due to each new design must be implemented and tested, and take some time. For that, this line of test board will not be implemented jet, up to all test finished.

4.2. Experience example

Remote lab experience which has been implemented as example is a DC power supply control using Standard Commands for Programmable Instruments (SCPI) (the base of the VISA protocol), connected with the General-Purpose Instrumentation Bus (GPIB). Communication with many lab instruments ends in a VISA commands exchange, throw different busses. This experience gives to students the chance to experience and understand this communication. As indicated before, the camera streaming is the base of the remote lab experience, an in this situation is recording the front panel of the DC power supply. Whit that, students can see in any moment the reaction of the Instrument to any command sent to it (errors, changes in display, changes in operation mode, etc.). For this remote lab experience, simple interaction page has been designed, which in seen in Figure 6.

Power supply connected in this situation is Agilent E3646A DC power supply, and it can be seen in the camera-streaming box, al left. This experience has steps, oriented as questions, which can be answered in any order. Those questions are listed in the combo-box under the camera streaming, and the selected one is the one, which must be answered. Depending on the question, should be answered in the proper answer box (at the right of the question combo-box), or interacting with the proper SCPI command with the instrument.

"Instrument Bus" text box is the point where the user must write the instructions to be sent to the Instrument, when the button "SEND" is clicked. In this field, responses taken from the instrument, when "RECEIVE" button is clicked, are shown too, being the SCPI interaction point. When "SEND" and "RECEIVE" are clicked, in addition to interact with "Instrument bus", Output Status is updated. This value indicated communication status, related to data send process (if information has been delivered, if Instrument can be reached, if there are data to be read, after an instruction with data return, in the instrument and another instruction has been sent), in other words, indicated communication OK, or ERROR (with ERROR specifier).

In case of ERROR for instructions, there are two options, "CLEAR ERROR" erase error vector in the Instrument, clearing error indicator in front panel. Other option is "RESET" completely the Instrument, clearing errors and putting all settings to defaults values. If ERROR is related not to be able to reach instrument, maybe GPIB configuration, which can be configured in "CONFIG" should be revised. There, GPIB address of the instrument, and for the controller should be set. Bus has to be "INITIALIZE", at experience start, or after each configuration change.

There only left "HELP" where a popup with information related with experience interaction and with the experience, procedure and steps are explained. In addition, "PROGRAMMING MANUAL", where a popup is opened which the programming manual of the Instrument, with all instructions needed for the experience.

When a proper answer has been done for a step (with the proper interaction with the Instrument or with the proper answer in the answer box), the background for it turn into green in the combo-box. When all steps are green, experience finished with a pop-up, but Instrument-Server still connected during reserved time-slot, for interaction with the Power Source.

With this, experience is explained. Evaluation is detailed in following section.

5. Learning Results

As part of experience design, goals are defined (pass all steps in experience time, pass steps with less than n tries, take less than a specific time in pass the whole experience, take at least two tries in at least three steps, for ensure student is not copy another student answers, etc.). Those goals can indicate a mark, related to experience evolution, as same as a when students fill a form with

measures and answering questions in Lab sessions. This is an auto-evaluation, during the remote Lab session, which is complemented with evaluation activities.

First, it is vital to know if there has been any problem or difficulty in user interaction with the remote system. For that, a quick survey, related to remote interaction is done, where "Easy to use", "Camera quality", "Response time" and "Proper time slot for experience" are asked, and must be evaluated in the range of: Not work, Very Poor, Poor, Adequate, Good, Very Good. In any moment, if system receive some Poor or Very Poor in an area, a warning is raised to administrator, in order to revise the experience or connection. If "Not work" is marked, an Error is raised to user experience register, in order to mark as not valid this try, and open an additional one. In addition, this must not be an anonymous survey, due to in case of Very Poor or even Poor, difficulties could be get to develop the experience, and that should be considered.

Auto-evaluation, gives information about how students follows experience steps. However, Lab experiences and remote Lab experiences, are oriented in reinforce theoretical concepts. Proper succeed in remote Lab experience is ensure those concepts has been understood. With that objective, a fast knowledge quiz is take by students, after take the remote Lab experience, related with the concepts reinforced in the experienced. This test is used for set the mark of the experience, joint to the auto-evaluation. Moreover, it is an indication for evaluate the experience, if none student passes this quiz, experience should be re-designed, due to it is not reinforcing the concept as desired.

As time during which can be in connection with Instrument-Server is limited by time-slots, survey and quiz could be a problem, due to they take time of experience. For that reason, these exercises are implemented in Instrument-Server, due to they are related to the remote Lab experience, but users perform them in Reception-Server. When experience ends, this is when user pass all steps, or time slot ends, Instrument-Server send to Reception-Server evaluation activities. From this moment, user can perform them during an assigned time (by default 1 day), without the needed of connection with Instrument-Server.

With all that, evaluation is take in two levels, during the experience, automatically, and after the experience, with survey and quiz. At the same time, student evaluation helps for the evaluation of the experience itself, detecting point to change or improve, or in the other hand, validating the functionality of the experience. With all these, remote lab experience is continuously evaluated, by survey, and using student's marks.

Additionally, all Student interactions with remote lab experience, are logged and saved in order to revise auto evaluation protocols, or if needed to manual revision.

6. Conclusions

A remote lab system has been designed, with the capacity of interact with almost any hardware, due to Instrument-Connector can be designed for interact with almost any kind of protocol or driver. It is easy to scale, adding or removing Instrument-Servers, with a simple enter in the Reception-Server, the single administration Server. Even the management of the system, composed with many servers, have been simplified, and almost all of it is done in the Reception-Server.

In relation with remote experiences, as Instrument-Connector can interact with almost any hardware and the combination of HTML+JavaScript can create a huge set of visual experiences in the web browser, almost any experience can be implemented.

Special network requirements (the most complicated part in complex infrastructure network) are only needed in Reception-Server, where external access are needed, with a domain and common web ports opened (80 and 443). Instrument-Servers only need to be in the same network than the Reception-Server.

As educational experience, evaluation system is designed, as a multi-step evaluation. First, goals are set during the experience, and that implies an auto-evaluation, join to a log of the user interaction with the experience. This is used to revise possible errors in steps or auto-evaluations. At experience ends, two fast exercises must be taken. One is a fast survey, about user interaction, in order to detect possible problems in camera quality, latency, time associated to the experience, information done, etc. The other one is a short quiz related to the theoretical concepts related with the experience, in order to confirm that experience has been useful and has reinforce those concepts.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The Spanish Ministry of Economy, Industry and Competitiveness [Grant No. supported this work TEC2016-77632-C3-3-R]. The authors would like to thank the Andalusian Government for funding the Research Unit PAIDI-TIC-168 in Computational Instrumentation and Industrial Electronics (ICEI) and the University of Cádiz.

References

[1] K. Jona, R. Roque, J. Skolnik, D. Uttal, D. Rapp, "Are Remote Labs Worth the Cost? Insights From a Study of Student Perceptions of Remote Labs," International Journal of Online and Biomedical Engineering (IJOE), **7**(2), 48–53, 2011, doi:10.3991/ijoe.v7i2.1394.

[2] I. Evangelista, J.A. Farina, M.I. Pozzo, E. Dobboletta, G.R. Alves, J. García-Zubía, U. Hernández, S.T. Marchisio, S.B. Concari, I. Gustavsson, "Science education at high school: A VISIR remote lab implementation," in Proceedings of 2017 4th Experiment at International Conference: Online Experimentation, exp.at 2017, Institute of Electrical and Electronics Engineers Inc.: 13–17, 2017, doi:10.1109/EXPAT.2017.7984378.

[3] J. Garcia-Zubia, J. Cuadros, V. Serrano, U. Hernandez-Jayo, I. Angulo-Martinez, A. Villar, P. Orduna, G. Alves, "Dashboard for the VISIR remote lab," in Proceedings of the 2019 5th Experiment at International Conference, exp.at 2019, Institute of Electrical and Electronics Engineers Inc.: 42–46, 2019, doi:10.1109/EXPAT.2019.8876527.

[4] S. Rigby, M. Dark, "Designing a flexible, multipurpose remote lab for the IT curriculum," in Proceedings of the 7th ACM SIG-Information Technology Education Conference, SIGITE 2006, ACM Press, New York, New York, USA: 161–164, 2006, doi:10.1145/1168812.1168843.

[5] L. De La Torre, M. Guinaldo, R. Heradio, S. Dormido, "The ball and beam system: A case study of virtual and remote lab enhancement with Moodle," IEEE Transactions on Industrial Informatics, **11**(4), 934–945, 2015, doi:10.1109/TII.2015.2443721.

[6] I. Titov, "Labicom.net - The on-line laboratories platform," in IEEE Global Engineering Education Conference, EDUCON, 1137–1140, 2013, doi:10.1109/EduCon.2013.6530251.

[7] I. Titov, A. Glotov, Y. Andrey, V. Petrov, "Labicom labs: Remote and virtual solid-state laser lab, RF & microwave amplifier remote and virtual lab: Interactive demonstration of Labicom labs in winter 2016," in Proceedings

of 2016 13th International Conference on Remote Engineering and Virtual Instrumentation, REV 2016, Institute of Electrical and Electronics Engineers Inc.: 336–338, 2016, doi:10.1109/REV.2016.7444496.

[8] M.D. Da, "Rasberry Pi Based Remote Lab Implementation," International Journal of Scientific & Engineering Research, **7**(8), 2016.

[9] W. Farag, "An Innovative Remote-Lab Framework for Educational Experimentation An Innovative Remote-Lab Framework for Educational Experimentation," International Journal of Online and Biomedical Engineering (IJOE), **13**(02), 68–86, 2017, doi:10.3991/ijoe.v13i02.6609.

[10] S. Odeh, E. Ketaneh, "A REMOTE ENGINEERING LAB FOR COLLABORATIVE EXPERIMENTATION," International Journal of Online and Biomedical Engineering (IJOE), **9**(3), 10–18, 2013, doi:10.3991/ijoe.v9i3.2500.

[11] I. Angulo, L. Rodriguez-Gil, J. Garcia-Zubia, "Scaling up the Lab: An Adaptable and Scalable Architecture for Embedded Systems Remote Labs," IEEE Access, **6**, 16887–16900, 2018, doi:10.1109/ACCESS.2018.2812925.

[12] 10 top Programming Languages in 2019 for Businesses.

[13] Top 10 Programming Languages of the World – 2019 to begin with… - GeeksforGeeks.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 110-122 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Kamphaeng Saen Beef Cattle Identification Approach using Muzzle Print Image

Hathairat Ketmaneechairat[1,*], Maleerat Maliyaem[2], Chalermpong Intarat[3]

[1]*College of Industrial Technology, King Mongkut's University of Technology North Bangkok, Bangkok, 10800, Thailand*

[2]*Information Technology and Digital Innovation, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand*

[3]*National Biobank of Thailand, National Science and Technology Development Agency, Pathum Thani, 12120, Thailand*

ARTICLE INFO

Article history:
Received: 21 March, 2021
Accepted: 23 June, 2021
Online: 10 July, 2021

Keywords:
Kamphaeng Saen Beef Cattle
Beef Cattle Identification
Muzzle Print Image
Machine Learning
SIFT Features
Gabor Filters
RANSAC Algorithm
Brute-Force Matchers
k-NN Algorithm
FLANN based Matcher

ABSTRACT

Identification of Kamphaeng Saen beef cattle is important of the registration and traceability purposes. For a traditional identification methods, Hot Branding, Freeze Branding, Paint Branding, and RFID Systems can be replaced by genius human. This paper proposed a Kamphaeng Saen beef cattle identification approach using muzzle print images as an Animal Biometric approach. There are two algorithms used in the system: Scale Invariant Feature Transform (SIFT) for detecting the interesting points and Random Sample Consensus (RANSAC) algorithm used to remove the outlier points and then to achieve more robustness for image matching. The image matching method for Kamphaeng Saen beef cattle identification consists of two phases, enrollment phase and identification phase. Beef cattle identification is determined according to the similarity score. The maximum estimation between input image and one template is affected from two perspectives. The first perspective applied SIFT algorithm in the size of the moving image with the rotating image and applied Gabor filters to enhance the image quality before getting the interesting points. For a robust identification scheme, the second perspective applied the RANSAC algorithm with SIFT output to remove the outlier points to achieve more robustness. Finally, feature matching is accomplished by the Brute-Force Matchers to optimize the image matching results. The system was evaluated based on dataset collected from Kamphaeng Saen (KPS; 47 cattle, 391 images), Nakhon Pathom and Tubkwang (TKW; 39 cattle, 374 images), Saraburi, Thailand. The muzzle print images database was collected between 2017 and 2019, in the total of 765 muzzle print images from 86 different cattles. The experimental result is given 92.25% in terms of accuracy which better than a traditional identification approach. Therefore, muzzle print images can be used to identify a Kamphaeng Saen beef cattle for breeding and marking systems.

1. Introduction

Kamphaeng Saen (KPS) beef cattle breed has been developed from a cross breed of Thai native cattle, Brahman and Charolais. Kamphaeng Saen beef cattle is suitable for tropical environment in Thailand and produce a high quality of meat [1]. The establishment of the breed is the results of the long-term effort of the research and development team of the Department of Animal Science, Faculty of Agriculture, Kasetsart University (KU), which was initially through the research project of the beef cattle by Prof. Chran Chantalakhana since 1969. The composite breed of 25% Thai native (N), 25% Brahman (B) and 50% Charolais (C) showed the superior genetic potential under the local Thai farming environment [2, 3, 4] shown in Figure 1. The native cattle are known to be superior in their high fertility in terms of regular estrous cycle and conception rates. The native cows produce high calving percentage even under sub-optimal feeding. Due to a small size and slow growth rate, the native cattle have not used in

*Corresponding Author: Hathairat Ketmaneechairat, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand, hathairat.k@cit.kmutnb.ac.th

https://dx.doi.org/10.25046/aj060413

commercial fattening system. Hence, crossbreeding between an exotic breed, the Brahman, and Thai native cattle was made to improve size and growth rate in the crossbred. In terms of beef quality, beef from Brahman is less desirable than beef from temperate cattle such as Charolais [5]. Although, Charolais is not well adapted to hot climate, but its growth performance and beef quality are favorable [6]. The breed formed by intersemating of the crossbred followed by selection will be named after the district where it was developed [7].

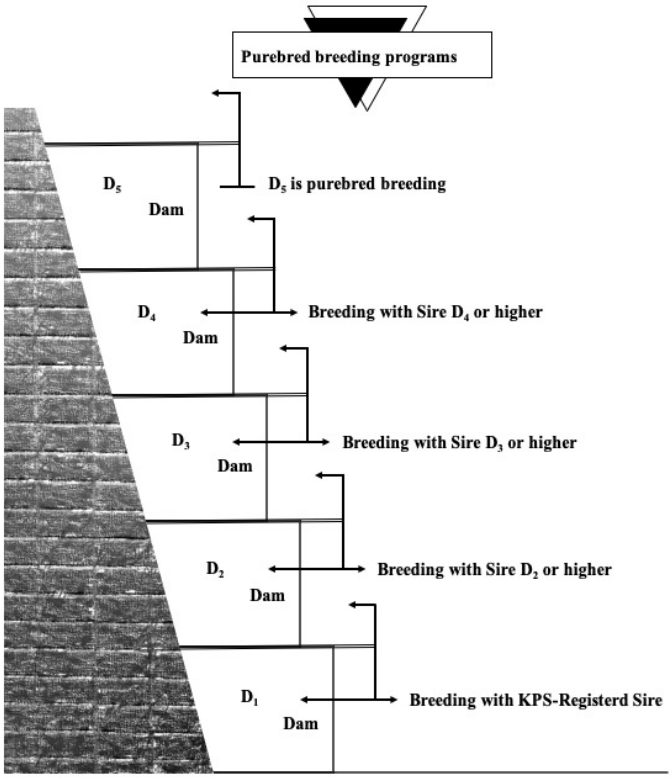


Figure 1: Kamphaeng Saen (KPS) beef cattle purebred breeding programs [7].

Since 1991, Kamphaeng Saen cattle have been distributed to farmers who are members of the KPS Beef Breeders Association. However, the uniformity of color and conformation of Kamphaeng Saen cattle has yet to be improved. In 1992, KPS Beef Breeders Association was founded, with 200 farmer members, registered the breed and the criteria for beef were established as follows:

1. Hair color is creamy to light yellow is the best suitable. However, some variation in color is accepted if good pedigree.
2. Genotype is 25% native, 25% Brahman, 50% Charolais; slight variation is acceptable.
3. Having birth date, farm brand and individual identification.
4. Having suspended testicles at six months of age.
5. Passing general assessment by association's officials [8].

Beef cattle identification is the method of recording a beef cattle with birth date, Sir/Dam, production, feeding programs, and health management. The identification system is an important method of livestock production. There are four types of identification:

1. Permanent methods: Ear Notches, Tattooing, and Hot/Freeze/Paint Branding.
2. Temporary methods: Chalk, Ear Tagging, and Neck Chains.
3. Electrical methods: Microchips, RFID Systems, Ruminal Boluses, and Injectable Transponders.
4. Biometrics methods: Muzzle Prints, Iris Patterns, Retinal Vascular, and DNA Profiling [9].

However, the main problems of these methodologies are low image quality, infliction of injury on the body of an animal, low-frequency coverage, loss of tags, and duplication respectively [10]. Hence, devising a robust means for cattle identification to mitigate the iterated challenges is a task that involves the state-of-the-art machine learning techniques in animal biometrics. DNA profiling is the process of determining an individual's DNA characteristics, focusing on short tandem repeat (STRs), nucleotide polymorphism (SNPs), mitochondrial DNA (mtDNA), and sequencing markers of an individual animals [11].

The measurements of biometric authenticate features are seven characteristics as follow:

1. Universality: a person has common characteristic.
2. Uniqueness: two persons have characteristic with a high of uniqueness.
3. Permanence: the characteristic never changed over time with advancing age.
4. Collectability: the characteristic easy to acquire can be measures.
5. Performance: how well a system factors include recognition accuracy, speed, and error rate.
6. Acceptability: how accept the characteristic into a system.
7. Circumvention: how easily which a system can be fooled by fraudulent biometric identifier.

A biometric system operates in two modes. First, verification mode: the system validates identity by biometric captured with own biometric template(s) stored in the database system. Second, identification mode: the system searching all template for a matching one-to-many comparison to establish an individual's identity. The system is designed four main modules: 1) Sensor module: which captures the biometric data of an individual. 2) Feature module: which acquired biometric data processing to extract a set of salient features. 3) Matcher module: which claimed identity is verification and identification based on the matching score and 4) Database module: which store the biometric templates of the registration.

A biometric system can measure from two types of verification errors: 1) False Match: an error from two different persons to be from the same person (False Acceptance Rate (FAR)). False Non-Match: an error from the same person to be from two different persons (False Rejection Rate (FRR)). A trade-off between FAR and FRR is the functions of the decision threshold in template matching for measuring the performance of a system; if it is decreased to make the system more tolerant to input variations and noise, then FMR increases. On the other hand, if it is raised to make the system more secure, then FNMR increases accordingly. The performance at all operating points (thresholds) can be referred to the concept of a Receiver Operation Characteristic (ROC) curve. A ROC curve is a plot of FMR against (1-FNMR) or FNMR for various threshold values [12].

The traditional muzzle printing method in Japan shown in Figure 2. while the basic procedure is started from making dry

muzzle by cloth, painting some ink on a muzzle by coverage, then rolling to lift paper print the area between upper lip and the top of the nostril. Then check a pattern of ridges and grooves for a complete pattern [13]. The identification process based on muzzle pattern shown in Figure 3, a ridges and grooves extracted from joint pixels as the features. Every joint pixel from two images is overlaid. Two joint pixels are matched if they are in a range of a pixel region [14].

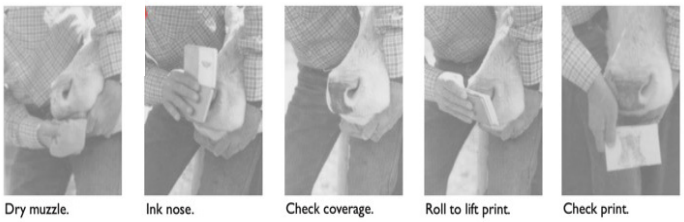


Figure 2: Muzzle printing procedure [13].

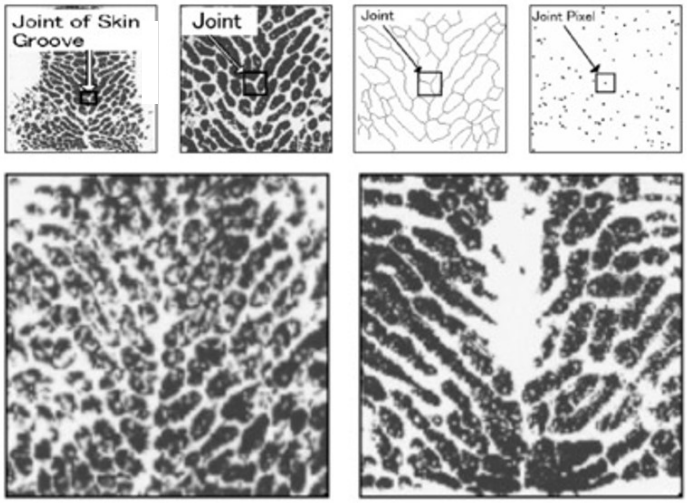


Figure 3: Muzzle pattern [14].

In this paper, a Kamphaeng Saen beef cattle identification approach using muzzle print image is proposed. The proposed applied the image matching method with machine learning techniques such as Scale-invariant feature transform (SIFT), Rectangular Gradients Histogram (R-HOG) which localize and detect the region of interest (ROI) in muzzle print images for the cattle identification and Random Sample Consensus (RANSAC) algorithm which used to remove the outlier points and improve the robustness of SIFT feature matching, RANSAC technique used with the SIFT in order to mitigate noises such a outliers points for better identification.

SIFT is a feature detection algorithm for image processing. It was published by David Lowe in 1999 and 2004. SIFT keypoints of objects are extracted from a reference image. An object in a new image is comparing each feature and finding candidate matching features based on Euclidean distance of their feature vectors. The full match keypoints that agreed up on object location, scale, and orientation are identified to a good match. The consistent clusters are performed by an efficient hash table of the Hough transform algorithm. Each cluster that agrees on object detailed verification and outliers has been discarded. Then, the probability that a set of features indicates is computed for the accuracy of false matches. That all pass object matches can be identified with high confidence.

RANSAC is a predictive algorithm for image processing. It was published by Dr.Martin A Fischler and Robert Bolles in 1981. RANSAC estimates by random sampling of observed data contain both inliers and outliers. Voting scheme implements the data elements for one or multiple models based on noisy features which will not vote for any single model (few outliers) and enough features to agree on a good model (few missing data). The compose of two repeated steps that are iteratively repeated until the consensus set in enough inliers: First, select randomly the minimum number of points to determine the parameters of the model. Second, determine how many points from all points fit with a predefined tolerance. If the number of inliers over the total number of points then re-estimate the model parameters.

The image matching for Kamphaeng Saen beef cattle consists of two phases, enrolment phase and identification phase. The beef cattle identification is determined according to the similarity score. The maximum estimation between input image and one template is affected from two perspectives. The first perspective uses SIFT algorithm in the size of the moving image with the rotating image, and uses Gabor filters for enhancing of image quality before getting the interesting points for a robust identification scheme, the second perspective uses the RANSAC algorithm is used with SIFT output to remove the outlier points and achieve more robustness. Finally, the feature matching is accomplished by using the Brute-Force Matchers for optimizing the image matching results.

The remainder of this paper is organized as follows: Section 2 about the related works, Section 3 explain the methodology, Section 4 is a proposed beef cattle identification approach, Section 5 the experimental scenarios, Section 6 shows the results and discussion. Finally, conclusion and future work is discussed in Section 7.

2. Related Works

The cattle identification using muzzle print image is proposed base on previous work that can be categorized into the image processing technique, machine learning technique and encouraging for a day of livestock management.

In [15], the author proposed a Principal Component Analysis and Euclidean distance classifier to evaluate and performed the muzzle ink prints with the training part from 3 images of 29 different cattle. The results showed that when using 230 eigenvectors (out of 290), the recognition rate was equal 98.85%. This technique as expected reduced the recognition rate when principal component less than 230, while training more images per cattle. In [16], the author using the fusion of texture feature that extracted from Webber Local Descriptor (WLD) and local binary pattern. The result showed that 96.5% in terms of identification accuracy. SURF (speeded-up robust features) and U-SURF (upright version) are the family with SIFT, SUR, it is better than SIFT in rotation and blur transform. SIFT is better than SURF in different scale images and SURF faster than SIFT. Both are good in illumination changes images. In [17], the author proposed SURF technique, the identification accuracy is 93% for 75% of training database. In [18], the author proposed U-SURF with the result of outstanding performance more than the original SURF.

In [19], the author proposed a Local Binary Pattern (LBP) to extract local invariant features from muzzle print images and applied including Nearest Neighbor, Naïve Bayes, SVM and KNN for cattle identification. The results shown that identification accuracy is 99.5%. In [20], the author proposed a multiclass support vector machines (MSVMs) in three phases: preprocessing used the histogram equalization and mathematical morphology filtering, feature extraction used the box-counting algorithm for detecting feature and classifications used the MSVMs. The results shown that 96% classification accuracy.

In [21] the author supported in precision livestock farming which focused on Image-based identification could be a promising non-intrusive method for cattle identification can be approach for deliver quantitative information and complete traceability of livestock in the food chain. In [22], the author shown the experimental results in the feature vector for different image of the same muzzle. It's highly symmetry and this technique can be applied in registration livestock to monitoring individual cattle management system.

3. Methodology

3.1. SIFT Features

Scale Invariant Feature Transform (SIFT) is a feature extraction method based on the extraction of local information. The features extracted are invariant to image scaling, rotation, and partially invariant to change in illumination and projective distortion. Four major stages to generate a set of features are shown in Figure 4:

Scale-Space Extrema Detection: the candidate keypoints can be obtained by detecting extrema from Difference of Gaussian (DoG) pyramid which an approximation of Laplace of Gaussian (LoG). The input data is transformed to the space $L(x, y, \sigma)$ as follows:

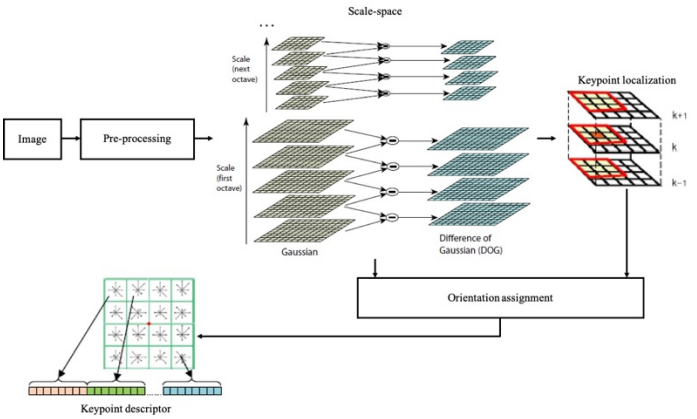


Figure 4: SIFT based on pre-processing [23, 24].

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{1}$$

where $*$ corresponds to convolution operator, $I(x, y)$ is the input image and $G(x, y, \sigma)$ is a Gaussian function with bandwidth σ.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2} \tag{2}$$

$$D(x, y, \sigma) = \big(G(x, y, \sigma) - G(x, y, \sigma)\big) * I(x, y)$$
$$= L(x, y, k\sigma) - L(x, y, \sigma) \tag{3}$$

1. Keypoint Localization: to get stable keypoints, three processes are applied in this step. The first process is to find the accurate location of keypoints using the 3rd order Taylor polynomial; the second process is eliminating low contrast keypoints; and the third process is to eliminate the keypoints in the edge using principal curvature. The interpolation is done using the quadratic Taylor expansion of the Difference-of-Gaussian scale-space function $D(x, y, \sigma)$ with the candidate keypoint as the origin. This Taylor expansion is given by:

$$D(x) = D + \frac{\partial D^T}{\partial x} + \frac{1}{2} x^T \frac{\partial^2 D^T}{\partial X^2} x \tag{4}$$

where

D and its derivatives are evaluated at the candidate keypoint.

$x = (x, y, \sigma)$ is the offset from this point.

2. Orientation Assignment: the orientation of keypoint will be calculated based on the gradient and orientation of a region around the keypoint. A keypoint may have more than one orientation. For an image sample $L(x, y,)$ at scale σ, the gradient magnitude, $m(x, y,)$, and orientation, $\theta(x, y,)$, are processed using differences pixel:

$$m(x, y) = \sqrt{\begin{matrix}\big(L(x+1, y) - L(x-1, y)\big)^2 + \cdots \\ +(L(x, y+1,) - L(x, y-1))^2\end{matrix}} \tag{5}$$

$$\theta(x, y) = tan^{-1}\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right) \tag{6}$$

3. Keypoint Descriptor: a window with the size of 16×16 centered, each keypoint is calculated with the orientation and gradient magnitude. The window is then divided into 4×4 sub regions. An orientation histogram which represented eight cardinal directions are calculated for each sub region based on gradient magnitude. The weight is calculated by a Gaussian window centered in the middle of the window. The keypoint descriptor consists of 128 elements from 16 sub regions where each sub regions consists of 8 features [23, 24].

3.2. Gabor Filters

Gabor filters are formed from two components, sinusoidal and Gaussian, The Gabor function was discovered by Gabor in 1946, where the function is defined in 1-D with t stating time and then developed unto 2-D in the spatial domain formulated [25].

3.3. RANSAC Algorithm

The RANSAC procedure is opposite to the conventional smoothing techniques: Rather than using as much of the data as possible to obtain an initial solution and then attempting to eliminate the invalid data points. RANSAC uses as small an initial data set as feasible and enlarges this set with consistent data if possible. For example, given the task of filtering an arc of a circle

to a set of two-dimensional points, the RANSAC approach will select a set of three points. Compute the center and radius of the implied circle and count the number of points that are close enough to that circle to suggest their compatibility with it. If there are enough compatible points. RANSAC will employ a smoothing technique such as least squares, to compute an improved estimate for the parameters of the circle [26].

3.4. Brute-Force Marchers

The brute-force descriptor matcher uses brute-force approach for feature matching. It takes the descriptor of one feature in the first image and compares it with descriptors of all features in the second image using some distance calculations. Then the closest one is returned in a resulting pair. The brute-force algorithm sometimes takes more time for highly precise. Its performance can be improved by setting specific parameters [27].

3.5. k-NN Algorithm

The *k*-Nearest-Neighbors algorithm (*k*-NN) is a well-known machine learning for pattern recognition method. *k*-NN is a non-parametric classification method, which is simple but effective in many cases. However, it needs to choose an appropriate value for *k* in order to success a classification model [28].

3.6. FLANN based Matcher

FLANN stands for Fast Library for Approximate Nearest Neighbors. It contains a collection of algorithms optimized for fast nearest neighbor search in large datasets and for high dimensional features. It works more faster than BFMatcher for large datasets. FLANN needs to pass two dictionaries which specifies the algorithm to be used : IndexParams and SearchParams [29].

4. The Proposed Beef Cattle Identification Approach

The proposed scheme for a Kamphaeng Saen beef cattle identification approach using muzzle print image is described from two perspectives, Enrollment phase and Identification phase:

Enrollment phase: to enhance input muzzle print image (template image) by Gabor filters and using SIFT features to extract the keypoint descriptor, then store muzzle template to database.

Identification phase: to enhance input muzzle print image (query image) by Gabor filters and using SIFT features to extract the keypoint descriptor. The query is matched against the templates stored in the database as (1:N) matching. RANSAC algorithm and Brute-Force matchers are applied in the matching process to remove the matching outliers, mismatched SIFT keypoints, data to ensure the robustness of the similarity score. The animal identity is then assigned according to the highest estimated similarity threshold score between the input image and the template one, all details as shown in Figure 5.

4.1. Enrollment Module

Muzzle print image was stored in the database folder. Each muzzle print has been registered with template id (i.e. template_001.jpg); cattle's info registered Location, Cow Tag, Gender, Type, and Owner. When cattle have been identified, then all about info of this muzzle print can be retrieved.

4.2. Identification Module

The matching modules was created by Python Script using Python 3.7.3, dependencies are required Numpy 1.16.5, SKimage 0.17.2 and OpenCV2 4.4.0. Module integrated with SIFT Features, Gabor Filters, RANSAC Algorithm, Brute-Force Matchers, k-NN Algorithm, and FLANN based Matcher using for experiments in the identification scheme steps as follow:

1. Place 2 muzzle print images that wanted to compare to the database folder.
2. Pass the names of images as arguments in the terminal console.

4.2.1. Matching Process

Start with Input Image, then Get Description from after Image Enhancement includes Ridge Segmentation (normalizing the image and find a ROI); Ridge Orientation (finding orientation of every pixel); Ridge frequency (finding the overall frequency of ridges extended), Frequent (estimate ridge frequency within image block); Ridge Filter is created Gabor filters and do the actual filtering. Then remove a border pixel under the conditions. BFMatcher is matching between descriptors and Calculate Score and compare with threshold. Finally, identification and decision making are done based on Algorithm 1.

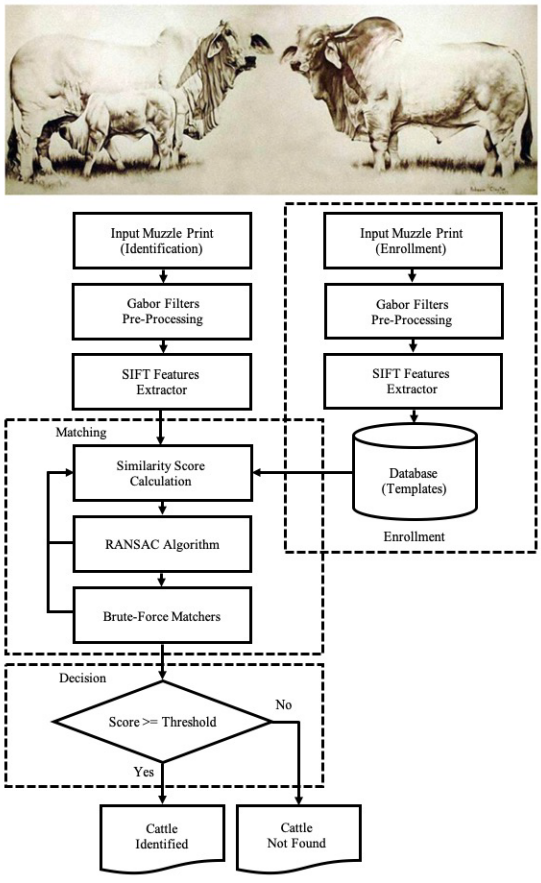


Figure 5: Completed module of a Kamphaeng Saen beef cattle identification approach using muzzle print image.

4.2.2.Decision Process

In this process, a similarity score will be compared with a threshold value to check if it either equal or greater than threshold, the result will be identified as cattle identified otherwise not identified.

Algorithm 1: KPS beef cattleis muzzle print image features *******matching.

```
image_name =sys.argv[1]
img1 =IMREAD_GRAYSCALE
kp1, des1 =get_description(img1)
image_name =sys.argv[2]
img2 =IMREAD_GRAYSCALE
kp2, des2 =get_description(img2)
bf =BFMatcher
matches =match(des1, des2)
score =0
for match in matches:
    score +=match.distance
score_threshold =33
if score/len(matches)< score_threshold:
    'Muzzle print matches.'
else:
   'Muzzle print does not matche.'
```

4. Experimental Scenarios

5.1.Data Collection

The database has been collected Kamphaeng Saen beef cattle between 2017 to 2019. From two locations: Cowboy Land, Nakhon Pathom, Department of Animal Science, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom, Thailand shown in Figure 6a and Tubkwang Reseaerch Center, Saraburi, Department of Animal Science, Faculty of Agriculture at Bangkhen, Kasetsart University, Bangkhen, Bangkok, Thailand shown in Figure 6b.

Figure 6::(a)Cowboy Land, Nakhon Pathom,Thailand and b)Tubkwang Research Center, Saraburi

The lack of an original muzzle print images database was a challenge for this research. Therefore, collecting a muzzle print images database was a crucial decision. The Dataset was collected from Kamphaeng Saen beef cattle with four periods started from May and November 2017, January and June 2018, and March and May 2019, from 2 locations, keep 4 collection per location, in round of 5-11 months period shown in Table 1.

Table 1: Muzzle print images database collected period from 2 locations.

Period	Cowboy Land, Nakhon Pathom (KPS)			Tubkwang Reseaerch Center, Saraburi (TKW)		
	Age (Month)	Month	Year	Age (Month)	Month	Year
1	Collect	May	2017	Collect	May	2017
2	(+)8	January	2018	(+)6	November	2017
3	(+)5	June	2018	(+)7	June	2018
4	(+)11	May	2019	(+)9	March	2019

47 KPS datasets from 47 cattle with muzzle print images each, include male/female and calf/puberty/breeders. 39 TKW datasets from 39 cattle with muzzle print images each, all female, and all breeders. KU 48/053 KPS and KU 52/23 TKW are dead after first period collected setup symbol is D. Photo takes by FUJIFILM X100T, OPPO Mirror 5 and Worker's camera with represented setup symbol are F $(4{,}896 \times 2{,}760 \times 24\text{b JPEG})$, O $(3{,}200 \times 2{,}400 \times 24\text{b JPEG})$ and W $(1{,}478 \times 1{,}108 \times 24\text{b JPEG})$, respectively. If images in the period is zero, its mean that cannot take a photo in this period because cattle stay in the stall.

Figure 7: Sample of muzzle print images database from KU 53/102 KPS.

Sample of muzzle print images database from KU 53/102 KPS shown in Figure 7. and KU 53/005 TKW shown in Figure 8. The image shows Cow Tag and location, Individual image cattle show in the top of left. The different testing method has been setup based on the quality of the collected images. Such as covering collected the muzzle print images based on quality level in different deteriorating factors include orientated, blurred, low resolution, and partial. The original muzzle print images have been taken from 4 periods of different cattle for experimental in the identification conditions.

Figure 8:Sample of muzzle print images database from KU 53/005 TKW

5.2.Data Analysis

Dataset of the muzzle images has been standardized in orientation and scale manually. In every muzzle images, a rectangle region centered on the minimum line between the nostrils is taken as the Region of Interest (ROI) may be in different size so that it is re-sized into 200 × 200 pixels. The image has been enhanced using intensity transformation function shown in Figure 9 and beads and ridges in a muzzle photo shown in Figure 10.

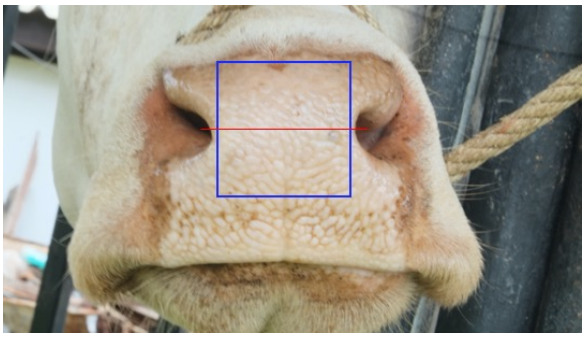


Figure 9: The blue rectangle region is the ROI of the muzzle photo, the red line is a minimum distance between the nostrils.

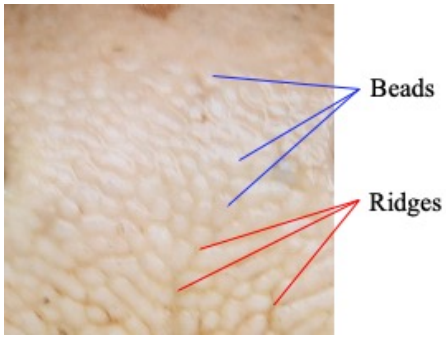


Figure 10: Beads and Ridges in a muzzle photo.

5.2.1.Scenario I

The Scenario I work as follows: 12 images of each cattle have been swapped between the enrollment phase an identification phase, the similarity score between all of images are calculated. Therefore, similarity score matrix with dimension of 200 × 200 pixels have been created. The cattle is correctly identified if the similarity score between the input image and the template image is greater than or equal a specific threshold shown in Figure 11 and Figure 12.

The template of a cattle has been created from 11 images which were marked as $T_1, T_2, T_3, \dots, T_{11}$. The remaining 1 image has been used as input, and was marked as I_1, S was a similarity function, and H was a similarity score. A correctly identified cattle should strictly follow the next equation as:

$$S(I_1, T_1) \,||\, S(I_1, T_2), \dots, ||\, S(I_1, T_{11}) \geq H \qquad (7)$$

From $I_1, T_1, T_2, T_3, \dots, T_{11}$ select one for the BS (Best Selected; good image quality)follow as:

$$S(BS, T_1) \,||\, S(BS, T_2), \dots, ||\, S(BS, T_{11}) \geq H \qquad (8)$$

5.2.2.Scenario II

For all 47 KPS datasets from 47 cattle, the template of a cattle has been created from 4 images which were marked as T_1, T_2, T_3, T_4 from each period and 1 BS image, total 5 images. Each BS from dataset are registered with ordered by name kps_template_001 to kps_template_047. Then 47 muzzle print images in KPS database will be as follow:

$$S(BS, KPS\ DATABASE) \,||\, S(T_1, KPS\ DATABASE), \dots, ||\, S(T_4, KPS\ DATABASE) \geq H \qquad (9)$$

5.2.3.Scenario III

For all 39 TKW datasets from 39 cattles, the template of a cattle has been created from 4 images which were marked as T_1, T_2, T_3, T_4 from each period and 1 BS image, total 5 images. Each BS from dataset are registered to *DATABASE* by order name tkw_template_001 to tkw_template_039. Then have 39 muzzle print images in TKW database, follow as:

$$S(BS, TKW\ DATABASE) \,||\, S(T_1, TKW\ DATABASE), \dots, ||\, S(T_4, TKW\ DATABASE) \geq H \qquad (10)$$

5.2.4.Scenario IV

The total 86 datasets from the 47 KPS datasets (47 cattle) and the 39 TKW datasets (39 cattle), the template of a cattle has been created from 1 image which were marked as T_1 its nearby BS and 1 BS image, total 2 images. Each BS from dataset are registered to *DATABASE* by order name template_001 to template_086. Then have 86 muzzle print images in database, follow as the next equation:

$$S(BS, DATABASE) \,||\, S(T_1, DATABASE) \geq H \quad (11)$$

5.2.5.Identification Time

For the evaluation of the identification time, "the number of image comparisons" and "the processing time of a single image comparison" will be considered in addition to "the total identification time". The total processing time T for an identification can be estimated by:

$$T = M \times (T_1 + T_2) \quad (12)$$

where M is the number of comparisons, T_1 is the processing time of a single comparison, and T_2 is the processing time for a search of the next candidate [30].

KU 53/005 TKW	Equation (7)		
Period 1	I_1	T_1	T_2
Period 2	T_3	T_4	T_5
Period 3	T_6	T_7	T_8
Period 4	T_9	T_{10}	T_{11}

Figure 12:The identification Scenario works as follows * 005 TKW.

5.2.6.Identification Accuracy

The performance metrics by contrast to traditional methods, biometric systems do not provide a cent percent reliable answer, it is quite impossible to obtain such a response. The comparison results between acquired biometric sample and its corresponding stored template is illustrated by a distance score. If the score is lower than the predefined decision threshold, then the system accepts the claimant, otherwise he is rejected. This threshold is defined according to the security level required by the application. Illustrates the theoretical distribution of the genuine and impostor scores. This figure shows that errors depend from the used threshold. Hence, it is important to quantify the performance of biometric systems. The International Organization for standardization ISO/IEC 19795-1 proposes several statistical metrics to characterize the performance of a biometric system [31].

In order to estimate the FMR, FNMR, and EER, suppose one biometric template is denoted by T, and one presented sample (input) is denoted by I. The similarity score S between the template and the input is measured by the function $S(I, T)$. The hard decision is made according to a similarity threshold h.

FMR is the rate that at which the decision is made as I matches T, while in fact I and T come from two different individuals. This means that the biometrics system accepts what should be rejected:

$$\mathrm{FMR}(h) = 1 - \int_{S=h}^{\infty} \mathrm{p_n}(s)\mathrm{ds} \quad (13)$$

where $\mathrm{p_n}(s)$ is the non-match distribution between two samples as a function of s.

FNMR is the rate which the decision made as I does not match T, while in fact I and T originated from the same individual. This means that the biometrics system rejects which should be accepted:

$$\mathrm{FMR}(h) = 1 - \int_{S=-\infty}^{h} \mathrm{p_m}(s)\mathrm{ds} \quad (14)$$

where $\mathrm{p_m}(s)$ is the match distribution between two samples as a function of s.

The Equal Error Rate (EER)is defined as the value of FMR and FNMR at the point of the threshold h where the two error rates are identical $h == EE$:

$$\mathrm{EER} = \mathrm{FMR}_{h=EE} = \mathrm{FNMR}_{h=EE} \quad (15)$$

The similarity threshold (h) should be chosen carefully in the system design phase according to the security level and the system's sensitivity. The similarity threshold should achieve a trade-off between FMR and FNMR errors. FMR and FNMR are not objective measurements because they are influenced by the selected threshold emerging from the system's application. However, FMR and FNMR are still possible to be used to measure performances of specific systems. The value of ERR can be used as a good indicator for measuring the system's performance, and can be selected though the Receiver Operating Curve (ROC) [32].

5. Results and Discussion

All scenarios is defined setup the best matcher method parameters that directed the number of keypoints with the follow best processing time in three matcher method include ORB, Ratio test, and FLANN as shown in Table 2. The analysis result of three

matcher methods when running analysis shown the console left side from KU 53/102 KPS and right side from KU 53/005 TKW. Result display to separate from three matcher methods name and the last one display a identification result. All methods show the Query, Template, Descriptors (Des.1, Des.2), Keypoints (Key.1, Key.2), Matches (number of matches between Key.1 and Key.2), Extraction Time(s), Matching Time(s), Score, Threshold, and Muzzle print (Matches or Not Matches). In Ratio Test method show the Ratio test, and Good matches. In FLANN method show the Ratio test, and Matches mask.

The results in this paper have been running using a MacBook Pro macOS Catalina, 2.3 GHz Dual-Core Intel Core i5, 16 GB 2133 MHz LPDDR3, Intel Iris Plus Graphics 640 1536 MB.

Table 2: Parameter setup of three matcher method.

Matcher method	Parameter setup
ORB	$score\_threshold = 55;$ $score < score\_threshold;$
Ratio test	$k = 2;\ for\ knnMatch;$ $lowe\_ratio = 0.8;$ $* apply\ ratiotest\ as\ per\ Lowe's\ paper$ $score\_threshold = 4;$ $score \geq score\_threshold;$
FLANN	$tree = 5;\ for\ FLANN\_INDEX\_KDTREE;$ $k = 2; for\ knnMatch;$ $lowe\_ratio = 0.8;$ $* apply\ ratiotest\ as\ per\ Lowe's\ paper$ $score\_threshold = 70;$ $score \geq score\_threshold;$

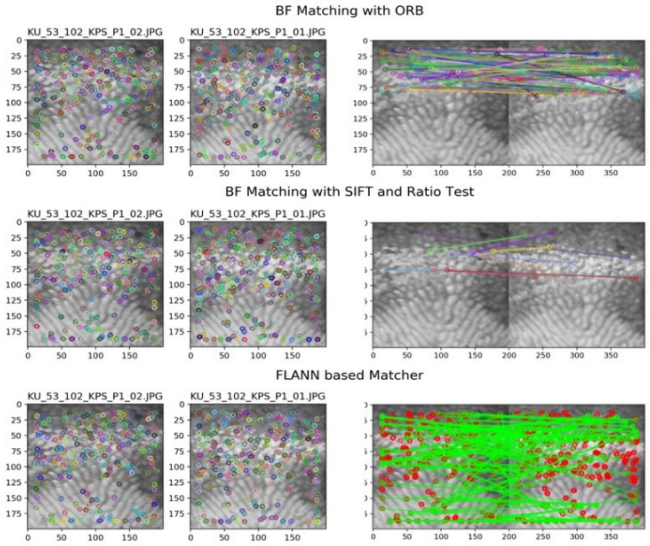


Figure 13: Image result with three matcher methods from KU 53/102 KPS.

Here, will see a result on how to match feature between two images. Then will try to find the query in template using feature matching. Using SIFT descriptors to match features with three matcher method are ORB, Ratio test, and FLANN shown in Figure 13 and Figure 14.

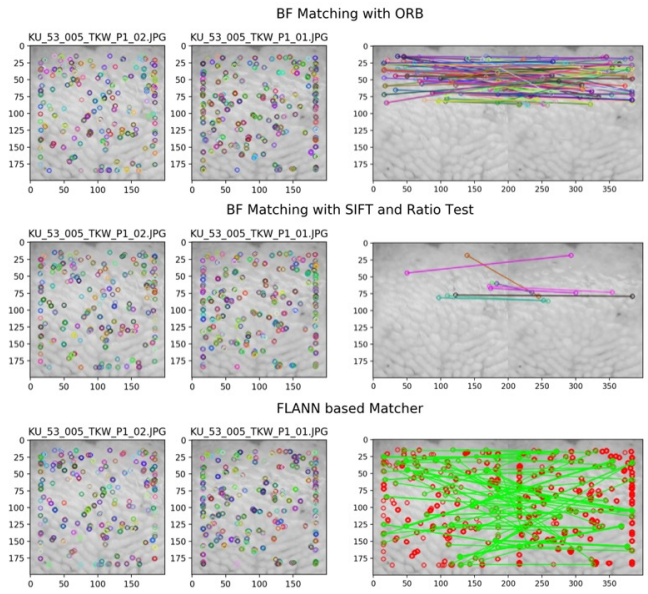


Figure 14: Image result with three matcher methods from KU 53/005 TKW.

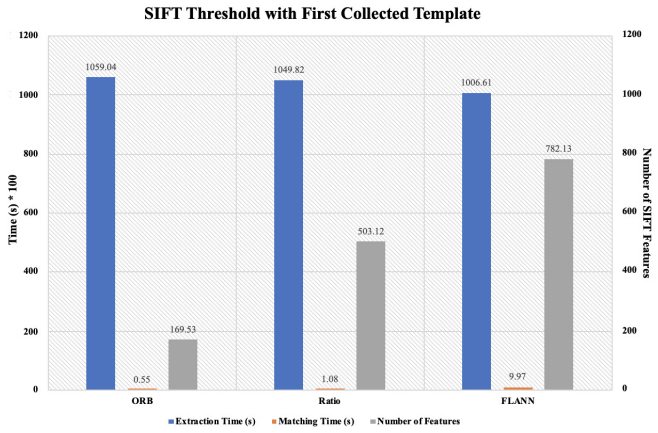


Figure 15: SIFT threshold with first collected template.

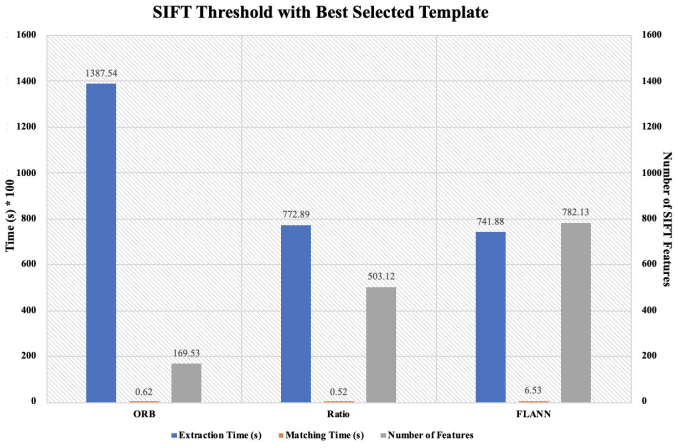


Figure 16: SIFT threshold with best selected template.

Figure 15 shown the result of Scenario I and Figure 16 shown is the SIFT threshold compared between the first collected template (period 1, I_1) and the best selected template (best of all periods, BS). BS was selected from clearer image with good light

condition and shape. The number of SIFT features is matched keypoints between query and template. The time based on second multiply 100 (s * 100). All values are average from number of queries, number of cattle, and number of two locations. In ORB method, the first collected template given extraction time and matching time better than the best selected template. So, the best selected template is given the reduced time of extraction time and matching time in Ratio Test and FLANN methods. The number of the features is equal compare because a list of queries was not changes, but a template has been changed.

Scenario II result in Figure 17 and Scenario III result in Figure 18 show the score threshold to compare between two locations is KPS and TKW, respectively. All values are average from number of queries, and number of cattle. Then, the score threshold with KPS give ORB = 30, Ratio test = 120, and FLANN = 294. The score threshold with TKW in our method give ORB = 32, Ratio test = 127, and FLANN = 306. Some average from number of two locations in ORB = 31, Ratio test = 123.5, and FLANN = 300. So can be estimated setup the score threshold in ORB = 38, Ratio test = 32, and FLANN = 170 in Scenario IV for find candidate and identify cattle.

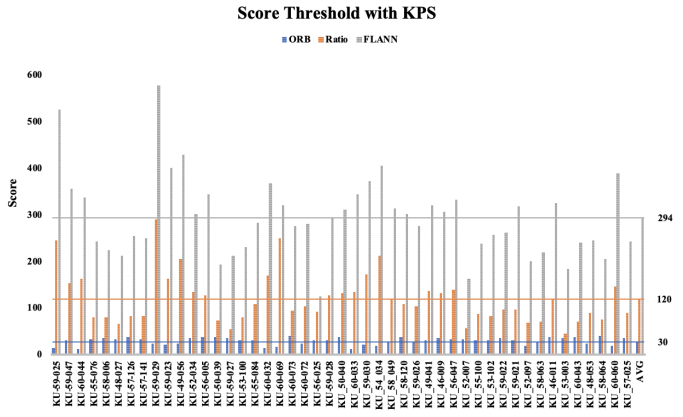


Figure 17: Score threshold with KPS.

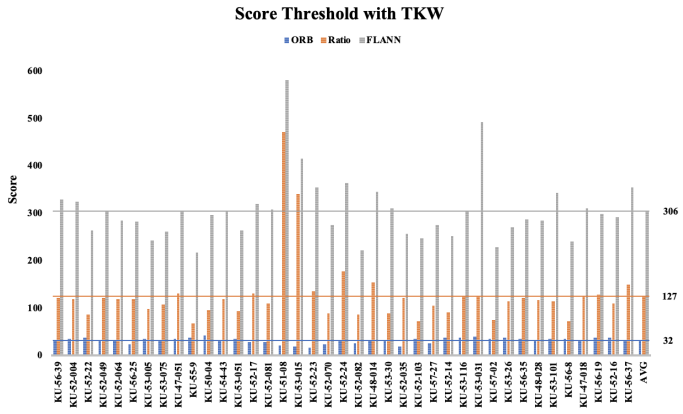


Figure 18: Score threshold with TKW.

Scenario IV result shows the characterization of the linear search. “Number of Comparisons” is the number of image comparisons conducted until the algorithm terminated, “Time for a Comparison” is the processing time required to conduct a single image comparison which includes feature extracting by SIFT, and “Time for a Search” is the processing time required to find the best template for the next comparison in the algorithm. The processing time for a single image comparison was computed separately same process in the linear search. The other values are computed from the results on the threshold of the optimum error rate.

In real time identification, one image of each individual cattle has been processed and enrolled in the database, the total images in the database were $(1 \times 86 = 86)$, and one image has been used as input to simulate the identification operation. According to Equation 11, in ORB method give 86 cattle out of 86 have been correctly identified which achieves equivalent identification accuracy value as 100%. It is worth notice that the average consumed feature extraction time is $12.23s$ and the average individual matching time is $0.01s$, in Ratio test method give 72 cattle out of 86 have been correctly identified which achieves equivalent identification accuracy value as 83.72%. It is worth notice that the average consumed feature extraction time is $9.11s$ and the average individual matching time is $0.01s$, in FLANN method give 80 cattle out of 86 have been correctly identified which achieves equivalent identification accuracy value as 93.02%. It is worth notice that the average consumed feature extraction time is $8.74s$ and the average individual matching time is $0.08s$, including RANSAC optimization, which are consistent with Figure 15 and Figure 16.

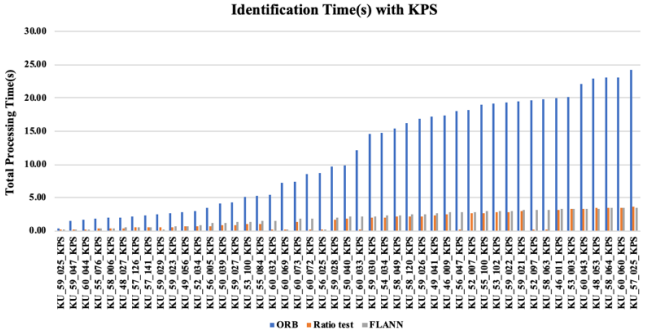


Figure 19: Identification Time(s)with KPS.

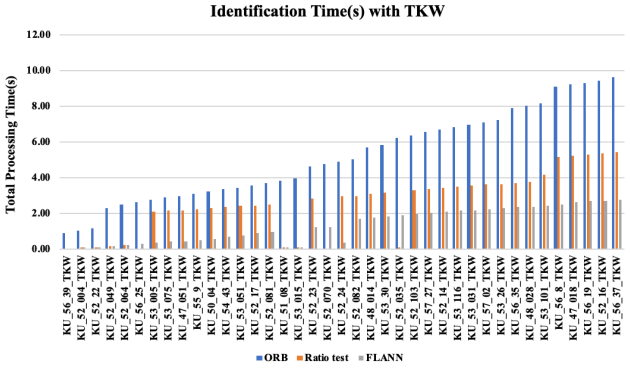


Figure 20: Identification Time(s)with TKW.

However, feature extraction time and matching time are considered very short in single point of feature extraction and matching operation. The identification time of each query cattle is shown in Figure 19 shows the identification time(s) with KPS, in ORB method give the total identification time still long, around $\approx 25s$, in Ratio test method give the total identification time still long, around $\approx 4s$, in FLANN method give the total identification time still long, around $\approx 4s$, at maximum. Figure 20 shows the

identification time*(s)* with TKW, in ORB method give the total identification time still long, around ≈ 10*s*, in Ratio test method give the total identification time still long, around ≈ 6*s*, in FLANN method give the total identification time still long, around ≈ 3*s*. Such a linear database search has been used identification time is based on the index of the template in the database. The template is matches by similarity score with score threshold to candidate list. Finally, identified is determined according to the best score and confirmation by Cow Tag.

For identification status, an image naming scheme works as template_XXX, whereas XXX is the image order (1 to 86) by enrolled with Cow Tag. The identified status with ORB shows that great identified of 86 cattle all correctness. The identified status with Ratio test shows that identified of 14 cattle false with Cow Tag {Query:

[6] KU_56_25_TKW, [16] KU_51_08_TKW,

[17] KU_53_015_TKW, [19] KU_52_070_TKW,

[24] KU_52_035_TKW, [42] KU_60_044_KPS,

[57] KU_60_032_KPS, [58] KU_60_069_KPS,

[60] KU_60_072_KPS, [61] KU_56_025_KPS,

[64] KU_60_033_KPS, [72] KU_56_047_KPS,

[78] KU_52_097_KPS, [79] KU_58_063_KPS}.

The identified status with FLANN shows that identified of 6 cattle false with Cow Tag {Query:

[16] KU_51_08_TKW, [17] KU_53_015_TKW,

[20] KU_52_24_TKW, [48] KU_59_029_KPS,

[58] KU_60_069_KPS, [61] KU_56_025_KPS}.

The identification status shows incorrect identified because the similarity score is less than the defined score threshold.

High performance evaluation in ORB is 100% identified, the incorrect identified cattle is considered as false matched or false accepted input because the match occurred with a template that does not correspond to the query image. The FAR in this case is Ratio test = 16.28%, FLANN = 6.98%, and it equal to the identification ER. The relation between FAR, FRR, and ERR are determined according to the similarity threshold. Figure 21 shows the FAR of Ratio test versus FRR related to the similarity threshold, the ERR is shown as the cross point between FAR and ERR. ERR is ≈ 0.18 with threshold is ≈ 35.0. Figure 22 shows the FAR of FLANN versus FRR related to the similarity threshold, the ERR is shown as the cross point between FAR and ERR. ERR is ≈ 0.007 with threshold is ≈ 183.0.

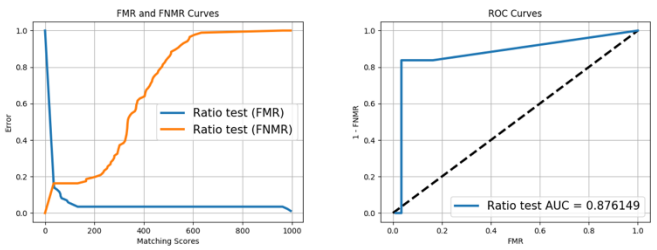


Figure 21:(a)FMR and FNMR curves and (b)ROC curves of Ratio test.

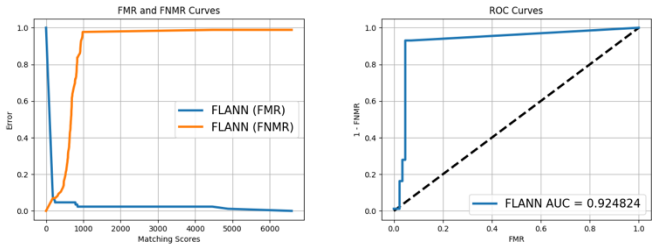


Figure 22:(a)FMR and FNMR curves and (b)ROC curves of FLANN.

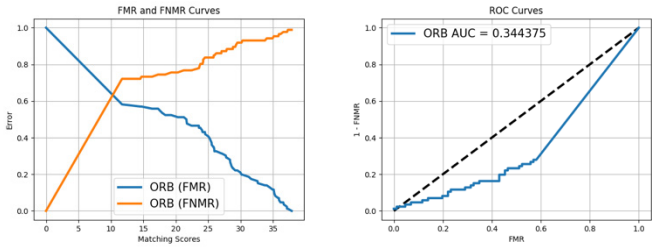


Figure 23:(a)FMR and FNMR curves and (b)ROC curves of ORB with Nearby BS.

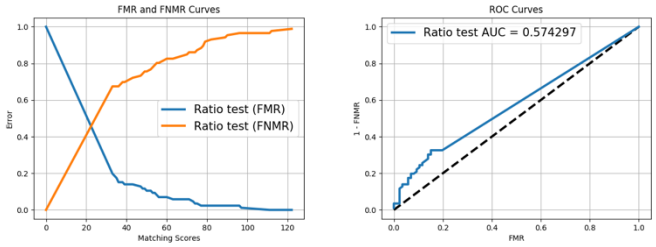


Figure 24:(a)FMR and FNMR curves and (b)ROC curves of Ratio test with Nearby BS.

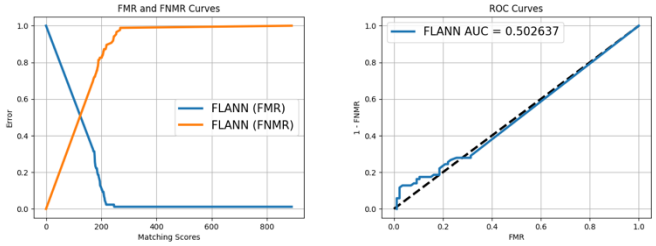


Figure 25:(a)FMR and FNMR curves and (b)ROC curves of FLANN with Nearby BS.

Nearby BS image results, Figure 23 shows the FAR of ORB versus FRR related to the similarity threshold, the ERR is shown as the cross point between FAR and ERR. ERR is ≈ 0.63 with threshold is ≈ 11.0. Figure 24 shows the FAR of Ratio test versus FRR related to the similarity threshold, the ERR is shown as the cross point between FAR and ERR. ERR is ≈ 0.44 with threshold is ≈ 23.0. Figure 25 shows the FAR of FLANN versus FRR related to the similarity threshold, the ERR is shown as the cross point between FAR and ERR. ERR is ≈ 0.52 with threshold is ≈ 120.0, because some of image from cattle not clearer.

Table 3 show ORB method is the best performance over Ratio test and FLANN method in term of performance evaluation.

Table 3.Comparison of three matcher methods.

Process	Methods		
	ORB	Ratio test	FLANN
Extraction Time	Normal	Good	Best
Matching Time	Best	Good	Normal
Number of Features	Low	Normal	High
Threshold	Low	Normal	High
Identification Time	Normal	Good	Best
Performance Evaluation	Best	Normal	Good

6. Conclusion

Kamphaeng Saen Beef Cattle Identification Approach using Muzzle Print Image was developed with SIFT feature extraction and matching. The identification scenarios considered a dimension of 200 × 200 pixels, which collected 765 images from 86 cattle (KPS; 47 cattle, 391 images and TKW; 39 cattle, 374 images). The muzzle print images of each cattle were swapped between the enrolment and the identification phase. The ORB method shown the best performance over Ratio test and FLANN method in term of performance evaluation. In order to evaluate the robustness of the scheme, the collected images cover different deteriorating factors. The superiority of the presented scheme comes from the coupling of SIFT with RANSAC as a robust outlier removal algorithm. The achieved identification accuracy is given 92.25%. Therefore, the proposed of muzzle print images can be applied to register the Kamphaeng Saen beef cattle for breeding and marking systems. In the future work, some machine learning techniques should be developed for Sire and Dam of Kamphaeng Saen beef cattle identification in Thailand. Additionally, the real time identification by using is smartphone also challenging.

Acknowledgment

This research was funded by College of Industrial Technology, King Mongkut's University of Technology North Bangkok (Grant No. Res-CIT0246/2020). We would like to thank the National Science and Technology Development Agency, Pathum Thani. Cowboy Land and KUBeef, Nakhon Pathom, Department of Animal Science, Faculty of Agriculture, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom, Department of Animal Science, Faculty of Agriculture, Kasetsart University, Bangkhen, Bangkok and Tubkwang Research Center Saraburi, Thailand for supporting data.

References

[1] P. Nilchuen, S. Rattanatabtimtong, S. Chomchai, "Superrovulation with Different Doses of Follicle Stimulating Hormone in Kamphaeng Saen Beef Cattle," Songklanakarin Journal of Science and Technology, **33**(6), 679-683, 2011, doi: 10.3923/javaa.2012.676.680.

[2] C. Chantalakhana, B. Rengsirikul, P. Prucsasri, "A report on performance of thai indigenous cattle and their crossbred from American Brahman and Charolais sires," Department of Animal Science, Faculty of Agriculture, Kasetsart University, Bangkok, Thailand, 11(4), 287-295, 1978, https://agris.fao.org/agris-search/search.do?recordID=TH19800531313. [Accessed 3 June 2020].

[3] S. Tumwasorn, K. Markvichite, P. Innurak, P. Prucsasri, C. Chantalakhana, S. Yimmongkol, P. Chitprasan, "Heterosis and additive breed effects on growth traits from crossing among Thai local, Charolais and American Brahman under Thai conditions," Department of Animal Science, Faculty of Agriculture, Kasetsart University, Bangkok, Thailand, **28**(1), 245-255, 1993, https://agris.fao.org/agris-search/search.do?recordID=TH1998000100. [Accessed 3 June 2020].

[4] P. Boonsaen, N. W. Soe, W. Maitreejet, S. Majarune, T. Reungprim, S. Sawanon, "Effects of protein levels and energy sources in total mixed ration on feedlot performance and carcass quality of Kamphaeng Saen steers," Agriculture and Natural Resources, **51**(1), 57-61, 2017, doi: 10.1016/j.anres.2017.02.003.

[5] P. Prucsasri. "Kamphaengsaen Beef Cattle," 3rd Edition, (in Thai) Kamphaengsaen Beef Breeders Association, Nakhon Pathom, Thailand, **57**, 1977.

[6] P. Bunyavejchewin, S. Sangdid, K. Hansanet, "Potential of beef production in tropical Asia," Proceedings of the 8th AAAP Animal Science Congress, Tokyo, Japan, Japan Society of Zootechnical Science, **1**, 404-403, 1996.

[7] P.Prucsasri, "Kamphaengsaen Beef Cattle," Neon Book Media, (in Thai), Nonthaburi, Thailand, 2015.

[8] P. Innurak, S. Yimmongkol, P. Skunmun, "Kamphaeng Saen synthetic Thai beef cattle breed: Its development, characteristics and prospects," Proceedings of the 11th AAAP Animal Science Congress, Kuala Lumpur, Malaysia, **2**, 51-53, 2004, https://agris.fao.org/agris-search/search.do?recordID=MY2014001336 [Accessed 10 July 2020].

[9] M. Neary and A. Yager. "Methods of Livestock Identification," Farm Animal Management@Purdue, Department of Animal Sciences, Purdue University, 1-9, 2002, https://www.extension.purdue.edu/extmedia/as/as-556-w.pdf. [Accessed 15 July 2020]

[10] R. -W. Bello, A. Z. H. Talib and A. S. A. B. Mohamed. "Deep Belief Network Approach for Recognition of Cow using Cow Nose Image Pattern," Walailak Jounal of Science and Technology (WJST), **18** (5), 2021, doi: doi: 10.48048/wjst.2021.8984.

[11] R. W. Bello, D. A. Olubummo, Z. Seiyaboh, O. C. Enuma, A. Z. Talib, A. S. A. Mohamed, "Cattle Identification: The History of Nose Prints Approach in Brief," The 6th International Conference on Agricultural and Biological Sciences, Conference Series: Earth and Environmental Science, **594**, 1-8, 2020, doi:10.1088/1755-1315/594/1/012026.

[12] A. K. Jain, A. Ross, S. Prabhakar, "An Introduction to Biometric Recognition," IEEE Transaction on Circuits and Systems for Video Technology. **14** (1), 4-20, 2004, doi:10.1109/TCSVT.2003.818349.

[13] B. Ebert. "Identification of Beef Animals," Alaba A&M and Aubum Universities, 1-8, 2006, https://ssl.acesag.auburn.edu/pubs/docs/Y/YANR-0170/YANR-0170-archive.pdf. [Accessed 17 July 2020]

[14] H. Minagawa, T. Fujimura, M. Ichiyanagi, K. Tanaka, M. Fang-quan, "Identification of Beef Cattle by Analyzing Images of their Muzzle Patterns Lifted on Paper," Proceeding of the Third Asian Conference for Information Technology in Agriculture, Asian Agricultural Information Technology & Management, Beijing, China, **28**(7), 596–600, 2002, https://eurekamag.com/research/003/801/003801309.php. [Accessed 24 July 2020]

[15] B. Barry, U. A. Gonzales-Barron, K. McDonnell, F. Butler, S. Ward, "Using Muzzle Pattern Recognition as a Biometric Approach for Cattle Identification," Transactions of the American Society of Agricultural and Biological Engineers, **50**(3), 1073-1080, 2007, doi: 10.13031/2013.23121.

[16] C. Sian, W. Jiye, Z. Ru, Z. Lizhi, "Cattle Identification using Muzzle Print Images based on Feature Fusion," The 6th International Conference on Electrical Engineering, Control and Robotics, Xiamen, China, IOP Conference Series: Materials Science and Engineering, **853**, 1-7, 2020, doi:10.1088/1757-899X/853/1/012051.

[17] A. I. Awad, M. Hassaballah, "Bag-of-Visual-Words for Cattle Identification from Muzzle Print Images," Applied Sciences, **9** (22), 1-12, 2019, doi:10.3390/app9224914.

[18] A. Noviyanto, A. M. Arymurthy, "Automatic Cattle Identification Based on Muzzle Photo Using Speed-Up Robust Features Approach," Recent Advances in Information Science, 110–114, 2012, http://www.wseas.us/e-library/conferences/2012/Paris/ECCS/ECCS-17.pdf. [Accessed 4 August 2020]

[19] A. Tharwat, T. Gaber, "Cattle Identification using Muzzle Print Images based on Texture Features Approach," the 5th International Conference on Innovations in Bio-Inspired Computing and Applications IBICA, Advances in Intelligent Systems and Computing, **303**, 217-227, 2014, doi:10.13140/2.1.3685.1202.

[20] H. A. Mahmoud, H. M. R. E. Hadad, "Automatic Cattle Muzzle Print Classification System Using Multiclass Support Vector Machine," International Journal of Image Mining, **1**(1), 126-140, 2015, doi:10.1504/IJIM.2015.070022.

[21] T. M. Gaber, "Precision Livestock Farming: Cattle Identification Based on Biometric Data," Faculty of Computers and Informatics Suez Canal University, Faculty of Agriculture, Ismailia, Egypt, 1-17, 2014, https://www.slideshare.net/Tarekgaber/precision-livestock-farming-cattle-identification-based-on-biometric-data-tarek-gaber-40692911.[Accessed 10

September 2020].
[22] H. M. El-Bakry, I. El-Hennawy, H. E. Hadad, "Bovines Muzzle Identification Using Box-Counting," International Journal of Computer Science and Information Security, **12**(5), 29-34, 2014, https://www.researchgate.net/publication/303960881_Bovines_Muzzle_Identification_Using_Box-Counting/stats. [Accessed 15 September 2020]
[23] D. G. Lowe, "Object Recognition from Local Scale-Invariant Features," the 7th IEEE International Conference on Computer Vision, Kerkyra, Corfu, Greece, 1150–1157, 1999, doi:10.1109/ICCV.1999.790410.
[24] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, **60**(2), 91–110, 2004, doi:10.1023/B:VISI.0000029664.99615.94.
[25] E. Erwin, N. N. Br. Karo, A. Y. Sasi, N. Aziza, H. K. Putra, "The Enhancement of Fingerprint Images using Gabor Filter," Journal of Physics Conference Series, 1196(1), 2019, doi:10.1088 /1742-6596/1196/1/012045.
[26] M. A. Fischler, R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Communications of ACM, **24**(6), 381– 395, 1981, doi:https://doi.org/10.1145/358669.358692.
[27] A. Jakubović, J. Velagić, "Image Feature Matching and Object Detection using Brute-Force Matchers," International Symposium ELMAR, Zadar, 83-86, 2018, doi: 10.23919/ ELMAR. 2018. 8534 641.
[28] G. Guo, H. Wang, D. A. Bell, Y. Bi, K. Greer, "KNN Model-Based Approach in Classification," On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. OTM 2003. Lecture Notes in Computer Science, 2888, 986-996, 2003, doi:10.1007/978-3-540-39964-3_62.
[29] M. Muja, D.G. Lowe, "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration," Proceedings of the 4th International Conference on Computer Vision Theory and Applications, Lisboa, Portugale, **1**, 331-340, 2009, doi:10.5220/00 01787803310340.
[30] A. I. Awad, K. Baba, "Evaluation of a Fingerprint Identification Algorithm with SIFT Features," Proceedings of the 3rd International Conference on Advanced Applied Informatics, Fukuoka, Japan, 129–132, 2012, doi: 10.1109/IIAI-AAI.2012.34.
[31] D. Maltoni, D. Maio, A. K. Jain, S. Prabhakar, "Handbook of Fingerprint Recognition, Second Edition," Springer-Verlag, 1-494, 2009.
[32] R. Giot, M. El-Abed, C. Rosenberger, "Fast Computation of the Performance Evaluation of Biometric Systems: Application to Multibiometrics," Future Generation Computer Systems, Special Section: Recent Developments in High Performance Computing and Security, **29**(3), 788–799, 2013, doi: 10.1016/ j.future. 2012.02.003

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 123-129 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Business Intelligence Budget Implementation in Ministry of Finance (As Chief Operating Officer)

Banir Rimbawansyah Hasanuddin*, Sani Muhammad Isa

Department of Computer Science, BINUS Graduate Program, Bina Nusantara University, Jakarta, 11480, Indonesia

ARTICLE INFO

Article history:
Received: 26 April, 2021
Accepted: 01 July, 2021
Online: 10 July, 2021

Keywords:
Business Intelligence
Data Mining
Data Warehouse
Data Analysis
Budget Implementation

ABSTRACT

The Ministry of Finance is the state ministry in charge of state financial affairs which has two functions, namely the Chief Financial Officer (CFO) as the State General Treasurer and the Chief Operating Officer (COO) as a Budget User. As COO, the Ministry of Finance is expected to be able to provide information related to budget implementation to leaders quickly and accurately. The problem that occurs is the implementation information is still done manually, so it takes time to process. In addition, there is no information regarding budget predictions for the next semester or year. This study uses Business Intelligence (BI) as a technique in the process of building budget execution information. The Business Intelligence Roadmap is a methodology used to produce budget implementation information in the form of a dashboard. To see the prediction of the realization of the budget for the next semester or year using the forecasting method with the neural network model. the Results is budget implementation information can be accessed easily and has accurate data and can provide information to the leaders as supporting material in making decisions at the Ministry of Finance.

1. Introduction

Ministry of Finance is public sectors that in charge of financial affairs and state wealth. Ministry of Finance has a dual role in terms of the power of managing state finances, first as Chief Financial Officer (CFO) who has the duty of the State General Treasurer (BUN). Second, the Chief Operating Officer (COO) who has a duty as a Budget User. Ministry of Finance as the COO has twelve echelon I unit that have responsibility to formulating ministry strategies, preparing work plans and budgets, using resources efficiently and effectively, reporting on the performance and use of available resources, and evaluating performance results.

Based on the above responsibilities, the Ministry of Finance as COO is expected to be able to provide information on budget implementation in a transparently to the leaders. Figure 1 is the process of how to present data to the leaders. Based on Figure 1, the problem occurs is, it takes a long time and process to produce budget implementation information

Besides that, there is no accurate data available as supporting material for the leaders to makes the policies. The length of a process in data processing because it i is worked manually using Microsoft Excel. The downloaded data then filtered according to the needs to be presented. The data that has been processed then presented on each sheet, so to see the results, it must be clicked one by one on the available sheets. This job can reduce the speed of time in providing information and difficulty of processing data if there are employee mutations.

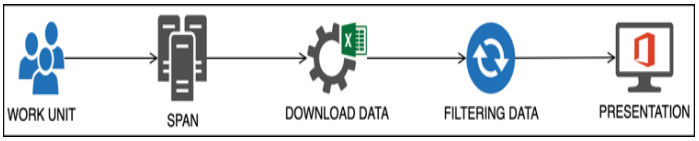


Figure 1: Process of Presenting Data

The right solutions for processing, analysis and presenting data is using Business Intelligence (BI). BI is a tools, technologies and solutions to extracting business information from a set of data [1]. Ministry of Finance has data that can be used to run BI, where the data is sourced from another system. The system is an integrated system of all processes related to the management. Recently the term "Business Intelligence" is referred to as "Business Analytics" [2]. Business Intelligence also provides stages and steps within a generate useful information for an organization.

This paper aims to produce budget implementation information using business intelligence which will be presented in the form of a dashboard and provide a data visualization of future prediction budget implementation trends.

*Corresponding Author: Banir Rimbawansyah Hasanuddin, banir.hasanuddin@binus.ac.id

2. Related Works

The role of Business Intelligence is constantly changing from what was previously only seen as an analytical application, now it is considered very important for organizational strategy. BI tools are considered a technology which results in efficient business operations by adding value to the company [3]. BI can help managers monitor and analysis quickly and efficiently [4]. BI is also the process an organizations or company take advantage of virtual and digital technology to collect, manage and then analysis data [5]. In another definition, BI is an applications and technologies for gathering, storing, cleanse, analysis, and providing access to data to help managers or leaders make sound business decisions on sound time [6].

According to [7], the most regularly used analysis are cross selling and up selling, customer segmentation and profiling, parameters of interest, survival time, customer loyalty and customer switching, credit assessment, fraud detection, logistics optimization, business process forecasts, service performance appraisals internet and internet content analysis.

BI projects are organized according to the same six stages common to every engineering project. Within each engineering stage, certain steps are carried out to see the engineering project through to its completion. Business intelligence roadmap describes sixteen development steps within six stages such as Justification Stage, Planning Stage, Business Analysis Stage, Design Stage, Construction Stage and Deployment Stage as in Figure 2 [8].

Data warehouse has a role as a data source in developing Business Intelligence. According [9], Data warehouse is a collection of data based on subject-oriented, integrated, not easy to change and datasets consist of varying times in support of management decisions. In a data warehouse schema, it usually consists of one fact table and several dimension tables, where the dimension table contains a more detailed description of the fact table [10]. In [11], the author said, some of the benefits provided by the data warehouse directly, that is users can perform extensive data analysis in various ways, consolidated data presentation, timely and better information, improved system performance results, and simplified data access. Extract, transform, loading (ETL) is a data integration framework that involves extracting data from data management systems and then cleaning it, transforming it according to business needs, and finally loading it into a database [12].

Data mining is a technology that is very useful in extracting helpful knowledge within hidden data collections [13]. In [14] the author state that data mining can be showed as a result of the natural evolution of information technology. Argue of [15] , said that data mining combined statistical analysis, machine learning techniques and database management in extracting forms from large databases. Other than that, data mining requires intensive computation for comparative data analysis [16].

Classification data mining is divided into two categories, that is predictive and descriptive [17]. In [18], the author states that predictive analysis is used to determine the future outcome of an event or possible situation, but it can also be used to automatically analysis large amounts of data with different variables. While the descriptive is presented in a short / summary form of data points and the main character is the data set [19].

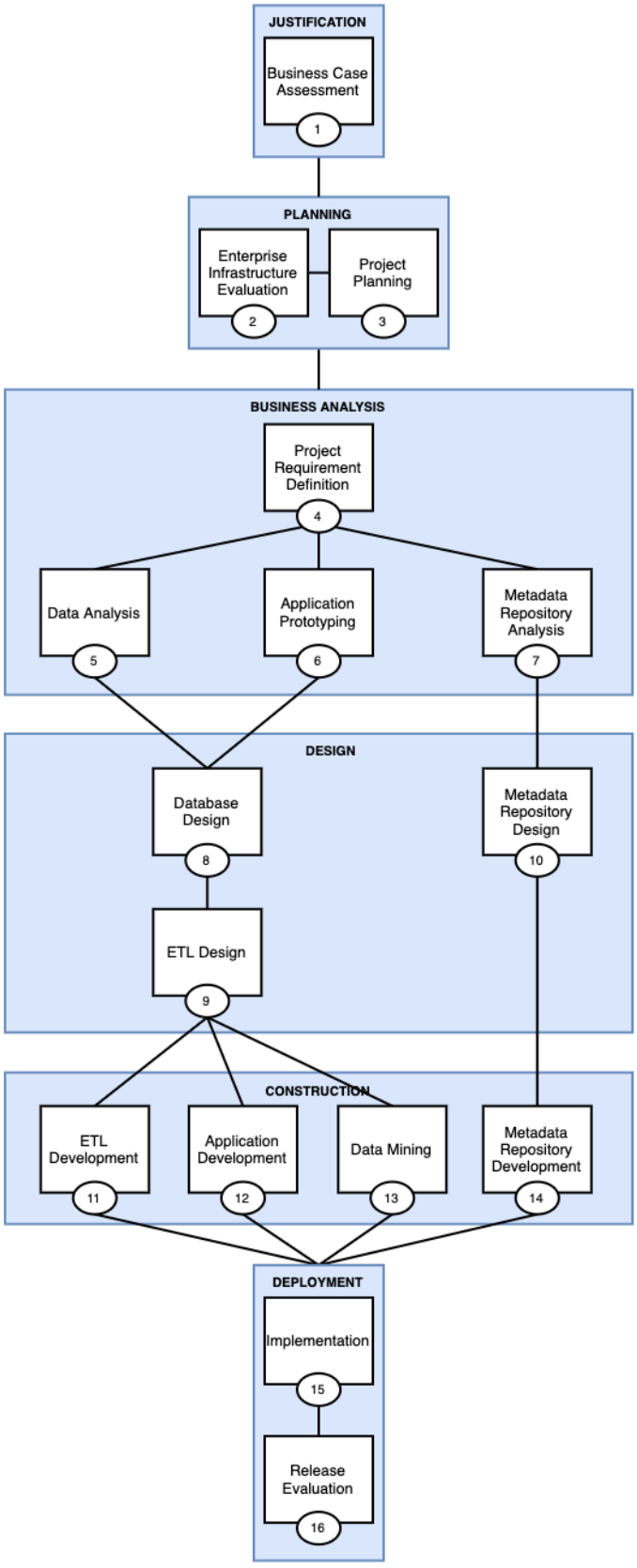


Figure 2: Business Intelligence Roadmap

One of the data mining techniques in conducting predictive analysis is Time Series Forecasting. Time Series Forecasting is a process that uses a model to predict future events based on known time [20]. This model is used because it develops a mathematical explanation that is similar to the biological processes of neurons [21]. Neural Network also has the ability to select all possibilities between variables and this technique is one of the best prediction methods [22]. The advantages of a Neural Network among others (1) has high accuracy for complex non-linear mapping approximations, (2) very flexible with noisy data, (3) not making priori assumptions about the distribution of the data (4) easy to update with new data and dynamic environment, (5) can be implemented in parallel hardware, (6) if there is a failure, it can proceed without problems due to its parallel nature.

The neural network can make an effective forecast for the financial market and the data can be taken directly from the Internet to provide real-time and off-line data processing and analysis [23].

3. Research Methodology

This research stage consists of thirteen steps. These stages have been simplified previously according to the business intelligence roadmap method. Figure 3 is a simplification step to produce business intelligence to be more effective and efficient. These steps are explained as follows.

3.1. Justification

Identify business needs. Determine business requirements, assessment decision-making solutions, competitor software that uses business intelligence, determine business intelligence application objectives, provide business intelligence solutions, perform risk measurement.

3.2. Planning

Plan the development of the project that will be completed and deployed. Determine technical specifications required for BI development, the source of data to be obtained, determine level of Critical success factor (CSF) and the project management level.

3.3. Business Analysis

Business analysis can help to formulating problems that will be developed of BI, with determining what results are desired from business analysis, such as the subject area, time, stage stages, detailed data and even what external data is needed to answer these business questions. Then [24] also argues that the business analysis approach is used as a quick decision making, where all stakeholders are involved through open discussions.

3.4. Design

Understand solutions to business problems or enables the business opportunity. Activities performed that is design BI database, monitoring and tuning database and query designs, design ETL process flow, set up the staging area.

3.5. Construction

Develop the product, which should provide a return on investment within a predefined time frame. Activities performed that is build and testing ETL process, build and testing the application program, datamining such as determine topology and activation function, perform initialization.

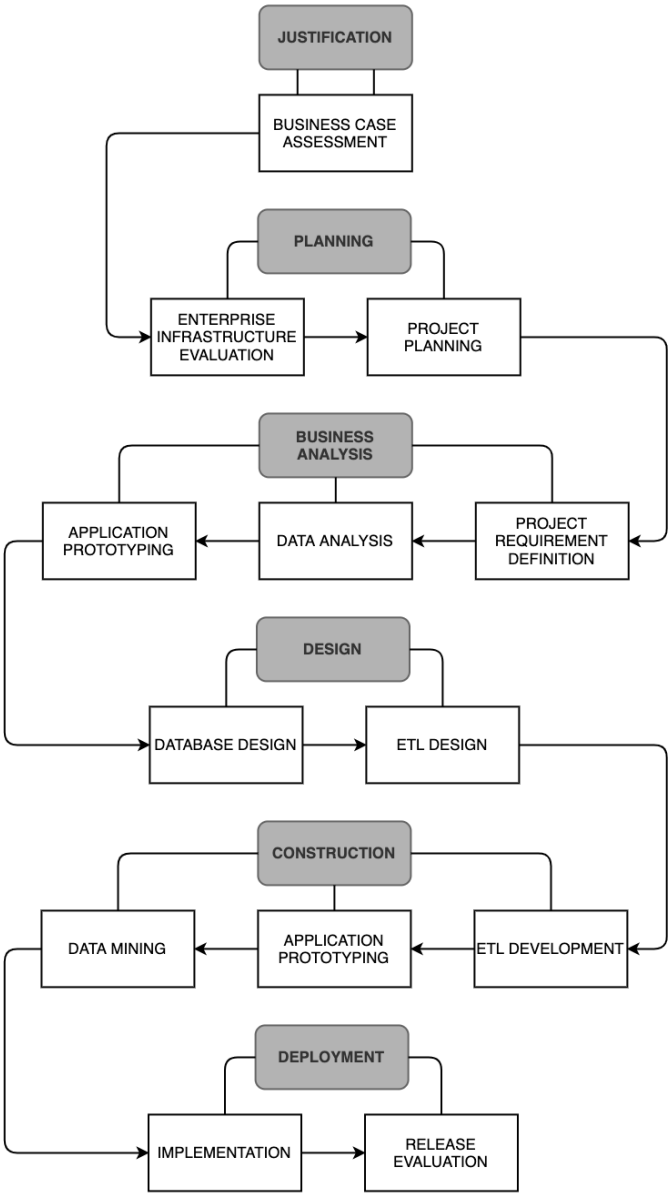


Figure 3: Research Steps

3.6. Deployment

Implement and finished the product, and then measure its effectiveness to determine whether the solution satisfy, exceeds, or fails within the expected return of investment. Activities performed that is planning for implementation, load the production database, set up the supporting, preparing a post implementation reviews, follow-up of meeting result after implementation.

4. Analysis and Result

Figure 4 is constellation schema because there is allocation and spending fact which correlated to dimension of register, fund, time, central, branch and region. A data warehouse is identified as a constellation, if the fact tables are linked [25]. In other words, constellations are schemes that have two or more facts connected to other dimensions. The source of the database is from the data

warehouse which is backed up every day and then performs the process of extract, transform, and load (ETL). The data is stored in the BI database, namely DWDashboardBI. Figure 5 describes the ETL design flow from the source data (database source) to the destination database (database target).

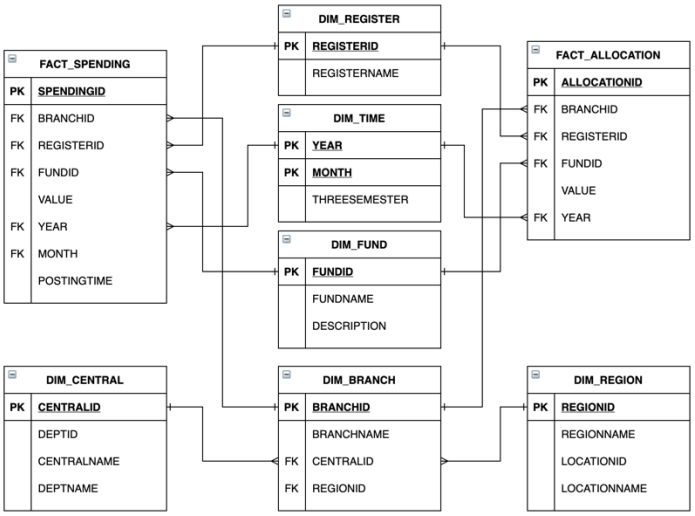


Figure 4: Constellation Schema

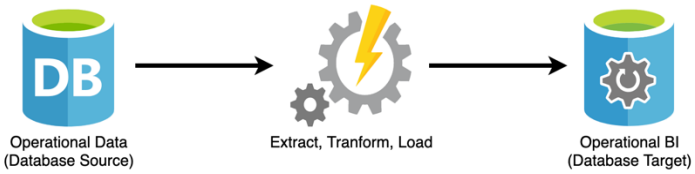


Figure 5: ETL Design

In data warehouse, several activities performed in the ETL process such as extracting, cleaning, conforming tables from and loading them into data warehouse [26]. The ETL process can be seen in Figure 6-13 and the software used to create ETL is Microsoft Visual Studio.

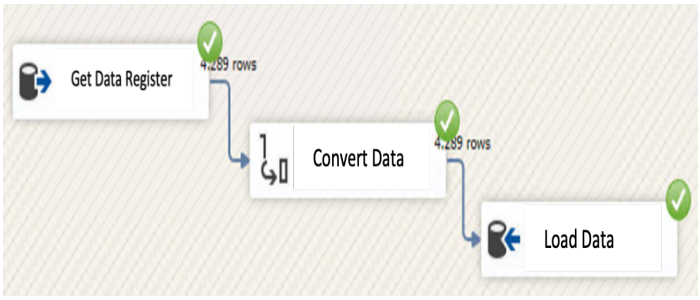


Figure 6: ETL Register Dimension

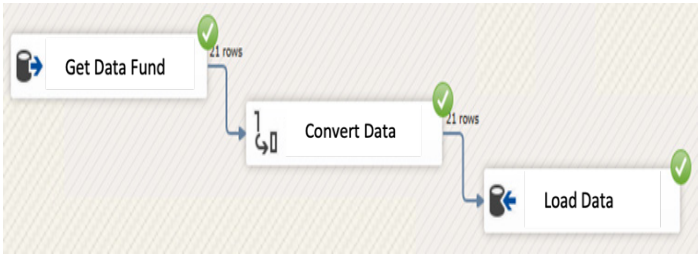


Figure 7: ETL Fund Dimension

Figure 6 describes the process of getting data from the M_Register table and then saving the data to the Dim_Register table.

Figure 7 describes the process of getting data from the M_Fund table and then saving the data to the Dim_Fund table.

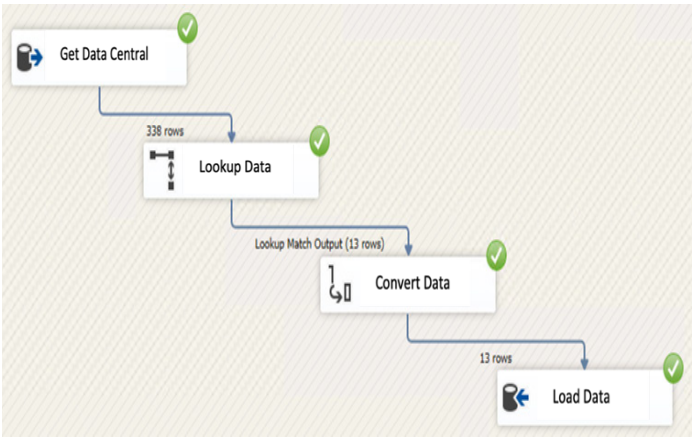


Figure 8: ETL Central Dimension

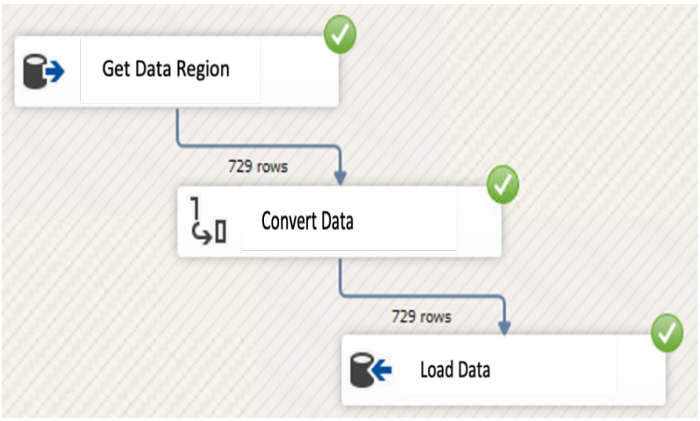


Figure 9: ETL Region Dimension

Figure 8 describes the process of getting data from the M_Central table and then saving the data to the Dim_Central table.

Figure 9 describes the process of getting data from the M_Region table and then saving the data to the Dim_Region table.

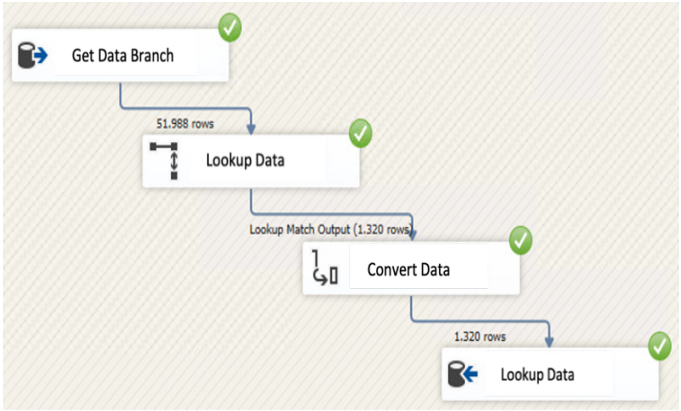


Figure 10: ETL Branch Dimension

Figure 10 describes the process of getting data from the M_Branch table and then saving the data to the Dim_Branch table.

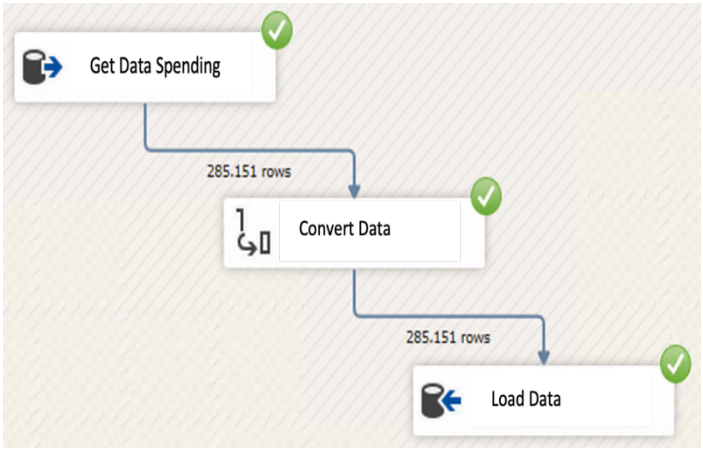


Figure 11: ETL Spending Fact

Figure 11 describes the process of getting data from the T_Spending table and then saving the data to the Fact_Spending table.

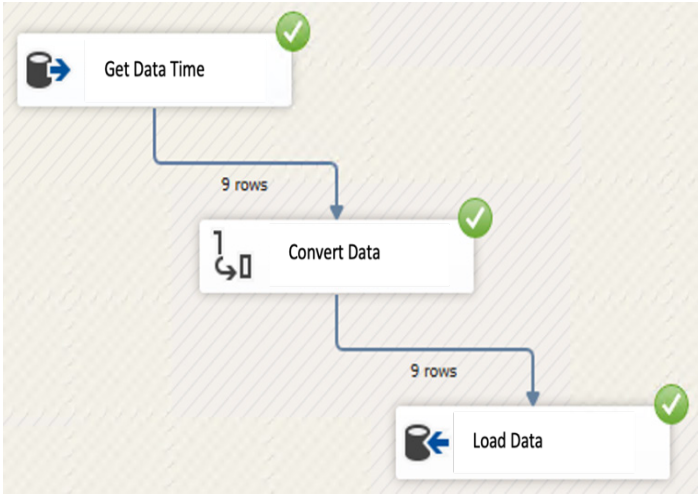


Figure 12: ETL Time Dimension

Figure 12 describes the process of getting data from the M_Time table and then saving the data to the Dim_Time table.

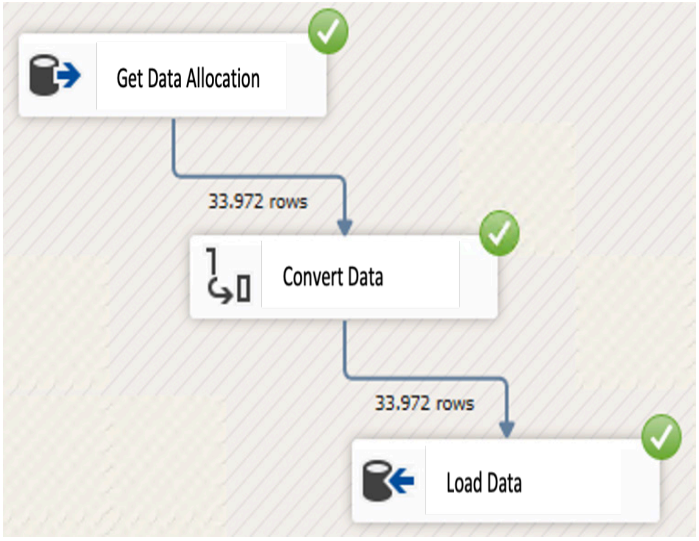


Figure 13: ETL Allocation Fact

Figure 13 describes the process of getting data from the T_Allocation table and then saving the data to the Fact_Allocation table.

Figure 14: Dashboard BI Report

After ETL process is executed, the user can see the report from existing data in data warehouse. This report can help the leaders to analysis data about budget implementation in current years. Figure 14 is dashboard BI report which can be used as an overview in decision making. There are five headlines, such as Budget Implementation until now which is shown with speedometer chart, realization of budget implementation by expenditure, Trend of spending monthly in year on year, budget implementation year on year by expenditure and spending budget by region in map chart. This dashboard was made using PHP with CodeIgniter framework, jQuery and CSS.

The use of the php programming language with the CodeIgniter framework, CSS and jQuery makes it easy to build a dashboard because it is a programming language that is easy to learn, open source, has a large community, easy to maintain and develop rapidly.

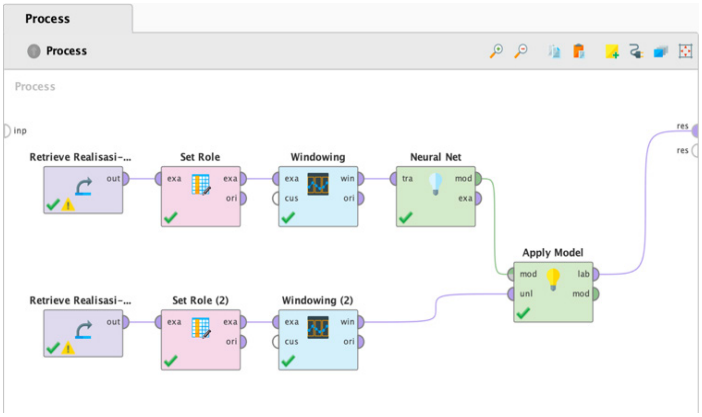


Figure 15: Figure Time Series with Neural Network Model

The Forecasting method is a method used to produce predictions of budget implementation values for the following semester and RapidMiner is a tool for performing this forecasting method. This method is used because the data contains time series information. While the algorithm model used is Neural Network as in Figure 15. Figure 16 shows the results of the prediction table which is divided into 4 columns according to the windowing parameters that have been set.

Row No.	Window id	TOTAL_REALISA...	prediction(TOTAL_R...	TGLPOST - 3	TGLPOST - 2	TGLPOST - 1	TGLPOST - 0	TOTAL_REALISASI - 3
1	0	[illegible]	6882128511748.463	Mar 31, 2017	Jun 30, 2017	Sep 30, 2017	Dec 31, 2017	[illegible]
2	1	[illegible]	16666095139742.078	Jun 30, 2017	Sep 30, 2017	Dec 31, 2017	Mar 31, 2018	[illegible]
3	2	[illegible]	28350540052128.875	Sep 30, 2017	Dec 31, 2017	Mar 31, 2018	Jun 30, 2018	[illegible]
4	3	[illegible]	38549766357374.860	Dec 31, 2017	Mar 31, 2018	Jun 30, 2018	Sep 30, 2018	[illegible]
5	4	[illegible]	9009117494467.275	Mar 31, 2018	Jun 30, 2018	Sep 30, 2018	Dec 31, 2018	[illegible]
6	5	[illegible]	21037868329446.164	Jun 30, 2018	Sep 30, 2018	Dec 31, 2018	Mar 31, 2019	[illegible]
7	6	[illegible]	34603761224431.625	Sep 30, 2018	Dec 31, 2018	Mar 31, 2019	Jun 30, 2019	[illegible]
8	7	[illegible]	41428140651032.910	Dec 31, 2018	Mar 31, 2019	Jun 30, 2019	Sep 30, 2019	[illegible]
9	8	[illegible]	11306134775737.035	Mar 31, 2019	Jun 30, 2019	Sep 30, 2019	Dec 31, 2019	[illegible]
10	9	[illegible]	24023390592252.060	Jun 30, 2019	Sep 30, 2019	Dec 31, 2019	Mar 31, 2020	[illegible]
11	10	[illegible]	35563079477494.010	Sep 30, 2019	Dec 31, 2019	Mar 31, 2020	Jun 30, 2020	[illegible]
12	11	?	41640478493418.020	Dec 31, 2019	Mar 31, 2020	Jun 30, 2020	Sep 30, 2020	[illegible]
13	12	?	16818679449688.310	Mar 31, 2020	Jun 30, 2020	Sep 30, 2020	Dec 31, 2020	[illegible]
14	13	?	34795797065856.066	Jun 30, 2020	Sep 30, 2020	Dec 31, 2020	Mar 31, 2021	[illegible]
15	14	?	45624223524732.810	Sep 30, 2020	Dec 31, 2020	Mar 31, 2021	Jun 30, 2021	[illegible]
16	15	?	35096271940404.094	Dec 31, 2020	Mar 31, 2021	Jun 30, 2021	Sep 30, 2021	?

Figure 16: Result of Prediction Table

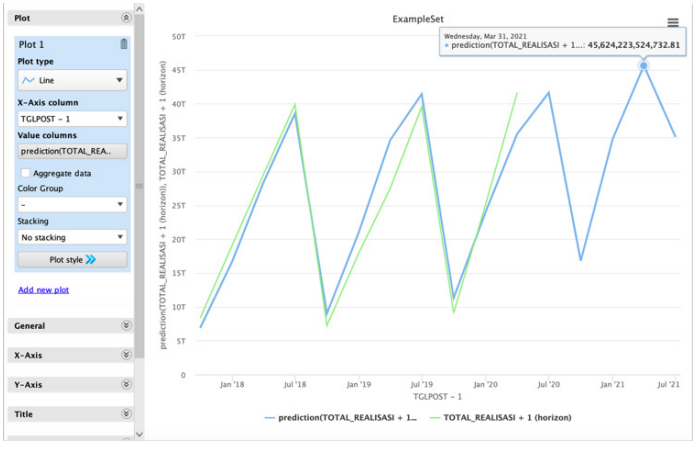


Figure 17: Chart of Spending and Allocation Prediction

In Figure 17 is a graph of its expenditure and prediction. The graph was shown in the form of a line graph by juxtaposing data on total expenditure and predicted of total expenditure. The green line represents the total expenditure data and the blue one is the prediction result. When we see in plain view, the result of predictions are close to with actual of total expenditure.

5. Conclusion

Based on the results of the data analysis which has been done, the following concluded:

- The Business Intelligence dashboard which was developed at the Ministry of Finance as COO, it can be known the trend of total budget implementation, per type of expenditure, and by region or province in Indonesia.
- The Business Intelligence dashboard can also provide predictions about the budget implementation that will happen for the following years, so that can help the leaders to analysis data descriptively within decision making appropriately.
- The Business Intelligence dashboard can speed-up the process of presenting data quickly and can be accessed anywhere.

The recommendation to the next developments of BI dashboard are:

- Can provide insight to the leaders about prediction of the budget ceiling by unit of echelon 1 with include the value of inflation that happened in Indonesia.
- Linked Business Intelligence with performance data for unit of echelon 1, so that can see the relationship between the result of performance values with the amount of budget received.

Conflict of Interest

The authors declare that there is no conflict of interests on this paper.

Acknowledgment

I am especially grateful for Dr. Sani Muhammad Isa for his advice in making this paper and I also wish to express my deep thanks to Department of Computer Science Bina Nusantara University for their kindness and helps to my studies.

References

[1] B. Moçka, G. Beqiraj, D. Leka, "Evaluation of Business Intelligence Maturity Level in Albania Banking Systems," International Journal of Advanced Technology and Engineering Exploration ISSN, (7), 2394–5443, 2015.

[2] D. Delen, G. Moscato, I.L. Toma, "2018 International Conference on Information Management and Processing, ICIMP 2018," 2018 International Conference on Information Management and Processing, ICIMP 2018, **2018-Janua**, 49–53, 2018.

[3] J. Ranjan, "Business Intelilgence: Concepts, components, techniques and benefits," Journal of Theorical and Applied Information Technology, **9**(1), 60–70, 2009.

[4] V. Khatibi, A. Keramati, G.A. Montazer, "A Business Intelligence Approach to Monitoring and Trend Analysis of National R&D Indicators," EMJ - Engineering Management Journal, **29**(4), 244–257, 2017, doi:10.1080/10429247.2017.1380578.

[5] S. Rouhani, M. Ghazanfari, M. Jafari, "Evaluation model of business intelligence for enterprise systems using fuzzy TOPSIS," Expert Systems with Applications, **39**(3), 3764–3771, 2012, doi:10.1016/j.eswa.2011.09.074.

[6] M.M. Nazier, D.A. Khedr, A.P.M. Haggag, "Business Intelligence and its role to enhance Corporate Performance Management," International Journal of Management & Information Technology, **3**(3), 08–15, 2013, doi:10.24297/ijmit.v3i3.1745.
[7] C.M. Olszak, "Toward Better Understanding and Use of Business Intelligence in Organizations," Information Systems Management, **33**(2), 105–123, 2016, doi:10.1080/10580530.2016.1155946.
[8] L.T. Moss, S. Atre, Business intelligence roadmap: the complete project lifecycle for decision-support applications, Addison-Wesley Professional, 2003.
[9] W.H. Inmon, Building the Data Warehouse,3rd Edition, 2002.
[10] A.S. Girsang, S.M. Isa, A.L. Haris, Arwan, K. Mandagie, L.R. Ariana, V. Ardinda, "Business Intelligence for Product Defect Analysis," IOP Conference Series: Materials Science and Engineering, **598**(1), 2019, doi:10.1088/1757-899X/598/1/012117.
[11] R. Sharda, D. Delen, E. Turban, "Business Intelligence: A Managerial Perspective on Analytics," **4**(1), 386, 2013.
[12] M. Madhikerrni, K. Främling, "Data discovery method for Extract-Transform-Load," 2019 IEEE 10th International Conference on Mechanical and Intelligent Manufacturing Technologies, ICMIMT 2019, (Icmimt), 174–181, 2019, doi:10.1109/ICMIMT.2019.8712027.
[13] S. Hajian, J. Domingo-Ferrer, "A methodology for direct and indirect discrimination prevention in data mining," IEEE Transactions on Knowledge and Data Engineering, **25**(7), 1445–1459, 2013, doi:10.1109/TKDE.2012.72.
[14] J. Han, J. Pei, M. Kamber, Data Mining: Concepts and Techniques, Elsevier, 2011.
[15] M.P. Bach, A. Čeljo, J. Zoroja, "Technology Acceptance Model for Business Intelligence Systems: Preliminary Research," Procedia Computer Science, **100**, 995–1001, 2016, doi:10.1016/j.procs.2016.09.270.
[16] R. Sowmya, K.R. Suneetha, "Data Mining with Big Data," Proceedings of 2017 11th International Conference on Intelligent Systems and Control, ISCO 2017, **26**(1), 246–250, 2017, doi:10.1109/ISCO.2017.7855990.
[17] S. Umadevi, K.S.J. Marseline, "A survey on data mining classification algorithms," Proceedings of IEEE International Conference on Signal Processing and Communication, ICSPC 2017, **2018-Janua**(July), 264–268, 2018, doi:10.1109/CSPC.2017.8305851.
[18] N. Mishra, S. Silakari, "Predictive Analytics : A Survey , Trends , Applications ," International Journal of Computer Science and Information Technologies, **3**(3), 4434–4438, 2012.
[19] N. Jain, "Data Mining Techniques:a Survey Paper," International Journal of Research in Engineering and Technology, **02**(11), 116–119, 2013, doi:10.15623/ijret.2013.0211019.
[20] S. Saigal, D. Mehrotra, "Performance Comparison of Time Series Data Using Predictive Data Mining Techniques," Advances in Information Mining, **4**(1), 57–66, 2012.
[21] V. Kotu, B. Deshpande, Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner, Morgan Kaufmann, 2014.
[22] A.M. Shahiri, W. Husain, N.A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," Procedia Computer Science, **72**, 414–422, 2015, doi:10.1016/j.procs.2015.12.157.
[23] X. Pang, Y. Zhou, P. Wang, W. Lin, V. Chang, "An innovative neural network approach for stock market prediction," Journal of Supercomputing, **76**(3), 2098–2118, 2020, doi:10.1007/s11227-017-2228-y.
[24] T. ur Rehman, M.N.A. Khan, N. Riaz, "Analysis of Requirement Engineering Processes, Tools/Techniques and Methodologies," International Journal of Information Technology and Computer Science, **5**(3), 40–48, 2013, doi:10.5815/ijitcs.2013.03.05.
[25] S. Sen, "Integrating Related XML Data into Multiple Data Warehouse Schemas," Computer Science Conference Proceedings, **4**(April), 357–367, 2012, doi:10.5121/csit.2012.2133.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 130-137 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Efficiency Comparison in Prediction of Normalization with Data Mining Classification

Saichon Sinsomboonthong*

Department of Statistics, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand

ARTICLE INFO

Article history:
Received: 30 May, 2021
Accepted: 05 July, 2021
Online: 10 July, 2021

Keywords:
Artificial Neural Network
Binary Logistic Regression
Decimal Scaling Normalization
Decision Tree
K-Nearest Neighbor
Naïve Bayes
Statistical Column Normalization
Support Vector Machine
Z-Score Normalization

ABSTRACT

In research project, efficiency comparison study in prediction of normalization with data mining classification. The purpose of the research was to compare three normalization methods in term of classification accuracy that the normalized data provided: Z-Score, Decimal Scaling and Statistical Column. The six known classifications: K-Nearest Neighbor, Decision Tree, Artificial Neural Network, Support Vector Machine, Naïve Bayes, and Binary Logistic Regression were used to evaluate the normalization methods. The six studied data sets were into two groups. Those data sets were data sets of White wine quality, Pima Indians diabetes, and Vertebral column of which data were 1-5 variables of the outlier coefficient of variation and data sets of Indian liver disease, Working hours, and Avocado of which data were 6-10 variables of the outlier coefficient of variation.
The result of comparison White wine quality and Vertebral column, the best efficiency method had many methods in a non-systematic way. For the data set of Pima Indians diabetes and Indian liver disease, Statistical Column and classification by K-Nearest Neighbor was the best efficiency. For the data set of Working hours, Decimal Scaling and classification by K-Nearest Neighbor was the best efficiency. For the data set of Avocado, Statistical Column and classification by K-Nearest Neighbor, Z-Score and Decimal Scaling and classification by Binary Logistic Regression were the best efficiency. All of normalization and classification methods, Statistical Column and classification by K-Nearest Neighbor was the best efficiency by precision.

1. Introduction

Nowadays, advances in information technology have conveyed to the storing of large amount of data. However, most of data usage is still extracting data from database. The knowledge gained from this data analysis can be of great use in organizational operations and decision making. Data mining can be operated in many forms depending on the objective of data mining. Classification is a modeling for categorical data from pre-classified data to use that model to classify new data that has not previously been classified [1]. In addition, data mining is a method of extracting knowledge from different data to utilize that knowledge in decision making. Such knowledge may be used to predict or create models for classifying or displaying relationships between different units, which data mining can be applied in many organizations, for example: finance, insurance, medical, etc. Today there is a lot of interesting research or exploration. In the process of working on those researches, researchers often use statistical methods to analyze data and draw conclusions for those researches in further revision or development. In order to obtain data, the data collected can be disorganized, often causing problems. Each variable has different values ranging from little, medium, and very different. If those data were analyzed, the result would be differed from the truth. As a result, the assumptions were not met and the data could not be used in the best way. One way to manage this problem is transformation or normalization, using a simple mathematical method to adapt the collected data to a new and standardized form, for example: Z-Score, Median, Min-Max, Decimal Scaling and Statistical Column [2].

From the first literature review, in Malaysia, there is investigate the use of three normalizations in prediction of degue, for example: Min-Max, Z-Score and Decima Scaling. These methods in prediction model are consisted of Support Vector Machine (SVM) and Artificial Neural Network (ANN). The comparison results considered the accuracy of prediction and mean square error (MSE). The results show that SVM and ANN had the maximum accuracy and the minimum MSE for Decimal Scaling, Min-Max, and Z-Score respectively. Nevertheless, SVM

*Corresponding Author: Saichon Sinsomboonthong,
E-mail: saichon.ss49@gmail.com

is a better prediction as compared to the ANN [3]. The second, comparative analysis of K-Nearest Neighbor (KNN) with various k using Min-Max and Z-Score with R programming. The average accuracy was about 88% for Min-Max and 79% for Z-Score [4]. Finally, the efficiency of normalizations was compared. The main objective of this research was to compare four normalization methods in terms of classification accuracy that the normalized data provided. Those methods were the following: Min-Max, Z-Score, Decimal Scaling, and Median. Four data sets and three classifications by K-NN, Naïve Bayes, and ANN were used to evaluate the normalization methods. For the conclusion of the dataset of White wine quality, normalization by Decimal Scaling and classification by K-NN were the best combination. For the dataset of Pima Indians diabetes, normalization by Decimal Scaling and classification by ANN were the best combination. For the dataset of Vertebral column, normalization by Decimal Scaling and classification by K-NN were the best combination. For the dataset of Indian liver patient, normalization by Decimal Scaling and classification by Naïve Bayes were the best combination. We assume that the best normalization method was the Decimal Scaling and classification by K-NN [5].

In this research, three normalizations were studied; Z-Score, Decimal Scaling, and Statistical Column and were carried out with four classification methods which were regularly use; K-NN, Decision Tree, ANN and SVM. The other two proposed classification methods were Naïve Bayes and Binary Logistic Regression to compared the most accuracy efficiency in prediction of normalization with classification by R programming.

2. Experimental Methods

The experimental methods are systematic and scientific approach to research. Here, they consisted of data collection and research procedures [5].

2.1. Data Collection

Data collection is three step methods: gathering, measuring and analyzing the accuracy of the data for research by standard checked methods [5]. Six secondary data sets were collected from website UCI.com, Kaggle.com and Mldata.com as followed:

- White wine quality, total number of data 1,500 values with 1-5 variables of the outlier coefficient of variation [6].
- Pima Indians diabetes, total number of data 768 values with 1-5 variables of the outlier coefficient of variation [7].
- Vertebral column, total number of data 310 values with 1-5 variables of the outlier coefficient of variation [8].
- Indian liver patient, total number of data 575 values with 6-10 variables of the outlier coefficient of variation [9].
- Working hours, total number of data 956 values with 6-10 variables of the outlier coefficient of variation.
- Avocado, total number of data 1,149 values with 6-10 variables of the outlier coefficient of variation [11].

The data set consisted of 2 parts: data sets 1-3 contained 1-5 variables of the outlier coefficient of variation and data sets 4-6 contained 6-10 variables of the outlier coefficient of variation.

2.2. Research Procedures

Research procedures are the specific methodology or techniques used to identify, select, process, and analyze information [5]. Here, they consisted of normalization, data sets partitioning method, data analysis and efficiency comparison in prediction of classification.

2.2.1. Normalization

Z-Score using R program, Decimal Scaling and Statistical Column using Excel program were performed normalization.

2.2.2. Data Sets Partitioning Method

Dividing the data set into 2 sets and randomly 5 rounds by specifying the random seed as 10, 20, 30, 40 and 50 in the ratio of 70:30 which is commonly used in the data mining research. Part 1, training data set was applied to build a model by 70 percent. For part 2, testing data set was applied to test a model by 30 percent [12]-[16] as followed in table 1.

Table 1: Result of six data sets partition.

Data set	Total number of data set	Total number of training data set (70 percent)	Total number of testing data set (30 percent)
White wine Quality	1,500	1,050	450
Pima Indian diabetes	768	537	231
Vertebral column	310	217	93
Indian liver Patient	575	402	173
Working hours	956	669	287
Avocado	1,149	804	345

2.2.3. Data Analysis

Data analysis is the method of applying statistical data to describe, explain and appraise data [5].

2.2.3.1. Normalization

Normalization is the method of improving values using measured on the different scale to the same scale. It permits analogy of related values of different data. There are many normalizations, for example: Z-Score, Median, Min-Max, Decimal Scaling and Statistical Column. In this research, we interested in three normalizations as follows [4].

1) Z-Score Normalization

This method, the data (X) are subtracted from the mean $(\overline{X})$ and divided by the standard deviation (SD) of sample for every style on training data to transform each input style into the new data (X^*). The normalization formula is as follows [17];

$$X^* = \frac{X - \overline{X}}{SD} \quad (1)$$

2) Decimal Scaling Normalization

The decimal scaling normalization method transforms the original value of the data as a decimal number. The decimal position is defined by the maximum absolute value as follows [18].

$$X^* = \frac{X}{10^j} \quad (2)$$

where j is the number of positions of the largest value.

3) Statistical Column Normalization

The statistical column normalization method transforms every column with a normalized column value, $n(c_a)$. Compute the normalization of every column by subtracting the data (X) with a normalized column value to a length of one. Then, compute every column by dividing a normalized column value and multiplied by 0.1 which is biased as follows. [17]

$$X^* = \frac{X - n(c_a)}{n(c_a)} \times 0.1 \quad (3)$$

2.2.3.2. Classification

Classification is the method of specifying and managing individual values into a set. Then, it is applied to predict a model of testing data after training data as follows [4].

1) K-Nearest Neighbor

K-Nearest Neighbor (KNN) is a very popular method as it is a simple and effective method that can be used to many tasks such as classification and missing value replacement. It uses the IBk algorithm [19]. The first, the data set must be prepared and scaled into a normalized scale. Then, the Euclidean distance is computed between two points [4].

2) Decision Tree

The tree used in decision support is an upside-down tree structure with roots at the top and leaves at the bottom. Within the tree there are nodes, each of which represents a decision based on the attributes. The branches of the tree represent the values or results obtained from the test, and the leaves at the bottom of the decision tree represent class or results. The top node is called the root node. Here, the decision tree decided to use the J48 (C4.5) algorithm [20].

3) Artificial Neural Network

Artificial Neural Network (ANN) is technology developed from artificial intelligence research to calculation of function values from data groups. ANN is the method for machines to learn from a prototype and then train the system to think and solve broader problems. The structure of ANN consists of input and output node. Processing is distributed in a layered structure, namely input, output and hidden layer. ANN processing relies on the transmission of work through the nodes of these layers. Here, the ANN decided to use the Multilayer Perceptron algorithm [21], [22].

4) Support Vector Machine

The goal of this method is a supervised learning that a highly general classifier can be built. That is, it can be work well with unknown database with the data formatting process from the low dimensional data set on the input space is in the high dimensional data set on the feature space using a function to format the data, known as the kernel function. This capability makes it easier to construct a quadratic data classifier on a feature space for classification. In addition, a good classifier should have a linear structure and be able to create the distance area between the classifier and the closest value of each group to be effective in separating each type of data set from one another. The appropriate line is called the optimal separating hyperplane. Here, the support vector machine decided to use the Sequential Minimal Optimization (SMO) algorithm [18].

5) Naïve Bayes

The first proposed classification method was Naïve Bayes. It will use an analysis of the probability of things that have not happened before, based on the predictions of what has happened before. A simple form of relationships is as follows [23];

$$P(C \mid A) = \frac{P(A \mid C) \text{ x } P(C)}{P(A)} \quad (4)$$

From Bayes equation, if one is to predict the class C when attribute A is known, it can be calculated from the probability of attribute A with the class C in training data set and probability of attribute A and class C.

6) Binary Logistic Regression

The second proposed classification method was Binary logistic regression. It is a regression analysis in which the dependent variable is a qualitative variable with only two values while the independent variable can be either a quantitative or a qualitative variable, or may be both a quantitative and qualitative variable. The binary logistic regression analysis method has no distribution conditions for independent variables, and there is no conditions of the variance and covariance matrix for each group, and this method predicts probabilities that each unit is in a specific group [24];

$$\begin{aligned} P(Success) &= P(Y=1) \\ &= E(Y) = p \\ &= \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\ldots+\beta_p X_p)}} \end{aligned}$$

and

$$\begin{aligned} P(Failure) &= P(Y=0) \\ &= 1-p \\ &= \frac{1}{1+e^{\beta_0+\beta_1 X_1+\ldots+\beta_p X_p}} \end{aligned} \quad (5)$$

From the above equation, the relationship between the independent and the dependent variables is nonlinear. Therefore, the relationship is adjusted in a linear form as follows:

$$Odd\ Ratio = OR = \frac{P(Success)}{P(Failure)}$$

$$= \frac{P(Y=1)}{P(Y=0)}$$

$$= \frac{p}{1-p}$$

$$= e^{\beta_0+\beta_1X_1+...+\beta_pX_p} \quad (6)$$

If the odd ratio is greater than 1, then the probability of an event of success is greater than an event of failure.

An estimate of the odd ratio is

$$\widehat{OR} = \frac{\hat{p}}{1-\hat{p}}$$

$$= e^{b_0+b_1X_1+...+b_pX_p} \quad (7)$$

From the above equation, find $\log_e(OR)$

$$\log_e(OR) = \log_e\left(e^{\beta_0+\beta_1X_1+...+\beta_pX_p}\right)$$

$$= \ln\left(e^{\beta_0+\beta_1X_1+...+\beta_pX_p}\right)$$

$$= \ln(OR)$$

$$= \beta_0+\beta_1X_1+...+\beta_pX_p \quad (8)$$

The right hand side of the above equation is in a linear form, called the logit response function.

If sample data is used,

$$\log_e\left(\widehat{OR}\right) = \ln\left(\widehat{OR}\right)$$

$$= b_0+b_1X_1+...+b_pX_p \quad (9)$$

2.2.4. Efficiency Comparison in Prediction of Classifications

The analysis results of three normalization methods were used by six classifications to compare the efficiency in prediction from the accuracy as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \text{ x } 100\% \quad (10)$$

where *True Positive* (*TP*) is the number of exactly classified as positive, the real value is positive. *True Negative (TN)* is the number of exactly classified as negative, the real value is negative. *False Positive (FP)* is the number of mistakenly classified as positive, the real value is negative and *False Negative (FN)* is the number of mistakenly classified as negative, the real value is positive [25].

Flowchart showed the step of experimental methods as follows in figure 1. The process started from six secondary data sets were collected from website. Therefore, normalization is the method of improving values using measured on the different scale to the same scale. There are three normalizations, for example: Z-Score, Decimal Scaling and Statistical Column. After that, data set were divided into 2 sets and randomly 5 rounds by specifying the random seed as 10, 20, 30, 40 and 50 in the ratio of 70:30. Part 1 the training data was applied to built a model using 70 percent. For part 2 the testing data was applied to test a model using 30 percent. Then, classification was applied to predict a model of testing data after training data. Classification consisted of six methods, for example: K-Nearest Neighbor, Decision Tree, Artificial Neural Network, Support Vector Machine, Naïve Bayes and Binary Logistic Regression. Finally, the analysis results of three normalization methods were used by six classifications to compare the efficiency in prediction from the accuracy.

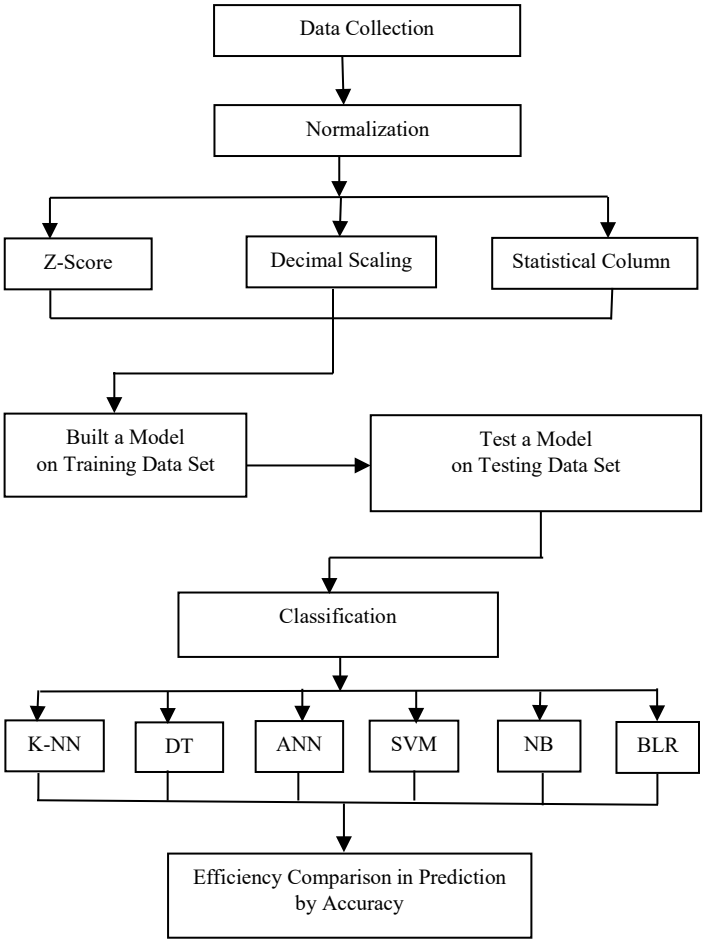


Figure 1: Flowchart of Experimental Methods

K-NN = K-Nearest Neighbor
DT = Decision Tree
ANN = Artificial Neural Network
SVM = Support Vector Machine
NB = Naïve Bayes
BLR = Binary Logistic Regression

3. Results and Discussions

3.1. White Wine Quality Data Set

As shown in Table 2, if Z-Score is used, classification by Decision Tree, Artificial Neural Network, Support Vector Machine and Binary Logistic Regression had the maximum accuracy at 100 percent. But if Decimal Scaling is used,

classification by K-Nearest Neighbor, Decision Tree, Artificial Neural Network, Support Vector Machine and Binary Logistic Regression had the maximum accuracy at 100 percent. If Statistical Column is used, classification by K-Nearest Neighbor, Support Vector Machine and Binary Logistic Regression Binary Logistic Regression had the maximum accuracy at 100 percent.

Table 2: The results of efficiency comparison in white wine quality data using Z-Score, Decimal Scaling and Statistical Column with Classification for K-Nearest Neighbor, Decision Tree, Artificial Neural Network, Support Vector Machine, Naïve Bayes and Binary Logistic Regression.

Classification	Normalization		
	Z-Score	Decimal Scaling	Statistical Column
K-Nearest Neighbor	93.4222	**100**	**100**
Decision Tree	**100**	**100**	92
Artificial Neural Network	**100**	**100**	56.1231
Support Vector Machine	**100**	**100**	**100**
Naïve Bayes	98.8446	99.0235	69.8728
Binary Logistic Regression	**100**	**100**	**100**

3.2. Pima Indians Diabetes Data Set

As shown in Table 3, if Z-Score is used, classification by Binary Logistic Regression had the maximum accuracy at 77.7320 percent. But if Decimal Scaling is used, classification by Decision Tree had the maximum accuracy at 79.2208 percent. If Statistical Column is used, classification by K-Nearest Neighbor had the maximum accuracy at 81.7316 percent. All the normalization and classification are compared, the Statistical Column Normalization and K-Nearest Neighbor classification had the maximum accuracy.

Table 3: The results of efficiency comparison in Pima Indians diabetes data using Z-Score, Decimal Scaling and Statistical Column with Classification for K-Nearest Neighbor, Decision Tree, Artificial Neural Network, Support Vector Machine, Naïve Bayes and Binary Logistic Regression.

Classification	Normalization		
	Z-Score	Decimal Scaling	Statistical Column
K-Nearest Neighbor	68.4848	69.4373	**81.7316**
Decision Tree	74.4589	**79.2208**	69.6969
Artificial Neural Network	77.0043	77.2824	65.3877
Support Vector Machine	76.9500	76.9500	67.8400
Naïve Bayes	73.6111	64.4787	66.3375
Binary Logistic Regression	**77.7320**	72.2247	69.0573

3.3. Vertebral Column Data Set

As shown in Table 4, if Z-Score is used, classification by Binary Logistic Regression had the maximum accuracy at 86.5807 percent. But if Decimal Scaling and Statistical Column are used, classification by K-Nearest Neighbor and Decision Tree had the maximum accuracy at 100 percent. All the normalization and classification are compared, Decimal Scaling, Statistical Column Normalization and K-Nearest Neighbor classification or Decimal Scaling, Statistical Column Normalization and Decision Tree classification had the maximum accuracy.

Table 4: The results of efficiency comparison in Vertebral column data using Z-Score, Decimal Scaling and Statistical Column with Classification for K-Nearest Neighbor, Decision Tree, Artificial Neural Network, Support Machine, Naïve Bayes and Binary Logistic Regression.

Classification	Normalization		
	Z-Score	Decimal Scaling	Statistical Column
K-Nearest Neighbor	81.9355	**100**	**100**
Decision Tree	67.7419	**100**	**100**
Artificial Neural Network	83.4513	83.9367	80.5782
Support Vector Machine	76.9500	76.9500	67.8400
Naïve Bayes	74.5348	76.6055	81.9705
Binary Logistic Regression	**86.5807**	73.9175	83.6670

3.4. Indian Liver Disease Data Set

As shown in Table 5, if Z-Score and Decimal Scaling are used, classification by Binary Logistic Regression had the maximum accuracy at 73.1029 and 73.1054 percent respectively. But if Statistical Column is used, classification by K-Nearest Neighbor had the maximum accuracy at 99.6531 percent. All the normalization and classification are compared, Statistical Column Normalization and K-Nearest Neighbor classification had the maximum accuracy.

Table 5: The results of efficiency comparison in Indian liver disease data using Z-Score, Decimal Scaling and Statistical Column with Classification for K-Nearest Neighbor, Decision Tree, Artificial Neural Network, Support Vector Machine, Naïve Bayes and Binary Logistic Regression

Classification	Normalization		
	Z-Score	Decimal Scaling	Statistical Column

K-Nearest Neighbor	64.1619	61.3341	**99.6531**
Decision Tree	68.2080	67.0520	69.3641
Artificial Neural Network	65.7435	70.2721	73.7657
Support Vector Machine	70.9700	70.9700	70.9700
Naïve Bayes	63.2700	71.1385	60.9860
Binary Logistic Regression	**73.1029**	**73.1054**	72.9447

3.5. Working Hours Data Set

As shown in Table 6, if Z-Score and Statistical Column are used, classification by Naïve Bayes had the maximum accuracy at 79.5518 and 99.7138 percent respectively. But if Decimal Scaling is used, classification by K-Nearest Neighbor had the maximum accuracy at 100 percent. All the normalization and classification are compared, Decimal Scaling Normalization and K-Nearest Neighbor classification had the maximum accuracy.

Table 6: The results of efficiency comparison in Working hours data using Z-Score, Decimal Scaling and Statistical Column with Classification for K- Nearest Neighbor, Decision Tree, Artificial Neural Network, Support Vector Machine, Naïve Bayes and Binary Logistic Regression.

Classification	Normalization		
	Z-Score	Decimal Scaling	Statistical Column
K-Nearest Neighbor	71.6376	**100**	99.5818
Decision Tree	73.5191	73.5191	72.8223
Artificial Neural Network	78.1346	78.0755	54.1547
Support Vector Machine	74.6300	74.4800	65.5200
Naïve Bayes	**79.5518**	78.8154	**99.7138**
Binary Logistic Regression	74.6093	74.9622	73.5478

3.6. Avocado Data Set

As shown in Table 7, if Z-Score and Decimal Scaling are used, classification by Binary Logistic Regression had the maximum accuracy at the same 100 percent. If Statistical Column is used, classification by K-Nearest Neighbor had the maximum accuracy at 100 percent. All the normalization and classification are compared, Statistical Column Normalization and K-Nearest Neighbor classification or Z-Score, Decimal Scaling Normalization and Binary Logistic Regression classification had the maximum accuracy.

Table 7: The results of efficiency comparison in Avocado data using Z-Score, Decimal Scaling and Statistical Column with Classification for K- Nearest Neighbor, Decision Tree, Artificial Neural Network, Support Vector Machine, Naïve Bayes and Binary Logistic Regression.

Classification	Normalization		
	Z-Score	Decimal Scaling	Statistical Column
K-Nearest Neighbor	99.7101	85.3333	**100**
Decision Tree	66.3768	66.3768	66.3768
Artificial Neural Network	99.2691	99.3303	99.4265
Support Vector Machine	96.0300	96.0300	99.6300
Naïve Bayes	90.0744	89.4525	99.5136
Binary Logistic Regression	**100**	**100**	99.9748

The result of study of efficiency comparison in prediction of normalization with data mining classification for data set with 1-5 variables of the outlier coefficient of variation were White wine quality, Pima Indians diabetes and Vertebral column. The best efficiency method was Statistical Column Normalization and classification by K-Nearest Neighbor, and Decimal Scaling Normalization and classification by Decision Tree and K-Nearest Neighbor respectively. For the dataset with 6-10 variables of the outlier coefficient of variation were Indian liver disease, Working hours and Avocado. The best efficiency method is Statistical Column Normalization and classification by K-Nearest Neighbor and Decimal Scaling Normalization and classification by K-Nearest Neighbor respectively which was similar to the research of T. Malai et al. (2021) found that the best method was Decimal Scaling Normalization and classification by K-Nearest Neighbor.

Table 8: The results of efficiency comparison all data using Z-Score, Decimal Scaling and Statistical Column with Classification for K-Nearest Neighbor, Decision Tree, Artificial Neural Network, Support Vector Machine, Naïve Bayes and Binary Logistic Regression.

Data	Classification	Normalization		
		Z-Score	Decimal Scaling	Statistical Column
White Wine Quality	- K-Nearest Neighbor		✓	✓
	- Decision Tree	✓	✓	
	- Artificial Neural Network	✓	✓	
	- Support Vector Machine	✓	✓	✓
	- Naïve Bayes			
	- Binary Logistic Regression	✓	✓	✓
Pima Indians Diabetes	- K-Nearest Neighbor			✓
	- Decision Tree			
	- Artificial Neural Network			
	- Support Vector Machine			
	- Naïve Bayes			
	- Binary Logistic Regression			
Vertebral Column	- K-Nearest Neighbor		✓	✓
	- Decision Tree		✓	✓

	- Artificial Neural Network - Support Vector Machine - Naïve Bayes - Binary Logistic Regression			
Indian Liver Disease	- K-Nearest Neighbor - Decision Tree - Artificial Neural Network - Support Vector Machine - Naïve Bayes - Binary Logistic Regression			✓
Working Hours	- K-Nearest Neighbor - Decision Tree - Artificial Neural Network - Support Vector Machine - Naïve Bayes - Binary Logistic Regression		✓	
Avocado	- K-Nearest Neighbor - Decision Tree - Artificial Neural Network - Support Vector Machine - Naïve Bayes - Binary Logistic Regression	✓	✓	✓

✓ = the best accuracy for each data set

As shown in Table 8, White wine quality data set, the highest efficiency methods were Decimal Scaling, Statistical Column and classification by K-Nearest Neighbor; Z-Score, Decimal Scaling and classification by Decision Tree and Artificial Neural Network and Z-Score, Decimal Scaling, Statistical Column and classification by Support Vector Machine and Binary Logistic Regression. Pima Indians diabetes data set, the maximum efficiency method was Statistical Column and classification by K-Nearest Neighbor. Vertebral column data set, the maximum efficiency method was Decimal Scaling, Statistical Column and classification by K-Nearest Neighbor and Decision Tree. Indian liver disease data set, the maximum efficiency method was Statistical Column and classification by K-Nearest Neighbor. Working hours data set, the maximum efficiency method was Decimal Scaling and classification by K-Nearest Neighbor. Avocado data set, the maximum efficiency method was Statistical Column and classification by K-Nearest Neighbor. The another maximum efficiency methods were Z-Score, Decimal Scaling and classification by Binary Logistic Regression.

4. Conclusion

In summary, White wine quality data and Vertebral column data, the maximum efficiency method have many methods in a non-systematic way. Pima Indians diabetes data and Indian liver data, the maximum efficiency method was Statistical Column and classification by K-Nearest Neighbor. Vertebral column data, the maximum efficiency method was Decimal Scaling, Statistical Column and classification by K-Nearest Neighbor and Decision Tree. Working hours data, the maximum efficiency method was Decimal Scaling and classification by K-Nearest Neighbor. Avocado data, the maximum efficiency method was Statistical Column and classification by K-Nearest Neighbor and the another maximum efficiency methods were Z-Score, Decimal Scaling and classification by Binary Logistic Regression. All of normalization and classification methods, Statistical Column and classification by K-Nearest Neighbor was the best efficiency by precision. This finding of Statistical Column and classification by K-Nearest Neighbor can be applied in many fields of medical, public health and science in real world problem.

Conflict of Interest

The author announce no conflict of interest.

Acknowledgment

I thank the School of Science, King Mongkut's Institute of Technology Ladkrabang for funding research project, help and support on efficiency comparison in prediction of normalization with data mining classification.

References

[1] S. Euawattanamongkol, Data mining, National Institute of Development Administration Publisher, 2016.
[2] N. Kratethong, Transformation to normal distribution, Master's Degree Thesis in Statistics, Department of Statistics, Faculty of Commerce and Account, Chulalongkorn University, 1999.
[3] Z. Mustaffa, Y.A. Yusof, "Comparison of normalization techniques in predicting dengue outbreak," in 2010 International Conference on Business and Economics Research, **1**, 345-349, IACSIT Press, Kuala Lumpur, Malaysia, 2011.
[4] A. Pandey, A. Jain, "Comparative analysis of KNN algorithm using various normalization techniques," International Journal Computer Network and Information Security, **11**, 36-42, 2017, doi:10.5815/ijcnis.2017.11.04.
[5] T. Malai, P. Ninthanom, S. Sinsomboonthong, "Performance comparison of transformation methods in data mining classification technique," Thai Journal of Science and Technology, **10**(1), 510-522, 2021. DOI: 10.1109/2018.2841987
[6] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, J. Reis, Wine quality data set, [Online], Available : https//:archive.ics.uci.edu/ml/datasets/Wine+Quality, 2009.
[7] J.W. Smith, J.E. Everhart, W.C. Dickson, W.C. Knowler, R.S. Johannes, Pima Indians diabetes database, [Online], Available : https//:www.kaggle.com/uciml/pima-indians-diabetes-database, 1988.
[8] H.D. Mota, Vertebral column data set, [Online], Available : https://www.kaggle.com/caesarlupum/vertebralcolumndataset, 2011.
[9] B.V. Ramana, Indian liver patient, [Online], Available : https//:www.mldata.io/dataset-details/indian_liver_patient/, 2012.
[10] L. Myoung, Working hours, [Online], Available : https://rdrr.io/rforge/Ecdat/man/Workinghours.html, 1995.
[11] J. Kiggins, Avocado prices, [Online], Available : https://www.kaggle.com/neuromusic/avocado-prices, 2018.
[12] R. Shams, Creating training, validation and test sets (data preprocessing), [Online], Available : https://www.youtube.com/watch?v=uiDFa7iY9yo, 2014.
[13] P. Thongpool, P. Jamrueng, R. Boonrit, S. Sinsomboonthong, "Performance comparison in prediction of imbalanced data in data mining classification," Thai Journal of Science and Technology, **8**(6), 565-584, 2019. DOI: 10.1109/TJST.2019.2841987
[14] S. Sinsomboonthong, "An efficiency comparison in prediction of imbalanced data classification with data mining techniques," Thai Journal of Science and Technology, **8**(3), 383-393, 2019.
[15] N. Phonchan, P. Jaimeetham, S. Sinsomboonthong, "Clustering efficiency comparison of outliers data in data mining," Thai Journal of Science and Technology, **9**(5), 589-602, 2020.
[16] S. Sinsomboonthong, "An efficiency comparison in prediction of outlier six classifications," Thai Journal of Science and Technology, **9**(3), 255-268, 2020.
[17] T. Jayalakshmi, A. Santhakumaran, "Statistical normalization and back propagation for classification," International Journal of Computer Theory and Engineering, **3**(1), 89-93, 2011.
[18] J. Han, M. Kamber, Data mining concepts and techniques, 2nd ed, Morgan Kaufmann, 2006.

[19] O.G. Troyanskaya, M. Cantor, G. Sherlock, O. Patrick, P.O. Brown, "Missing value estimation methods for DNA microarrays," Bioinformatics, **17**(6), 520-525, 2011.
[20] R. Thammasombat, Decision support system for mobile internet package selection using decision tree, Ph. D Thesis, Business Computer, Faculty of Business Administration, Ratchapruek College, 2012.
[21] K. Waiyamai, C. Songsiri, T. Rakthammanon, "Using data mining techniques to improve the quality of education for students of the faculty of engineering," The NECTEC Technical Journal, **11**(3), 134-142, 2011. DOI: 10.1109/2011.7508132
[22] D.T. Larose, Discovering knowledge in data : an introduction to data mining, John Wiley & Sons, 2005.
[23] D.T. Larose, Data mining methods and models, John Wiley & Sons, 2005.
[24] K. Wanichbancha, Multivariate data analysis, Thammasarn Co Ltd, 2009.
[25] S. Sripaaraya, S. Sinsomboonthong, "Efficiency comparison of classifications for chronic kidney disease : a case study hospital in India," Journal of Science and Technology, **25**(5), 839-853, 2017. DOI: 10.1109/CONFLUENCE.2016.7508132

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol.6, No.4, 138-146 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

The Gamification Design for Affordances Pedagogy

Wilawan Inchamnan[1,*], Jiraporn Chomsuan[2]

[1]*College of Creative Design & Entertainment Technology, Dhurakij Pundit University, Bangkok, Thailand*

[2]*College of Innovation Business and Accountancy, Dhurakij Pundit University, Bangkok, Thailand*

ARTICLE INFO

Article history:
Received: 12 May, 2021
Accepted: 22 June, 2021
Online: 10 July, 2021

Keywords:
Gamification
Affordance
Growth Mindset
Design method
Affordances Pedagogy
Motivation and Engagement

ABSTRACT

This study aims to design a gamification affordances pedagogy. Affordances are the ways in which we perceive environments to support the needs of learners in the educational system. The main questions are how gamification elements can influence student engagement to improve their affordances. Affordance behavior is a human behavior that refers to a mindset; an attitude or opinion, especially a habitual one. Motivational activities can change a learner's behavior. A skill-based mindset can be created through the use of affordance motivation. Affordance refers to the points, badges, and leaderboards in gamification elements. This research aims to improve the affordance mindset design of interactive systems with gamification. The affordance design will improve the pedagogy related to engagement. The research focuses on the mindset factors and the relationship between the factors that promote the desired learning outcomes. The findings may help in designing the gamification affordance design method for affordance pedagogy. The expected model could improve learners' affordances and instructional activities.

1. Introduction

In the 21st century, there is extensive research on growth mindset and intrinsic motivation in learning. The constructs of mindset and motivation are important for educators to determine the impact on student learning and outcomes [1]. Understanding the two constructs, mindset and motivation, and the relationship between them is necessary as it provides insight into student motivation and drive. Gamification is a tool that can increase and promote user motivation, especially in education. The educational concept requires that teaching and learning activities are more fun and interesting [2]. Today, learner engagement is still a challenge in the education system. Design is employed in education to increase the student desire to focus on the educational task [3], as an affordance mindset. Educational games and various forms of edutainment have gained more attention in the discipline of learning and teaching strategies. Educators believe learning can be enhanced through play and fun [4]. The increasing motivation to learn may affect the learner's affordances. An affordance concerns the possible actions that an item offers while learning. The term affordance is a somewhat ambiguous term [5]. and affordance could be improved in terms of its ability to influence learning outcomes.

Learner can conduct their learning lives using advanced technology that engender an affordance behavior mindset. The mindset is a crucial factor in leaner motivation. Affordances are a core opportunity for action [6]. A mindset can change the attitude to learning in the education system, which represents a step in the right direction.

This paper reviews several recent gamification studies that focus on growth mindset and motivation. The theoretical frameworks of Affordance Mindset and Motivation reflect how they are applied in educational gamification. The research design is divided into two parts. The first focuses on the engagement elements related to the gamified classroom activities. The gamification strategies are then designed by using the mindset factors and learner characteristics. Then the gamification affordances design method is applied through points, badges,

*Corresponding Author: Wilawan Inchamnan, E-mail: wilawan.inn@dpu.ac.th

leaderboards and ranks in gamification activities for the pedagogy strategies.

2. Literature Review

2.1. Mindset

The mindset is a set of both conscious and unconscious human beliefs, which relates to how humans view what they consider to be their personality. Mindset can be divided into two types, Fixed Mindset and Growth Mindset. These mindsets refer to the way people think about the nature of intelligence and learning. People with a growth mindset value effort, tend to set learning goals (e.g., mastery) rather than performance goals (e.g., grades), and attribute failure to lack of effort rather than lack of ability [7]. Learnersì mindsets can be influenced by school-based activities to help improve academic outcomes [7] through motivation and engagement.

2.2. Motivation and Engagement

Motivation and engagement indicate passion and emotional involvement in learning activities [8]. Engagement permits meaningful learning, which includes the quality of student effort, student interaction and their immersive experiences [8]. Some research divided engagement into three dimensions: behavioral, emotional, and cognitive engagement [3].

2.3. Motivation

Motivation is an abstract construct used to explain people behavior. The behavior represents the basis for peopleìs actions, desires, and needs. Motivation can be named as oneìs behavioral direction, or what justification a person to want to repeat a behavior [9]. Motivation can be allocated into two different types known as intrinsic (internal) motivation and extrinsic (external) motivation [10]. Intrinsic motivation is the desire to seek new things and new challenges, to analyze one's abilities, to observe and to gain knowledge [11]. It is driven by interest or enjoyment in the task itself and exists within the individual rather than relying on external factors or the desire for reward. Intrinsic motivation arises from within the individual, just as the idea of an affordance draws attention to a possible action [12]. Extrinsic motivation mentions to the performance of an activity to attain a desired consequence and is the opposite of intrinsic motivation [11]. Extrinsic motivation is the type of motivation that comes from outside the individual and often involves rewards such as trophies, money, social recognition, or praise. In education, motivation is a major cause of differences in student learning outcomes, considered a possible predictor of a student's academic performance; students with high academic motivation are more likely to succeed academically [13]. Students who are intrinsically motivated are more likely to be curious and inquire about the process, focusing on the task itself rather than just the outcome. In contrast, students who are extrinsically motivated are more concerned with the outcome (e.g., grades, prizes) than with the process of completing the task itself [14]. Game activities could encourage player experiences, which is called immersive engagement. Engagement also influences peopleìs adoption. Behavior usage is how frequently or for what purpose the behavior is used while behavior adoption is the degree to which the behavior is utilized. Psychological outcome is a measure of effort. The level of effort is influenced by emotional engagement (pleasure, excitement, and persistence), behavioral engagement (effort), cognitive engagement (attention, reflection), and learning performance (perceived competence, perceived improvement) [3]. The student can repeat desired behaviors by reinforcement providing. Reinforcement incentives people into two forms of motivation. The first one is intrinsic motivation that refers to engage in behaviors for enjoyment, challenge, pleasure, or interest [5]. Then, the extrinsic motivation can engage in an activity to earn an external reward when learner is motivated to perform their behavior [1].

2.4. Engagement

Engagement in the extent to which a learner connects with the gaming environments and indicates a positive psychological state of mind when so doing [15]. Games' activities provide immediate feedback, which is more effective and efficient than traditional learning strategies. Gamification strategies can encourage the learner to be acquire experience during play. Game experiences drive personal change and transformation by generating an attitude of acceptance about the challenge, motivation to achieve, and constant innovation by simulation. Simulation encourages the participant to immerse themselves in learning [16]. Adaptation is the process by which strategies are moderated by engagement. Adaptation is a consequence of activities and events that are enhanced, developed, and implemented. Success in learning depends on the learner's desire to learn, which is known as behavioral intention [17]. Learners can change their behavior as a result of motivation. Engagement may encourage behavior. Usage can be influenced by behavioral assumption. The focus of usage behavior is on specific activities performed using specific sources of information. These include general knowledge acquisition, learning, and the pursuit of purposes [18].

2.5. Gamification Affordances

As a concept, affordance provides a useful bridge to explain the interplay between the artefact and the human user [5]. Gamification elements could provide a thematic evaluation of subsequent formulations of the term affordance. This is the method designed to provide a preliminary overview of how the notion of affordance can be interpreted. The figure 1 show the direct perception of affordances for the user along with a consideration of the usersì skills. The researcher refers to the "flow channel" as a linear function on a plane with skills and challenges as axes. An increase in the learner's skills is due to learning, and an increase in the challenges of performing a task is due to novelty [19]. The pedagogy design may keep the two in balance between challenges and skills. Subjects experienced Flow when they first encountered a task with a high balance between skills and challenges [20]. Skills is due to learning and an increase in the challenges of performing tasks [21], which relate to an affordance. The design of the

pedagogy allows for a practical mapping flow onto the gameplay. The approach enhances the player's interaction with the game elements and provides useful insights into the learner's skills [21]. The next section presents gamification affordance strategies including points, badges, and leaderboards.

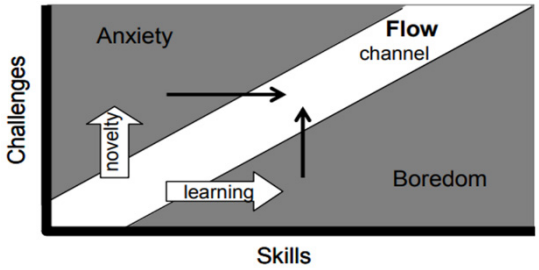


Figure 1:The Flow Channel

2.5.1. Points

Point is the activity outcome. Game point can trigger and reinforce situational awareness competencies. Points can provide information about the playerìs progressive such as response times, correct answers, and the procedures followed during play. The awarding of points is connected to behaviors that require crucial competencies [22].The competencies can develop as a training tool for learning goals [23].

This paper aims to design affordance mindset processes through the player data.The affordance for learning is measured by questionnaire questions regarding learnersì sense of pride and community acceptance. The satisfaction of basic psychological needs (competence, autonomy, and relatedness [24]), is a fundamental requirement for being autonomously motivated. Competence refers to the experience of success by fulfilling challenging tasks and gaining mastery within an environment [25] and is reflected by the number of points scored.

2.5.2. Badges

Badges reflect performance results. Digital badges indicate the achievements or skills acquired while playing the game. These badges are collected and displayed to the other players [25]. Gamification elements are used to encourage performance and skill acquisition, which are the desired learning outcomes.

2.5.3. Leaderboards

The most-used game element is leaderboard that refers to a ranking board of the players in a competitive event. The leaderboard is to illustrate player where they are ranked in a gamified system. Leaderboardsì mechanic can be employed in various ways to offer goals and to increase motivation [26]. Leaderboard ranking can motivate players to compete, which increases participation [27] and facilitates comparison and competition.Leaderboard is the basic elements that make up games that combined to deliver a system of mastery to end users [28].

2.6. Cognition

Cognition refers to the mental processes involved in the acquisition of knowledge and understanding. These cognitive processes include thinking, knowing, remembering, judging, and problem solving [29]. Cognitive processes affect every aspect of life, from school to work, to relationships. Some specific uses for these cognitive processes include the following.

- Learning New Things that require being able to take in new information and form new memories. The learner makes connections with other things that they already know.

- The formation of memories is a major topic in the field of cognitive psychology. Memories refer to how people remember, what they remember, and what they forget, and reveal much about how cognitive processes work.

Making decisions means making judgments about things you have experienced and processed. This may involve comparing new information with previous knowledge or integrating new information with new knowledge before making a decision [30]. Behavioral action refers to use, i.e. the fact that it is used, or habit. Relationships have been found between adoption, post-adoption variables, and usage behavior in the post-adoption process. Continued use based on experience and satisfaction in the post-adoption process represents high quality use [31]. Post-adoption, each individual is engaged in change behaviors to varying degrees. In this context, several factors may influence the relationships between adoption and post-adoption variables [31].

2.6.1. Positive Feedback

Positive feedback is a game mechanism. This mechanism is designed to accelerate or enhance ongoing output [32]. Gamification can be applied to stimuli of increasing intensities through intrinsic rewards and feedback. The user-centered design activities require an interactive feedback. The learning activity feedback is determined by peopleìs motivation and cognitive mindset.

2.6.2. Growth Mindset

The term mindset refers to implicit beliefs that have been shown to influence the thoughts and actions of individuals [33]. A learner's mindset has been shown to influence his motivation and academic performance [34], [35]. Growth mindset type means that a learner can improve his talents and abilities through effort. Growth mindset belief is a type of intelligence that can be improved through hard work and the use of strategies [36]. Growth Mindset can be generated through motivation and achievement [37], [38], could promote learner's engagement.

A growth mindset helps learners improve their skills and knowledge over time, and mindset research studies the power of such beliefs in influencing human behavior [39]. Mindset is a soft skill of great importance [40]. For example, athletes are driven by success and can realize their potential through effort, practice, and instruction. In the education system, some research has shown that students with a growth mindset can greatly improve their success and achievement [41].

2.7. The Growth Mindset Stimulus Model

According to the previous research, the figure 2 illustrates the gamification model, which encourages learner to conduct their learning by using gamification elements. The learning behavior refers to the taking steps in the right direction guideline. Gamification activities conceptual model can be influenced by gamified activities [42]. The growth mindset is fostered by the positive feedback through cognition and motivation. The motivation can be stimuli by competency and leaderboard. The cognition could be influence by level [42].

The figure 2 shows the gamification workflow for growth mindset processes that influence the positive feedback and growth mindset by gamified activities. The learnerìs growth mindset is measured through feedback, level, motivation, their position on the leaderboard, and their competencies. The model shows the gamification activities can increase the likelihood of achieving the learning objectives by motivating students to learn. This illustrates the relationship between the stimulus activities and growth mindset processes [43]. Game activities can encourage the learner to practice their skills and their position on the leaderboard can enhance their pride and social acceptance. Motivation and cognition can enhance their performance through positive feedback and higher scores (GPA).

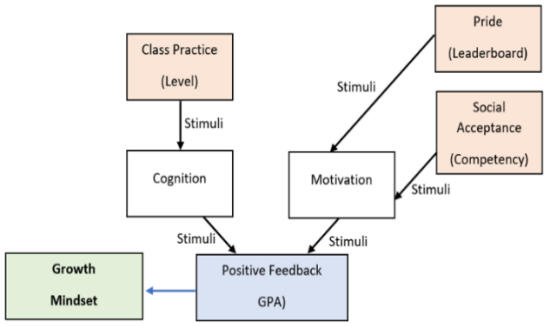


Figure 2: The Growth Mindset Stimulus Model [42]

3. Methodology

Having reviewed a scholarly source on a topic concerning mindset then mindset questionnaire is developed to classify participants into a different mindset characteristic. This research aims to link the mindset factors and gamification elements. The experiment design aims to cluster the mindset criteria and determine the relationships between the factors and the learning outcomes. The data collection uses a mindset questionnaire to facilitate comparisons with learner subjects. The pedagogy design gathers the university lecturer for the strategiesì class activity. The advisor experiences help to map the teaching or pedagogy style with gamification elements. Figure 3 illustrates the processes of data evaluation.

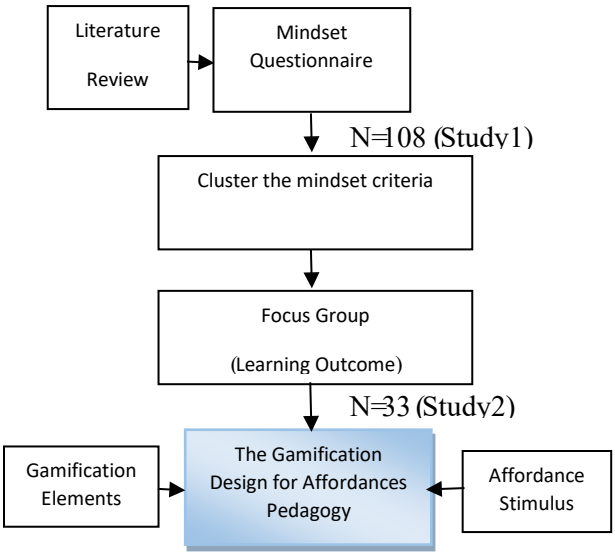


Figure 3: The Methodology of the gamification design for affordances pedagogy

3.1. Participants

Study1: According to factors finding, the participants in this study were 108 university Dhurakij Pundit University students. Participants included 72%from the Business faculty, 24%from the Creative Design faculty and 4%from the Journalism faculty all aged between 18 and 23. All participants in this study were volunteers. The average GPA was 3.55.

Study2: Then, the pedagogy design participants in this study were 33 university lecturers, randomly. The questionnaire design used the factors from the first findings. The average of experiences was 13 years. Normally, their teaching style were lecture and practical (50:50). They preferred to use gamification in the classroom and showed the score or progress to their student.

3.2. Clustering the mindset criteria: Factor Analysis (Study1)

This study uses the factor analysis. The mindset measures by questionnaires survey. The students may respond to questions about their opinions, which are all associated with the latent fix and growth variable mindset as below.

- Fix1: It is difficult for my intelligence /level of intelligence to change.
- Fix2: There are certain activities /subjects that I cannot be good at.
- Fix3: I think people do not have to try to do what they are capable of.
- Fix4: I can always learn new things, but I do not think that learning can increase my intelligence.
- Fix5: I am frustrated to see that people can do better than me.
- Fix6: If I have try something new and fail, I am not good at it.

- G1: I believe I can develop in all areas, albeit a little.
- G2: When I do something wrong, I feel that I would learn more from it.
- G3: I feel great when people see that I am good /talented, that means I am successful.
- G4: I am inspired by successful people.
- G5: I feel that I can help others to be successful.
- G6: If I try something new and fail, I want to repeat it until I can.

3.3. The Pedagogy Design (Study2)

The design process aims to link the pedagogical design and gamification elements and this paper focuses only on the factors affecting the mindset. The conceptual model can be employed in future work. The model can apply to the pedagogical context design includes the pedagogy plan.

The second findings from the university lectures show the relationship between factors.

- G1: The learnerìs skills can develop
- G2: Learnerìs practicing in the classroom
- G3: The level of learnerìs activities during learning
- G4-1: lectureriìs Comment for lesson activities individual
- G4-2: lectureriìs Comment for lesson activities to all of students
- Gamification: The class activities during learning
- Learn for smart:- Learn new things can increase the intelligence.

4. Results

4.1. Clustering the mindset criteria

The relationship of each variable to the underlying factor is expressed by the factor loading. The result of factor analysis, which deals with indicators of mindset, with twelve variables (opinions) and four resulting factors is shown below. Factor loadings can be interpreted like standardized regression coefficients, so the opinion in response 1 (Fix1) has a correlation of 0.72 with Factor 1. Five others, responses 2,3,4,5, and 6 (Fix 2,3,4,5,6), all representing fixed-attitude characteristics, are also associated with Factor 1. Based on the loading of the variables, the finding points high on factor 1, it could be considered as "Fixed Mindset".

However, opinions on questions 9 and 10 (Growth3 and 4), have high factor loadings on the other factor, Factor 2. They seem to indicate an external stimulus; that is, students are driven by other people, so Factor 2 would be classified as "Exogenous growth mindset." Similarly, Factor 4, is constituted by an answer of Growth 5 and 6 showing that students have positive feelings whenever they help others, which is an external stimulus, hence Factor 4 would also be regarded as "Exogenous growth mindset".

Table 1: Component Analysis

Rotated Component Matrix[a]

	Component			
	1	2	3	4
Fix1	.721	-.353	.071	.170
Fix6	.708	.174	-.121	-.193
Fix5	.690	.235	-.206	-.221
Fix2	.675	-.091	-.105	-.077
Fix3	.673	-.246	.091	.106
Fix4	.611	.131	.060	-.243
G3	.072	.784	.178	.011
G4	-.077	.638	.328	.221
G2	-.231	.164	.827	.081
G1	.106	.185	.796	.070
G6	-.107	.021	.176	.853
G5	-.167	.496	-.040	.647

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.[a]

a. Rotation converged in 11 iterations.

Opinions on questions 7 and 8 (Growth 1 and 2) have strong factor loadings to Factor 3. Students agreed they are likely to do anything as a result of intrinsic motivation, thus Factor 3 is regarded as îEndogenous growth mindsetï.

Table 2: KMO test

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.730
Bartlett's Test of Sphericity	Approx. Chi-Square	292.755
	df	66
	Sig.	.000

The Kaiser-Meyer-Olkin test (KMO- Table 2) is a measure of how appropriate the data are for factor analysis. The KMO value in this study was .730, which indicates that the sampling is appropriate. Bartlett's test is performed before applying factor analysis to cheque whether the data reduction technique can reasonably compress the data. In this study, the test statistic Chi-Square was 292.755 and the corresponding p-value was 0.000, which is less than the significance level (0.05). Thus, the data are suitable for factor analysis.

In column labelled (table3) " Extraction Sums of Squared Loadings. The" Total" column shows the Eigenvalues for each factor extracted which higher than 1. The second column "%of Variance" indicates the variance is explained by each factor (or component). The "Cumulative %"column shows the percentages of the total variance explained by the factors (62.418%).

Table 3: Total Variance Explained

Total Variance Explained									
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.276	27.298	27.298	3.276	27.298	27.298	2.893	24.105	24.105
2	2.059	17.158	44.456	2.059	17.158	44.456	1.626	13.546	37.651
3	1.134	9.451	53.906	1.134	9.451	53.906	1.575	13.123	50.774
4	1.021	8.511	62.418	1.021	8.511	62.418	1.397	11.644	62.418
5	.857	7.140	69.558						
6	.694	5.782	75.340						
7	.643	5.358	80.697						
8	.599	4.991	85.688						
9	.538	4.486	90.175						
10	.457	3.809	93.983						
11	.391	3.255	97.238						
12	.331	2.762	100.000						
Extraction Method: Principal Component Analysis.									

4.2. Learning Outcome: Mindset Assessment Processes

The results (Table 4) show that learning outcomes have a positive relationship with a fixed mindset ((Sig, .008) Fix3: Try to do what they are capable of doing. Fix4: Learn new things, but I don't think learning can increase my intelligence).

Fix3 refers to the affordances of learning that learners believe they can develop in all areas, even if only a little (Growth Mindset: G1). The results show a significant relationship between Fix3 and G1 (.016). Affordance of Learning is also positively related to Growth Mindset (sig. .001) that the learner believes they will learn more if they do something wrong (G2). Having a Fixed Mindset has a positive influence on G4, which refers to the learner being inspired by successful people (Sig. .000).

Table4 refers to what learners think about intelligence. They love to learn new things, but intelligence is fixed and cannot be increased by learning. The results show significant relationships between Fix4 and G1, G2, G3, and G4 (sig. .025,.013,.000,.002).

Table 4: Chi-Square tests

Variable	Value	df	Asymptotic Significance (2-sided)
GPA and Fix3	32.692a	16	.008
GPA and Fix4	26.695a	16	.045
G1 and Fix3	30.486a	16	.016
G1 and Fix4	28.788a	16	.025
G2 and Fix3	31.970a	12	.001
G2 and Fix4	25.368a	12	.013
G3 and Fix4	38.382a	12	.000
G4 and Fix3	50.059a	16	.000
G4 and Fix4	37.716a	16	.002

Having a Growth Mindset means believing in skill development, having a desire to learn more, seeing others, and being inspired by successful people. The mindset can promote the fixed mindset, which refers to how people think about the nature of intelligence and learning.

The findings (table5) show the positive relationship between G1 and G2. The lectures believe that the students can develop and practice in the classroom (Sig. .015). The pedagogy design such as practical tasks (G1) will encourage with comment or feedback (G4) during the activities (Sig. .033, .026).

The level design for learner's activities is positive impact on the feedback (G4) (Sig. .028) and Gamification strategy (Sig. .045). The class activities during learning such as gamification can encourage learners in terms of the new things can increase the intelligence (Sig. .039).

Table 5: ANOVA Tests (The lecture participants)

Variable	Sum of Squares	Df Between Groups	F	Sig.
G1 and G2	.742	1	6.576	.015
G1 and G4-1	.970	1	5.010	.033
G1 and G4-2	1.227	1	5.471	.026
G3 and G4-1	1.478	2	4.036	.028
Gamification and G3	3.976	2	3.436	.045
Gamification and learn for smart	3.480	2	3.625	.039

The results show in the table 6 that include the activities in the classroom. The participants illustrate the teaching experience to catch up the students' attention. The advisor experiences help to map the teaching or pedagogy style with gamification elements.

Table 6: Means comparison

Variable	Mean	Std. deviation
Gamification	4.36/5	.603
The feedback of progress	4.33/5	.816
The percentage level of progress(100)	37.33	29.49
The Practical activities	4.85/5	.364
The percentage of practical activities(100)	54.24	20.620
Inborn Intelligence	2.33/5	.924
Learning the New things can improve the intelligence	4.06/5	.747
Time in the Activities	4.15/5	.667
Time for Activity/each (minute)	41.66	18.819

Quest in the Activity	3.96/5	.728
Time for Quest/each activity (minute)	44.54	17.781
Social media feedback	3.60/5	1
Reward for the activity	4.15/5	.618
Percentage for reward/each activity (100)	15.64	12.36
Comment for individual	4.69/5	.466
Comment for all of students	4.54/5	.505

5. The Gamification Design for Affordances Pedagogy

From the previous study that map to the results in this study, the finding shows the relationship between factors. The significance of fixed mindset measures was positive impact through the growth mindset (Fig 4.). According to the research background and findings, the gamification mechanics can helps develop deeper insights into the capacity for pedagogy design. The findings support the ideas to develop the optimal psychology or flow theory [19] to suggests a remarkable activity in lesson plan. The gamification elements including points, badges, and leaderboards can derive her/his optimal experience [21]. For instance, the level and point can motivate people for engagement. An individualìs capacity to concentrate will impact their ability to experience flow [21].

5.1. The Gamification Elements

The design process aims to link the pedagogical design and gamification elements and this paper focuses only on the factors affecting the mindset. The conceptual model can be employed in future work. The model can apply to the pedagogical context design includes the promotion of tasks in each faculty, increasing motivation and encouraging desirable learning behavior. Data are collected through classroom observations and stimulated recall interviews. The core characteristics of growth mindset pedagogy include focus on process, mastery orientation, persistence, and individual student support [44].

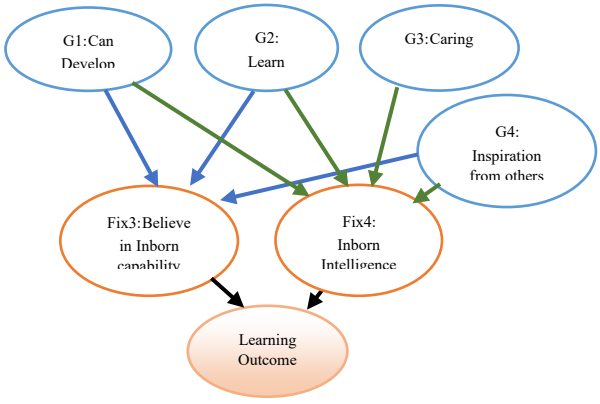


Figure 4: The relationships between factors

According to figure2, figure 5 shows the stimulus model for a growth mindset through motivation activities. The growth mindset factors are the stimuli, which can influence the learnerìs affordances.

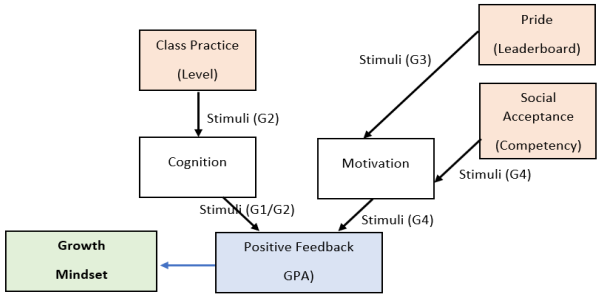


Figure 5: The Conceptual Gamification Affordances Design Model

5.2. The Affordances Pedagogy

The table 7 shows the findings that include the flow and pedagogy elements. These methods refer to the lecturersì experiences (Study 2) and studentsì feedback (Study 1). The gamification design for affordances pedagogy examines the literature review in terms of affordance, motivation and gamification. The research background aims to design the pedagogy elements below:

- Class assignment
- Quiz
- Group discussion
- Discussion
- Case study
- Play game

The pedagogy elements focus on the engagement that enable a practical mapping flow onto gamification. The approach enhances the studentìs interaction with the gamification elements such as points, badges, and leaderboard.

Table 7: The pedagogy elements of flow and gamification design (Adopted from Jones [45]).

	Element of Flow	Manifestation in Pedagogy Elements
1.	Task that we can complete	Class Assignment: The feedback of progress during class **Game element: Point**
2.	Ability to concentrate on task	Quiz (test): The feedback of progress or show the score **Game element: Point**
3.	Task has clear goals	Group Discussion: Quest in the Activity with friends and lecturer **Game element: Badges**
4.	Task provides immediate feedback	Class Assignment: Comment for individual/Comment for all of students

		Game element: Leaderboard
5.	Deep but effortless involvement	Discussion: social media/browsing the internet **Game element: Leaderboard**
6.	Exercising a sense of control over their actions	Case Study: The Practical activities/The percentage of practical activities around 50% per class **Game element: Badges**
7.	Concern for self disappears during flow, but sense of self is stronger after flow activity	Play game: The activity provides for a class environment as a simulation of life that refers to gamification such as reward, point and leader board in the class activity **Game element: Badges, point and leader**
8.	Sense of duration of time is altered	Class Assignment: Time in the Activities/Time for Activity and Time for case study for each activity around 40-45 minute **Game element: Leaderboard**

6. Discussion/Conclusion

Gamification design for affordances pedagogy focuses on engagement that provides a practical mapping flow to gamification activities during instruction. This approach could enhance student interaction through points, badges, and a leaderboard. A learning mindset could be created using motivation and affordance pedagogy. Affordance refers to points, badges, leaderboards and ranks in gamification elements. This research aimed to determine the affordance mindset factors of gamification interactive systems. The results show the relationships between factors that lead to desired learning outcomes. In accordance with Sailer, M., Homner, L. examined research topic "The Gamification of Learning: a Meta-analysis", the meta-analysis supports the claim that gamification of learning works because the results showed significant, positive effects of gamification on cognitive, motivational, and behavioral learning outcomes [46]. Similarly, there is a large body of work that clearly shows that incorporating gamification into the instructional process can lead to better student learning outcomes and helps to increase student achievement [47]-[50]. The findings derived from the current research could be used in a gamification affordances design method. The use of the model could help to improve the learner's affordances. The pedagogical guide could encourage the learner in the classroom. Engagement enables meaningful learning, which includes the quality of student effort, student interaction, and their immersive experiences during activities. The framework idea is a theoretical construct used to shape pedagogy and learner behavior. It represents the reasons for learners' actions, desires, and needs during instruction. Future research could test and revise the gamification design for the pedagogy of affordances.

References

[1] B. Ng, îThe neuroscience of growth mindset and intrinsic motivation,ï Brain Sciences, **8**(2), 20, 2018, doi: 10.3390/brainsci8020020.

[2] G.P. Kusuma, E.K. Wigati, Y. Utomo, L.K.P. Suryapranata, îAnalysis of gamification models in education using MDA framework,ï Procedia Computer Science, **135**, 385ó392, 2018, DOI: 10.1016/j.procs.2018.08.187.

[3] C. Silpasuwanchai, X. Ma, H. Shigemasu, X. Ren, îDeveloping a comprehensive engagement framework of gamification for reflective learning,ï in Proceedings of the 2016 ACM Conference on Designing Interactive Systems, 459ó472, 2016, doi: 10.1145/2901790.2901836.

[4] K. Rapeepisarn, K.W. Wong, C.C. Fung, M.S. Khine, îThe relationship between game genres, learning techniques and learning styles in educational computer games,ï in International conference on technologies for E-learning and digital entertainment, 497ó508, 2008, doi: 10.1007/978-3-540-69736-7_53.

[5] S. Harwood, N. Hafezieh, îíAffordanceí-what does this mean?,ï in 22nd UKAIS Annual Conference, St Catherineís College Oxford, UK, 4th-5th April 2017, 2017.

[6] T. McClelland, îThe mental affordance hypothesis,ï Mind, **129**(514), 401ó427, 2020, doi: 10.1093/mind/fzz036.

[7] S.K. Patrick, E. Joshi, îSet in Stoneï or íWilling to Growì? Teacher sensemaking during a growth mindset initiative,ï Teaching and Teacher Education, 83, 156ó167, 2019, doi: 10.1016/j.tate.2019.04.009.

[8] R.S. Alsawaier, îThe effect of gamification on motivation and engagement,ï The International Journal of Information and Learning Technology, 2018, doi: 10.1108/IJILT-02-2017-0009.

[9] A.J. Elliot, îApproach and avoidance motivation and achievement goals,ï Educational Psychologist, **34**(3), 169ó189, 1999, doi: 10.1023/A:1009009018235.

[10] M.R. Lepper, J.H. Corpus, S.S. Iyengar, îIntrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates.,ï Journal of Educational Psychology, **97**(2), 184, 2005, doi: 10.1037/0022-0663.97.2.184.

[11] R.M. Ryan, E.L. Deci, îIntrinsic and extrinsic motivations: Classic definitions and new directions,ï Contemporary Educational Psychology, **25**(1), 54ó67, 2000, doi: 10.1006/ceps.1999.1020.

[12] J. Tranquillo, M. Stecker, îUsing intrinsic and extrinsic motivation in continuing professional education,ï Surgical Neurology International, **7**(Suppl 7), S197, 2016, DOI: 10.4103/2152-7806.179231.

[13] F.A. Hodis, L.H. Meyer, J. McClure, K.F. Weir, F.H. Walkey, îA longitudinal investigation of motivation and secondary school achievement using growth mixture modeling.,ï Journal of Educational Psychology, **103**(2), 312, 2011, doi: 10.1037/a0022547.

[14] P.R. Clinkenbeard, îMotivation and gifted students: Implications of theory and research,ï Psychology in the Schools, **49**(7), 622ó630, 2012, doi: 10.1007/978-981-13-3021-6_15-1.

[15] D. Sharek, E. Wiebe, îMeasuring video game engagement through the cognitive and affective dimensions,ï Simulation \& Gaming, **45**(4ó5), 569ó592, 2014, doi: 10.1177/1046878114554176.

[16] A. Ahmed, M.J.D. Sutton, îGamification, serious games, simulations, and immersive learning environments in knowledge management initiatives,ï World Journal of Science, Technology and Sustainable Development, 2017, doi: 10.1108/WJSTSD-02-2017-0005.

[17] F.D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q., 319ò340, 1989, doi: 10.2307/249008.

[18] What is Usage Behavior. (Access March, 31 2021) https://www.igi-global.com/dictionary/usage-behavior/71901, 2021.

[19] M. Csikszentmihalyi, Play and intrinsic rewards, Springer: 135ó153, 2014, doi: 10.1007/978-94-017-9088-8_10.

[20] M. Csikszentmihalyi, M. Csikzentmihaly, Flow: The psychology of optimal experience, Harper \& Row New York, 1990, doi: 10.1080/00222216.1992.11969876.

[21] B. Cowley, D. Charles, M. Black, R. Hickey, îToward an understanding of flow in video games,ï Computers in Entertainment (CIE), **6**(2), 1ó27, 2008, doi: 10.1145/1371216.1371223.

[22] E.Kuindersma, J.van der Pal, J.van den Herik, A.Plaat, îBuilding a game to build competencies,ï in International Conference on Games and Learning Alliance, 14ó24, 2017, DOI:10.1007/978-3-319-71940-5_2.

[23] P.Wouters, E.D.der Spek, H.Van Oostendorp, îCurrent practices in serious game research: A review from a learning outcomes perspective,ï Games-Based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and Effective Practices, 232ó250, 2009, DOI: 10.4018/978-1-60566-360-9.ch014.

[24] R.M. Ryan, E.L. Deci, îSelf-determination theory and the facilitation of intrinsic motivation, social development, and well-being.,ï American Psychologist, **55**(1), 68, 2000, doi: 10.1037/0003-066X.55.1.68.

[25] O.Perski, A.Blandford, R.West, S.Michie, îConceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis,ï Translational Behavioral Medicine, **7**(2), 254ó267, 2017, doi:10.1007/s13142-016-0453-1.

[26] M.K. Pedersen, N.R. Rasmussen, J.F. Sherson, R.V. Basaiawmoit, îLeaderboard effects on player performance in a citizen science game,ï ArXiv Preprint ArXiv:1707.03704, 2017.

[27] R.Farzan, J.M.DiMicco, D.R.Millen, C.Dugan, W.Geyer, E.A.Brownholtz, îResults from deploying a participation incentive mechanism within the enterprise,ï in Proceedings of the SIGCHI conference on Human factors in computing systems, 563ó572, 2008, doi: 10.1145/1357054.1357145.

[28] G.Zichermann, J.Linder, îGamification revolution,ï 2013.

[29] A.P.A. STYLE, îGUIDE TO THE 6TH EDITION PUBLICATION MANUAL OF THE AMERICAN PSYCHOLOGICAL ASSOCIATION.ï

[30] R.R.Irwin, Cognition, Springer:65ó78, 2002.

[31] K.Park, îInnovative product usage behavior in the post-adoption process,ï ACR Asia-Pacific Advances, 1998.

[32] K.A. Abdel-Sater, îPhysiological positive feedback mechanisms,ï Am J Biomed Sci, **3**(2), 145ó155, 2011, ; doi:10.5099/aj110200145.

[33] R.Ronkainen, E.Kuusisto, K.Tirri, îGrowth mindset in teaching: A case study of a Finnish elementary school teacher,ï 2019, DOI: 10.26803/ijlter.18.8.9.

[34] P.Bouvier, E.Lavoué, K.Sehaba, îDefining engagement and characterizing engaged-behaviors in digital gaming,ï Simulation \& Gaming, **45**(4ó5), 491ó507, 2014, DOI:10.1177/1046878114553571.

[35] K. Cherry, îPositive reinforcement and operant conditioning,ï VeryWell Mind, 2018.

[36] G.Norman, J.Norcini, G.Bordage, Competency-based education:milestones or millstones?, 2014, doi:10.4300/JGME-D-13-00445.1.

[37] C.M.Mueller, C.S.Dweck, îPraise for intelligence can undermine childrenis motivation and performance.,ï Journal of Personality and Social Psychology, 75(1), 33, 1998, doi: 10.1037/0022-3514.75.1.33.

[38] E.OiRourke, E.Peach, C.S.Dweck, Z.Popovic, îBrain points:A deeper look at a growth mindset incentive structure for an educational game,ï in Proceedings of the third (2016)acm conference on learning@ scale, 41ó50, 2016, doi: 10.1145/2876034.2876040.

[39] R.P.French II, îThe fuzziness of mindsets:Divergent conceptualizations and characterizations of mindset theory and praxis,ï International Journal of Organizational Analysis, 2016, doi: 10.1108/IJOA-09-2014-0797.

[40] A.Derler, Growth Mindset Culture, Harvard Business Review, July 23.2018.

[41] C.S.Dweck, D.S.Yeager, îMindsets:A view from two eras,ï Perspectives on Psychological Science, **14**(3), 481ó496, 2019, doi: 10.1177/1745691618804166.

[42] W.Inchamnan, J.Chomsuan, îGamification Workflow for Growth Mindset Processes,ï in 2020 18th International Conference on ICT and Knowledge Engineering (ICT\&KE), 1ó6, 2020, DOI: 10.1109/ICTKE50349.2020.9289879.

[43] A.Hochanadel, D.Finamore, others, îFixed and growth mindset in education and how grit helps students persist in the face of adversity,ï Journal of International Education Research (JIER), **11**(1), 47ó50, 2015, DOI:10.19030/JIER.V11I1.9099.

[44] I. Rissanen, E. Kuusisto, M. Tuominen, K. Tirri, îIn search of a growth mindset pedagogy: A case study of one teacherìs classroom practices in a Finnish elementary school,ï Teaching and Teacher Education, **77**, 204ó213, 2019, doi: 10.1016/j.tate.2018.10.002.

[45] M.G.Jones, îCreating Electronic Learning Environments:Games, Flow, and the User Interface.,ï 1998.

[46] M.Sailer, L.Homner, The gamification of learning:A meta-analysis, 2020, DOI:10.1007/s10648-019-09498-w.

[47] J.C. Burguillo, îUsing game theory and competition-based learning to stimulate student motivation and performance,ï Computers \& Education, 55(2), 566ó575, 2010, DOI:10.1016/j.compedu.2010.02.018.

[48] C.-H. Chen, C.-H. Chiu, îEmploying intergroup competition in multitouch design-based learning to foster student engagement, learning achievement, and creativity,ï Computers \& Education, **103**, 99ó113, 2016, doi: 10.1016/j.compedu.2016.09.007.

[49] A.Dominguez, J.Saenz-de-Navarrete, L.De-Marcos, L.Fernández-Sanz, C.Pagés, J.-J. Martinez-Herráiz, îGamifying learning experiences: Practical implications and outcomes,ï Computers \& Education, **63**, 380ó392, 2013, doi: 10.1016/j.compedu.2012.12.020.

[50] W. Frkacz, îAn empirical study inspecting the benefits of gamification applied to university classes,ï in 2015 7th Computer Science and Electronic Engineering Conference (CEEC), 135ó139, 2015, DOI: 10.1109/CEEC.2015.7332713.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 147-153 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Vibration and Airflow Tactile Perception as Applied to Large Scale Limb Movements for Children

Hung-Chi Chu[1], Fang-Lin Chao[*,2], Liza Lee[3]

[1]*Department of Information and Communication Engineering, Chaoyang University of Technology, 436, Taiwan R.O.C.*

[2]*Department of Industrial Design, Chaoyang University of Technology, 436, Taiwan R.O.C.*

[3]*Department of Early Childhood Development & Education, Chaoyang University of Technology, 436, Taiwan R.O.C.*

ARTICLE INFO

Article history:
Received: 26 April, 2021
Accepted: 21 June, 2021
Online: 20 July, 2021

Keywords:
Vibrators
Zigbee module
Visual impaired
Airflow
Large movement

ABSTRACT

This study aimed to develop an airflow-vibrator motivated facility and assess exercise behaviors. The combination design involved computer-controlled airflow/ vibrators, a user interface program, and an adjustable structure presenting interaction options. The teacher and the participants can choose specific music with adjustable speed. The researcher did interviews during the initial test and field study. During the intervention, all participants succeeded in following the impinged flow with a positive emotional display. A wireless module and gas flow clue lifted the distance limitation of the vibration connection and enabled prompts in a larger area covered by radio waves. The flexible structure fit individuals ergonomic and the affordance consideration. After practicing, the students knew exactly how to pass and asked for the ball from the classmate. Wireless switch and signals give students more confidence in pitching — the participants successfully swap the body to follow the airflow.

1. Introduction

Rhythmic activities involve a combination of music, rhythm, and movement, which is intended to relax the mind and body. Today's teenagers show their creativity through hip-hop, and visually impaired individuals learn physical activities through touch and spoken instructions. For the visually impaired, spoken instructions are often unclear, and physical contact may cause discomfort to the other party. This study aimed to develop an airflow-vibrator-motivated facility and assess exercise behaviors. Although these technologies have gradually matured, there is no appropriate combination of teaching materials to form a reasonable price auxiliary tool. Thus, there is a gap connecting the technology elements and assistive teaching need of motivated facility and assessing exercise behaviors in school. The research work's contribution extends assistive technology of tactile perception to vibration and airflow, the combination of IoT devices, and the haptic feedback with a better user experience.

The combination of technology and design through tactile perception has been used to reduce the rhythmic learning difficulties of visually impaired children. Vibro-tactile feedback enhancement for orientation and obstacle avoidance can be obtained through the use of discreet actuators and obstacle detector sensors [1]. It provides frequent indications of useful dynamic information, such as level of proximity or distance. People can also perceive rhythm through tactile senses. In [2], the author described rhythm combined with frequency and amplitude to systematically produce 84 distinguishable tactile stimuli icons to help user perceptual tactile rhythm. In [3], the author presented unimodal and cross-modal rhythm perception with auditory, tactile, and visual modalities. Based on these findings, auditory and tactile modalities are suitable for presenting rhythmic information. Many assistive technologies have been successfully applied in education. Games that require physical activity are commonly used in rehabilitation to help restore physical function and balance [4]. Rehabilitation at home can reduce medical expenses and treatment time.

Visually impaired individuals often exhibit limited ability to participate in physical activities. Studies have found that visually impaired individuals take 50% longer to complete tasks with equivalent accuracy compared to sighted participants [5]. In [6], the author explained that movement is often limited by requiring outside guidance, fear of injury, and ridicule of others. However, when visually impaired individuals participated in bowling and

[*]Corresponding Author: Fang-Lin Chao, E-mail: flin@cyut.edu.tw

tennis video games with vibration and sound prompts, they exhibited improved mood and satisfaction following the activities.

Further study by [7] demonstrate the feasibility of real-time sensory substitution as a cost-effective approach for making gesture-based video games. In [8], the author presented a method of real-time video analysis to detect the presence of a particular visual cue as a cost-effective approach for making gesture-based video games. Users can express meaning in terms of vibration intensity or frequency, and haptics is also used for control, such as feedback when tapping an electronic pet. Feedback can be used as a reminder of movements in different parts, and studies have found that tactile sensation improves user engagement [9]. In [10], the author described the evaluation of human-computer interaction with the decision-making model. The information flow is helpful to monitor the effectiveness of the collective's activities. In [11], the author presented ubiquitous touch interaction with haptic feedback and show the movable world object supplies an accurate detection through a user study.

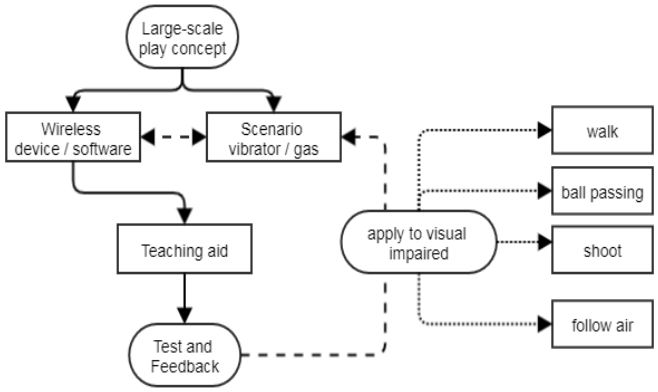
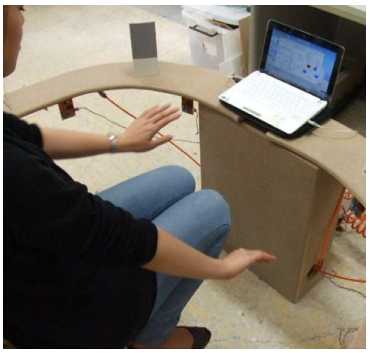


Figure 1: Design flow and applications for visually impaired

2. Detail Design

With the shrinking of the original size of the Internet of Things and the popularization of open platforms, researchers try to present the above ideas from the sensors and control components available on the market. This assistive tool involved a software development platform in performing human-machine interaction. Figure 1 indicates the design flow and possible applications for the visually impaired. First, start from the technical side, let those components be combined into the required system, and then look at the field of application and the idea of what kind of teaching materials teachers need. After the experiment in school, we extended it to the visually impaired (on the right) to apply it to different situations in their school life. The development platform and detailed parts in the design also show below.

As seen in Figure 2a, the vent is placed under the table for interacting with the user. The seated height of the participant was adjustable so that their legs remained at the same height as the air outlet so that their body could fully sense the airflow. As seen in Figure 2b, a large air compressor accommodates more compressed air. Figure 2b, top right shows the electromagnetic valve that controls the air outlet; bottom right shows the outlet of an exhaust pipe, which was an elongated slit that adjusted to the subject's position



(a)

(b)

Figure 2: Prototype of airflow assisted design: (a) use state; (b) compressor and spout.

A Phidget PC-USB interface board are utilized to provide an on/off sequence. As displayed in Figure 3, the interface control program has a graphical user interface using Microsoft Visual Studio with the current state of each switch. The button with a dark background indicates that it is currently activated. The text area on the right-hand side lists the sequence of the switching actions. A user can also manually control the switching by directly clicking on the buttons.

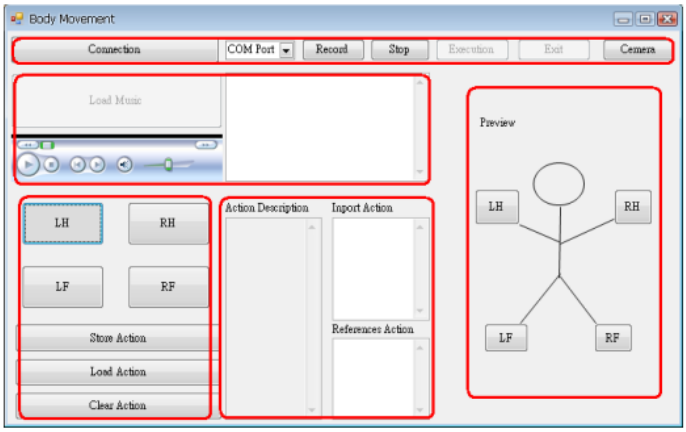


Figure 3. Software interface layout design with a set of the airflow outlet and timing control.

The total weighing of the vibrator is below 150 grams through circuit miniaturization to prevent additional loading during bodily movement. The vibrator is located in the skin contact surface with adjustable belts for proper contact. The control command could turn on/off the vibrator at a specific time. Each sensor module uses a 4.2 V lithium battery power. The vibrator is connected to a low-speed I/O pin on the sensor node to extend the module's usage. The ZigBee interface is IP-Link5501, and the sensor node is IP-Link1223. An NPDU consists of a network header and a network payload. The Network header contains frame control of 2 octets,

routing fields of 6 octets, and data payload NSDU (network service data unit) of variable length.

3. Collocation and Research of Auxiliary methods

After observing a class of visually impaired children, the researchers proposed tactile perception to assist in directing physical activity. Engineer submitted proposals for the method of tactile prompting: vibration and airflow through software control. For example, during the design of the airflow, the compressor air was controlled by the researchers. In addition, musical elements ensemble games required the synchronization of music with a pair of wireless vibrators.

Research simulated visual impairment with blindfolds and made detailed adjustments to the study after testing the design setup. Participants were expected to don vibrators on their left and right hands and beat a tambourine with the help of cues from the vibrators (Figure 4).

Figure 4: Wireless vibrators on group's dual hand and wireless control software

Vibration and airflow cues complement each other: Vibration is more subtle and can prompt specific body movement. In contrast, airflow is more generally perceivable and can direct a wide range of group or individual actions. Experiments show that vibration stimulates activity, while the sound field of an airstream can indicate the direction of the movement. This study attempts to implement the combination of vibration and airflow into teaching: first by using vibration cues to instruct physical movement and airflow cues to guide group activity. Second, apply these findings to educate visually impaired individuals in physical activities.

3.1. Method

Although there are many ways to combine assistive technology and applications, this study selected vibration /airflow solutions for testing. We tested two separate groups:

(1) Children between the ages of 9 and 10 without visual impairments. These children had previously participated in rhythmic activities.

(2) Visually impaired children with general amblyopia or total blindness between the ages of 9 and 12 whose physical ability was close to that of children with normal vision. However, careful observation revealed that these students exhibited limited balance and restricted movement. Their teacher reported hoping to guide them to participate in more activities.

Eight independent wireless vibration modules were set up in a system where a single host controlled the signal. When configured between different people in the same group, they could be guided by the system to work together. The advantage of wireless communication was that subjects using it could move freely.

3.2. Group teaching

During group teaching, the participants were given different individual activities through the prompts of the vibrator after being linked. As a result, the following ideas were put forward to expand applications in teaching.

Activity 1: Ensemble: increase children's concentration and sensitivity to rhythm through touch. The teacher instructed children to strike a tambourine when they felt a vibration, forming an ensemble rhythm.

Activity 2: AB mode: Increase children's focus through two touching ways: A for a single short vibration and B for continuous vibration. Participator performs specific actions corresponding to the different vibrations when prompted at irregular intervals (Figure 5). The goal was to see if the child could concentrate and distinguish between different vibration patterns.

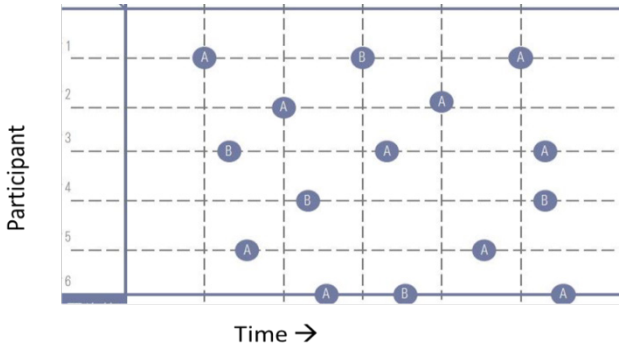


Figure 5: Simultaneous A/B vibration test

Activity 3: Consecutive A/B: Promote children's memory through tactile sensation and movement; promote children's creativity through music activities. One child was instructed to perform a motion after feeling the vibration signal. The next child was asked to imitate the action of the previous child in response to the vibration signal and then add a second motion to the sequence. A third child was then asked to perform the two last moves and add a third motion when they felt the vibration signal.

The jet sound is received through tactile and auditory cues. Air rhythmic action is the primary stimulus, supplemented by vibrations to prompt a change in the application, such as in the three activities above. When the child's back was facing the airflow, the study design proposed a scenario of moving a wall to instruct the child to push the invisible wall with his hand so that the child could express creativity during movement. When vibration is combined with airflow, large-scale airflow can guide actions in large areas to different locations, such as forward or backward ambulation, walking around a circle, or pointing in a specific direction.

3.3. Evaluation Criteria and Behavioral Assessment Tools

Researcher used the following pediatric behavioral assessment tools:

(1) Wenlan Adaptation Behavior Scale

The Wenlan Adaptation Behavior Scale (Motor Skills Section), which is used to measure motor skills development in children,

was completed at baseline. The test was conducted to track the changes in the children's physical ability.

(2) Observation and Evaluation Record Form for children's activities

The observation records of children were also used as a research tool for participatory observation research. Under controlled circumstances, systematic observations of phenomena or individual behaviors were based on established research goals. Observers recorded the performance of the assessed scores in writing. Finally, observer made an objective explanation of the phenomenon or individual behavior. The observation variables are defined as following: week-activity-participant-vibration mode; so that notation (1-2-M-B) means teaching week-1, playing activity-2, with participant-M, using continuous vibration mode-B.

The content of the observation table was divided into structured assessments, supplemented by available text supplements, and recorded by observing the child's reaction and concentration during movement. The teacher adopted a Likert five-point scoring method for evaluation. A score of 5 indicated frequent behavior observed during the activity. Three observers rated the frequency of each behavior according to the situation described in the items and finally calculated the average result of the observation score. The evaluation items included:

(1) Movement:

Students respond to vibrator and waves limbs. Uses instrument correctly according to instructions and beats instrument in time to the music. Responds to and moves limbs according to airflow position.

(2) Attention

Senses change in vibration mode. Uses hearing and physical perception to pay attention and engage to the airflow.

(3) Emotion

The student expresses positive emotion through spoken words, movement, and expression. Maintains positive feeling throughout the airflow process.

(4) Creativity

Moves in varied ways in the designated space according to the vibration prompts. Swings body and waves limbs were corresponding to airflow.

4. Results and Discussion

During the first week, observers found that the children were unsure how to react to the vibrators, which caused them to respond incorrectly. In addition, they did not make prominent movements in response to a prompt for free creative actions because of personal shyness. In the second week, the teacher added explanations, introduced movement simulation, and carried out group activities using airflow stimulation. We found that children exhibited significantly different motion feedback and even actively added new motions. The research found that children were integrated into the activity as a whole and actively with new movements in third week. Further details are as follows:

4.1. Week 1 record

Activity 1: Ensemble

1. The children were unfamiliar with the vibrator and could not accurately sense it. Therefore, the performance is not apparent (1-1-M-A) (Figure 6).

2. The children immediately shook their arms when both hands were shaken and showed their arms after a pause (1-1-M-B).

3. After putting on the vibrator, the children kept looking at it. They started by making fists and frowning. (1-1-A-B).

4. The children smiled before the start of the activity. They laughed and jumped on the tambourines. (1-1-E-B)

5. The children beat the tambourines correctly from beginning to end, maintaining a single movement pattern without change. (1-1-E-B)

(a)

(b) (c)

Figure 6: Student behaviors of three events: (a) ensemble, (b) A/B mode, and (c) A/B two-point.

Activity 2: A/B mode

1. The children could not clearly distinguish between A and B vibration patterns and demonstrated a high rate of action repetition. (1-2-M-A)

2. A small number of children could make the correct movements but could not clearly distinguish patterns when making inappropriate movements (1-2-M-B).

3. When the children wore the vibrator, they frowned and looked at it. They appeared quiet during the activity. (1-2-A-B)

4. Most of the children smiled during the activity, indicating excitement to participate. (1-2-E-B)

5. The child exhibited no change in movement and remained in place. (1-2-C-B)

Activity 3: Consecutive A/B

The instructions were not followed precisely and the children required cueing from the instructor. (1-3-M-A)

Some children responded quickly to the signal, but some needed a reminder from their peers. (1-3-M-B) (Figure 6b)

The children devoted themselves to the activity and paid attention to their vibrator. One child said, "Yeah, I'm here," immediately after the vibration. (1-3-A-B)

The children exhibited happiness through expression and body language engaged in the activity. (1-3-E-B)

The movements of the children were almost all performed while running (one was jumping). (1-3-C-B)

In response to wearing a wireless vibrator, the children's overall body language was stiff, and their movement rarely changed. While some exhibited creativity and came up with their ideas, most of them repeated the body movements of others. Some children could focus on the vibrator on their wrists and listen to the instructions of the tester. The children looked nervous but exhibited no negative emotions.

4.2. Week 2 record

Activity 1: Ensemble

Under the instructor's guidance, the children's ability to tap to the music improved, but the tapping method did not change. (2-1-C-A)

The children correctly responded to changes in vibration and beat the tambourines at the same time. When only one vibrator signaled them, most of them only moved one hand. (2-1-M-B)

The children were fully involved in the activity and could primarily respond correctly. (2-1-A-B)

The children laughed frequently and waved their arms, exhibiting positive emotions. (2-1-E-B)

The children's movements were primarily similar, and the rhythm was maintained when the facilitator participated. (2-1-C-B) (Figure 7a)

Activity 2: AB Mode

The children were still unable to create new movements because of nervousness. (2-2-C-A)

The children were able to perform AB mode correctly and respond to different prompts. (2-2-M-B)

The children demonstrated a high willingness to participate. Children were smiling and showing positive emotions during activities. (2-2-E-B)

They maintained the same movement with no other limb movements demonstrated. (2-2-C-B)

(a) (b)

Figure 7. Week 2 study: (a) ensemble (b) A/B two-point

Activity 3: Consecutive A/B

The children correctly judged the location of the airflow. (2-3-M-A) (Figure 7b)

The children made significant progress in focus and emotions. (2-3-A, E-A)

There were innovative movements. (2-3-C-A)

Most children could feel the vibration correctly, but one child happily ran around with the airflow; others only had physical movements. (2-3-M-B)

All of the children could focus on activities and follow the action. (2-3-A-B)

The children demonstrated a high willingness to participate and showed positive emotions throughout. (2-3-E-B)

The children exhibited 3 distinct movements; in addition to the original running and jumping, they were also sidestepping. (2-3-C-B)

After the participating children donned wireless vibrators and performed activities with airflow, the children occasionally need to be reminded by the testers. The overall magnitude of limb response became more significant, and the children were willing to cooperate actively to maintain rhythm (Figure 8). After growing familiar with the rhythm, they could focus on vibrational changes and react instantly. When students improved their sense of familiarity, the children had a high willingness to participate, showing positive emotions and smiling.

4.3. Week 3 record

The child feels the way of shaking (3-1-M-A). They demonstrated a high rate of repetitive movements, so they did not perform well in creativity. (3-1-C-A)

Most children could correctly sense the vibration pattern, but a few children exhibited slightly longer reaction times. (3-1-M-B)

The children engaged in the activity. (3-1-A-B)

The children demonstrated a high willingness to participate, showing positive emotions. (3-1-E-B)

Two children performed somersaults, swinging the tambourine from top to bottom. (3-1-C-B)

The children's energetic performance improved significantly, and they could respond to tactile vibrators, and air prompts. They were also more engaged in activities. During the exercise, they cooperated with a high degree of concentration. The children talked naturally with the tester and performed activities with enthusiasm; at the end, they actively asked when they could participate again (Figure 8). During this period, children engaged in simple competition with their peers, so more innovative actions appeared, such as somersaults and trunk rotations. Overall, the children's performance in terms of limb movement, concentration, emotions, and creativity improved significantly.

(a) (b)

Figure 8: Week 3 study: (a) airflow + vibrator, (b) ensemble and creative move

4.4. Overall Performance

Haptic perception successfully assisted motor skill development in children. In terms of overall performance, children have noticeably changed four goals: movement, attention, emotion, and creativity. The children's performance may have been limited due to unfamiliarity with the tester in the early stages. In the middle period, movements and expressions changed gradually as the children grew more comfort. The sensation of airflow was highly attractive to the children. When airflow appears in activities, the children seemed excited and did not want to leave so they could continue playing.

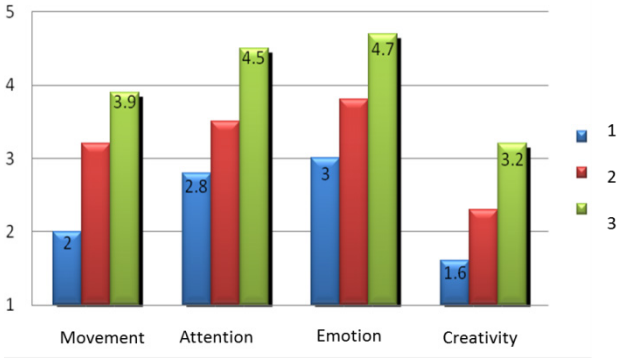


Figure 9: Children's overall performance during three weeks

The quantitative data (using the SPSS) of three class activities are analyzed and discussed below. Figure 9 shows the evaluation results using the Likert five-point scoring method. The children's overall performance in the later period improved, not only in more physical movements but also in the position of the vibrator and airflow. New actions appeared, such as somersaults. After the activity, the children asked if they could participate again, indicating that the activity was attractive. With appropriate syllabus design, children can demonstrate positive physical and emotional merges.

5. Applications for Visually Impaired Children

Visually impaired children with general amblyopia or total blindness, aged between 9-12 years, possess physical ability close to children with regular sight. Before the activity, however, observation revealed that the visually impaired students exhibited limited balance ability. In addition, they usually exhibit restricted movement. Because of safety concerns, the teacher gave them many constraints; for example, they had to go in pairs one by one when they went outside the classroom. When the teacher asked them to stand on one foot, they appeared unstable and eager to reach for the objects next to them for support. Their movements were less varied, and they presented with stiff posture. The teacher hoped to guide them to engage in more activities using the assistive devices. The wireless module lifted the distance limitation of the vibration connection and enabled prompts in a larger area covered by radio waves signal.

The scenario items included: ensemble, orientation, and passing ball, shoot the ball, and follow the change of airflow. Firstly, a small-scale practice using the previous ensemble and A/B game. The students were given different instruments instructed to tap the instrument through music and vibrator cues. The vibrator cues provided a rhythmic tacit understanding of ensemble participation. Visually impaired students have a good sense of sound and rhythm, students successfully played with musical instruments and A/B game.

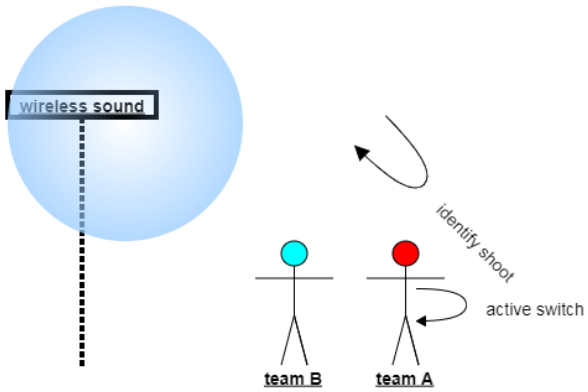


Figure 10: Schematic drawing of the large-scale activities: vibrator cues provided walking instructions, students improved pitching in a different location.

Secondly, six pupils with amblyopia in special schools participated in four kinds of large-scale movement play for 60 minutes (Figure 10).

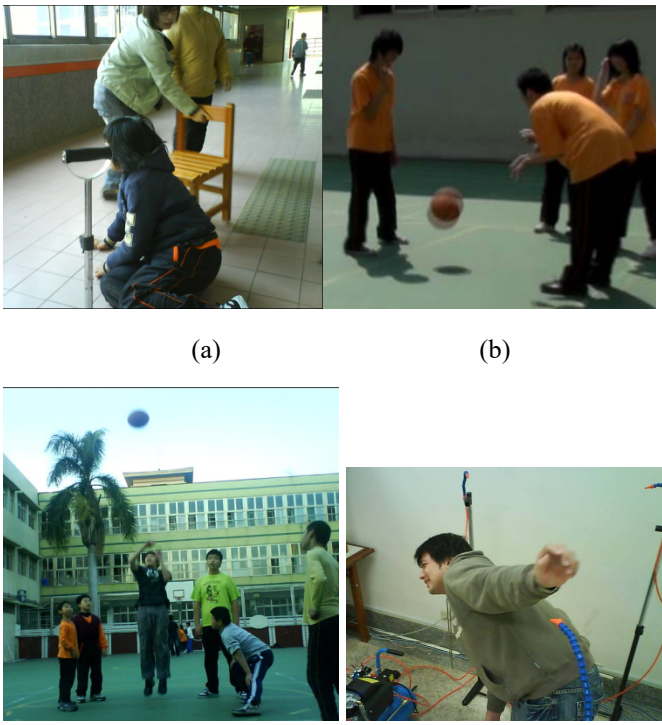
(1) Orientation: Vibrator cues provided directions and walking instructions. Testers used vibrators to indicate right and left to help children reach their destination.

(2) Passing practice: Vibrator cues are used to instruct children to throw a ball back and forth and improve communication efficiency. The vibrator provided the teacher's instructions to pass and receive the ball.

(3) Shooting: We set up a wireless sounder on the basketball frame. When students receive the ball and want to shoot, they start the sound source with the wireless switch and judge the direction base on the sound from target area. We found that wireless signals give students more confidence in pitching.

(4) Quickly follow the change of airflow: promptly change the air outlet point, the participants rotate the body to monitor the airflow to form a continuous action creatively.

Sometimes students were out of sync during orientation due to the speed and circuit delay. After practicing, the students knew exactly how to pass and asked for the ball from the team classmate (Figure 11). Wireless switches give students more confidence in pitching it. The participants swap the posture to follow the airflow successfully. Through the interview, the vibrator and gas clues enabled more interaction in large-scale exercises. Wireless technology and moving gas help remote instructions to assist users in interacting with other people, guidance from teacher, and the signal in the surrounding environment.



(a) (b)

(c) (d)

Figure 11. Large scale activities: (a) vibrator cues provided walking instructions in orientation practice (children walk along the path), (b) students pass the ball according to vibration cues, (c) students improved pitching in a different location, and (d) dynamic gas clues for posture following.

6. Conclusion

The aim of the research is developing tactile perception assistive technology to activate the physical activity of children. The design method utilized vibration and airflow actuators and a control system to provide proper feedback and acknowledgment. The study also developed the corresponding instructions and introducing context to promote children's participation. Three weeks evaluation process conducted through field practice in elementary school with guiding activities. Both explaining hints and free creative activities were encouraged in classroom group cooperation activities. They were also able to recognize different vibration patterns that respond to the corresponding limb movements. The field testing results indicated children's positive responses and spontaneous creative actions. Their overall performance in exercise, attention, emotion, and creativity improved significantly during the evaluation instructions. The research extends assistive technology of tactile perception to vibration and airflow. The combination of new IoT devices of wireless sensor networks and remote control of haptic feedback are helpful to visually impaired individuals, which support better instruction in the future classroom.

Acknowledgment

This work was supported in part by the National Science Council, Taiwan, ROC, under grant NSC 101-2221-E-324-028, and 99-2221-E-324-026-MY2.

References

[1] G. Ghiani, B. Leporini, F. Paternò, "Vibrotactile feedback to aid blind users of mobile guides," Journal of Visual Languages & Computing, **20**(5), 305-317, 2009, doi: 10.1145/1409240.1409306.

[2] D. Ternes, K. E. Maclean, "Designing large sets of haptic icons with rhythm," in 2008 International Conference on Human Haptic Sensing and Touch Enabled Computer Applications, 199-208, Heidelberg, 2008, doi: 10.1007/978-3-540-69057-3_24.

[3] M. Jokiniemi, R. Raisamo,J. Lylykangas, "Crossmodal rhythm perception," in International Workshop on Haptic and Audio Interaction Design, 111-119, Heidelberg, 2008, doi: 10.1007/978-3-540-87883-4_12.

[4] B. Lange, C.Y. Chang, E. Suma, B. Newman, "Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor," in 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1831-1834, 2011, doi: 10.1109/IEMBS.2011.6090521.

[5] S.A. Douglas, S. Willson, "Haptic comparison of size (relative magnitude) in blind and sighted people," in 2007 Proceedings of the 9th international ACM SIGACCESS Conference on Computers and Accessibility, 83-90, 2007, doi: 10.1145/1296843.1296859.

[6] D. Morelli, E. Folmer, "Real-time sensory substitution to enable players who are blind to play video games using whole body gestures," Entertainment Computing, **5**(1), 83-90, 2014, doi: 10.1016/j.entcom.2013.08.003.

[7] D. Morelli, J. Foley, E. Folmer, "VI-Tennis: a vibrotactile/audio exergame for players who are visually impaired," in 2010 Proceedings of the Fifth International Conference on the Foundations of Digital Games, 147-154, 2010, doi: 10.1145/1822348.1822368.

[8] L. Zoccolillo, D. Morelli, M. Iosa, "Video-game based therapy performed by children with cerebral palsy: a cross-over randomized controlled trial and a cross-sectional quantitative measure of physical activity," Eur J. Phys Rehabil Med, **51**(6), 669-76, 2015, doi: 10.14288/hfjc.v10i1.225.

[9] J. Yim, N. Graham, "Using games to increase exercise motivation," in 2007 Proceedings of Conference on Future Play, 166-173, 2007, doi: 10.1145/1328202.1328232.

[10] A.H. Hoppe, F. Marek, F.v.d Camp, R. Stiefelhagen, "Extending movable surfaces with touch interaction using the virtualtablet: an extended view," Advances in Science, Technology and Engineering Systems Journal., **5**(2), 328-337, 2020, doi: 10.25046/aj050243.

[11] N. Bakanova, A. Bakanov, T. Atanasova, "Modelling human-computer interactions based on cognitive styles within collective decision-making," Advances in Science, Technology and Engineering Systems Journal, **6**(1), 631-635, 2021, doi: 10.25046/aj060169.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 154-160 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

A Novel De-rating Practice for Distributed Photovoltaic Power (DPVP) Generation Transformers

Bonginkosi Allen Thango*, Jacobus Andries Jordaan, Agha Francis Nnachi

Dept. of Electrical Engineering, Tshwane University of Technology, Emalahleni, 1034, South Africa

ARTICLE INFO

Article history:
Received: 08 March, 2021
Accepted: 28 April, 2021
Online: 20 July, 2021

Keywords:
Transformer
Losses
Hotspot temperature rise
Finite Element Method
De-rating

ABSTRACT

Transformers are habitually designed and manufactured for operation at a fundamental frequency of 50Hz and sinusoidal load current. Transformers are susceptible to non-linear loads. The inception of switching action characterises Non-linear loads and consequently nonsinusoidal load current which brings about higher transformer service losses, hotspot temperature rise, and degradation of cellulosic and liquid insulation, and consequently untimely failure of transformers during service. This phenomenon yield current with different components that are multiples of the fundamental frequency of the distributed photovoltaic power (DPVP) generation system. In order to obviate these challenges, the continuous power rating of the transformer, which is intended to facilitate non-linear loads must be minimised using procedure ascribed by the standards as de-rating. This work, an extension of previous work, proposes a novel procedure by means of Finite Element Method (FEM) for the de-rating of DPVP transformers serving non-linear loads during their service life. The proposed procedure considers parameters such as skin effect, proximity effect, and the magnetic flux leakage on the windings that were not included in the IEEE recommended de-rating procedure. The theoretical examination is substantiated on a 500kVA, three-phase, oil-filled transformer.

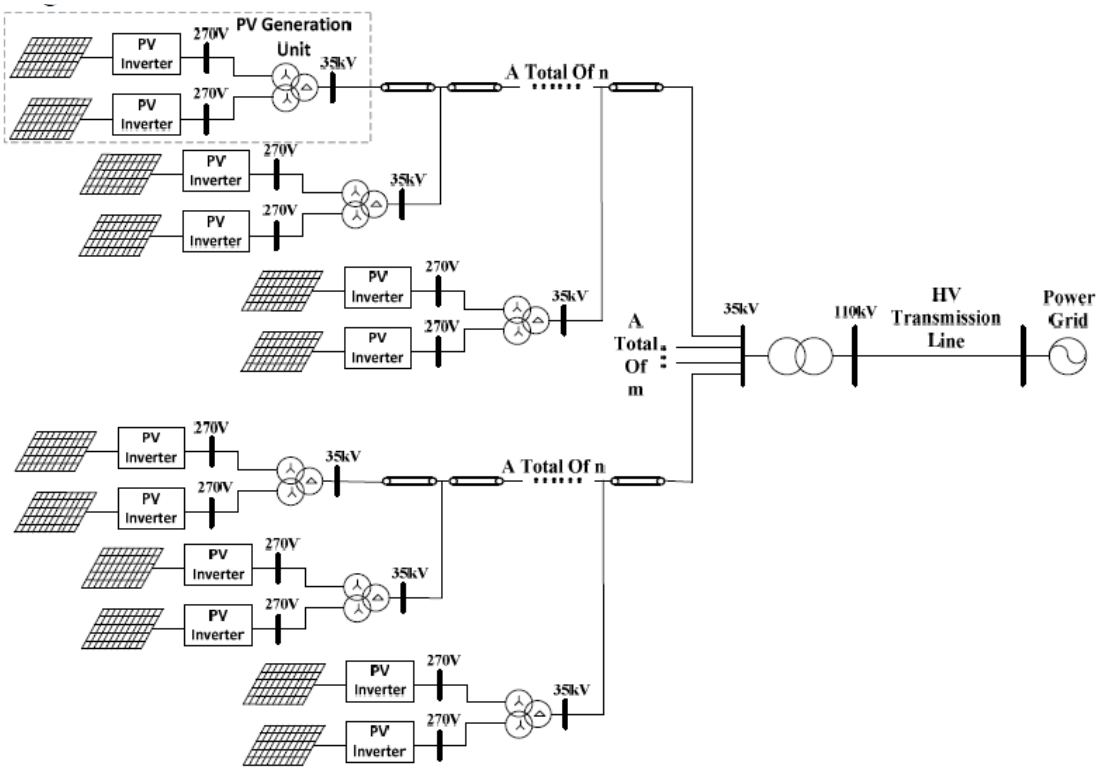
1. Introduction

In this day and age, distributed photovoltaic power (DPVP) generation producers are apprehensive concerning allotting of power ratings for transformers that are projected to operate in harmonically contaminated environment. The use of regular distribution transformers has so far proven to be impotent in handling the operational requirements of DPVP generation during service. To pledge future reliability of DPVP generation, transformers facilitating this application must embed these requirements, in which includes sporadic loading cycle, harmonic and distortion and de-rating in the design philosophy. Relying on conventional regular distribution transformer design philosophies can result in conditions where surged voltages and current could abbreviate a DPVP transformer's service life. Owing to the economic status of competitive rates for DPVP projects and since transformers are ordinarily not operated at rated loading, there is a tendency to de-rate the transformer. There may even be a propensity to oversize the transformers to increase the winding Eddy losses due to harmonic currents seen by the transformer during service. Considering that DPVP transformers are not rated at regular distribution transformer ratings, effort to adjust it to the standard kVA rating is laborious. Standard DPVP generating ratings generally lie between standard distributing transformer ratings, so there appears to be an inclination to select the nearest in spite of the power being de-rated.

Studies that consider the de-rating of transformer ratings when supplying harmonic currents are prevalent in the publications [1] – [7]. These publications are based on case studies of measured results and analytical formulations. The approach on this study have a similar downside of not accounting for significant parameters suchlike skin effect, proximity effect, and the magnetic flux leakage on the windings at fundamental and under harmonic conditions.

A configuration of a DPVP plant is demonstrated in figure 1 [8] and comprises of a cluster of PV generation inverter schemes, in which the generated power is collected by the power retrieval system into a 35kV bus. In the PV generation inverter schemes, two PV arrays are connected with PV inverters and into a 500kVA transformer through a LCL filter and then accessible to the station PV energy retrieval system. The generated energy is fed into the national grid through a high voltage (HV) transmission line.

*Corresponding Author: Bonginkosi Allen Thango, thangotech@gmail.com



In view of the intermittent nature of the DPVP plant and the switching action of inverters, in case the harmonic current output of an individual PV generation inverter scheme is minimal, the output harmonic current of the cluster of PV generation inverter schemes has the potential to exceed the limitations recommended by the standard [9].

In [10], transformer losses that occur when facilitating solar PV farm environment were investigated based on seasons. The study in particular draw attention to the injection of harmonics and distortion using the IEEE Std. C57.110-2018 recommended practice. This work, an extension of the previous work in [10], examines the influence of harmonics and distortion on transformers during their service lifetime. The work further review the de-rating procedure recommended by IEEE for transformers under harmonic conditions. The work then proposes a novel procedure for de-rating transformers using the computational power of FEM. The proposed procedure takes into account of parameters suchlike skin effect, proximity effect, and the magnetic flux leakage that the analytical method (AM) recommended by the IEEE fails to take into consideration.

2. Effect of harmonic currents

During service the total transformer losses (P_{TOT}) are comprised of the no-load loss (P_{NL})and the load losses (P_{LOAD}) as expressed in eq. (1) below.

$$P_{TOT} = P_{NL} + P_{LOAD} \quad (1)$$

The service no-load losses are as a result of the core excitation when the transformer is connected to the supply voltage and is independent of the loading profile. Under harmonic conditions, the harmonic current passing through resistance and leakage reactance of the transformer may deform the output voltage. Practical experience has demonstrated that the rise in temperature in the core is not a factor limiting the evaluation of the permissible current under harmonic conditions. Evidently, the C57.110-2018 [11] recommended practice for establishing transformer capability under harmonic conditions does not take the account the no-load losses under such conditions.

The load losses (P_{LOAD}) are comprised of the copper losses (I^2R) and the winding stray losses in which are constituted by the winding Eddy losses (P_{WEC}) and stray loss in structural parts (P_{OSL}) as expressed in eq. (2) below [11].

$$P_{LOAD} = I^2R + P_{WEC} + P_{OSL} \quad (2)$$

The copper losses signifies the heat dissipated by the load current in the transformer winding conductors. Under harmonic conditions if the load current increases, then the copper losses will also experience an increase. During the factory acceptance testing of a transformer there is no methodology available for the testing of the winding stray losses. Although, the impedance test can be employed to ascertain the total transformer losses. Then the winding stray losses can be acquired by deducting the copper losses from total transformer losses. In the event that the rated winding Eddy loss of a transformer is known, then this loss under harmonic conditions can be evaluated as expressed in in eq. (3) [11].

$$P_{WEC} = P_{WEC\_R} \times \sum_{h=1}^{h_{max}} \left[\frac{I_h}{IR}\right]^2 \times h^2 \quad (3)$$

The stray loss in structural parts can be evaluated for harmonic conditions by applying a similar procedure. In the 4th edition of

the UL 1561 standard published in 2011[12], the Factor-K for de-rating a transformer is specified as expressed in eq. (4) below.

$$K_{factor} = \left[\sum_{h=1}^{hmax} {I_h}^2\right] \times F_{HL} \quad (4)$$

The Factor-K indicates the effect of harmonic conditions upon the increase in the winding Eddy losses of the transformer. In the C57.110-2018 standard [11], the harmonic loss factor to account for the increase in the winding Eddy losses is expressed as follows in eq. (5).

$$F_{HL} = \frac{P_{WEC}}{P_{WEC\_R}} = \frac{\sum_{h=1}^{hmax} {I_h}^2 \times h^2}{\sum_{h=1}^{hmax} {I_h}^2} \quad (5)$$

On the empirical studies conducted in [13], for evaluating the winding Eddy losses, eq. (6) and eq. (7) are established.

$$P_{WEC\_R} = \frac{0.8 \times P_{WEC\_R}}{{I_2}^2 I_L} \quad (6)$$

$$P_{WEC\_R} = \frac{2.8 \times P_{WEC\_R}}{3 \times {I_2}^2 I_L} \quad (7)$$

The maximum allowable harmonic load current formula recommended by the C57.110-2018 standard [11] of a transformer is the current whereupon the maximum winding Eddy loss ratio as expressed in eq. (8) below.

$$I_{max} = \sqrt{\frac{1+P_{WEC\_R}}{1+F_{HL} \times P_{WEC\_R}}} \quad (8)$$

3. Case scenario

In this case scenario, the technical characteristics of the studied 500kVA, three-phase, oil-immersed, DPVP transformer are presented in Table 1. The objective herein is to investigate the operational performance of the studied transformer that facilitate a harmonically distorted load current during service.

Table 1: Technical specification

Item	Value
kVA rating	500kVA
HV Voltage/HV Current	11kV/1.44A
LV Voltage/LV Current	400V/79.2A
Load Losses (@75% loading)	1375W
HV winding resistance	121.5 ohms
LC winding resistance	0.03 ohms

The corresponding harmonic load current considered in the analysis is presented in Figure 2. Further, the de-rating information of the studied transformer according to the analytical method and the proposed FEM procedure will be computed based upon the supplied harmonic load current.

In preparation to shed light on the concept of skin effect on the winding conductors at the fundamental frequency and under harmonic conditions, the two simulations with a round cross-section were investigated. These simulations substantiate the behaviour of the load current due to the applied frequency. At fundamental frequency, the effect of the skin effect upon the winding conductors will be as presented in Figure 3.

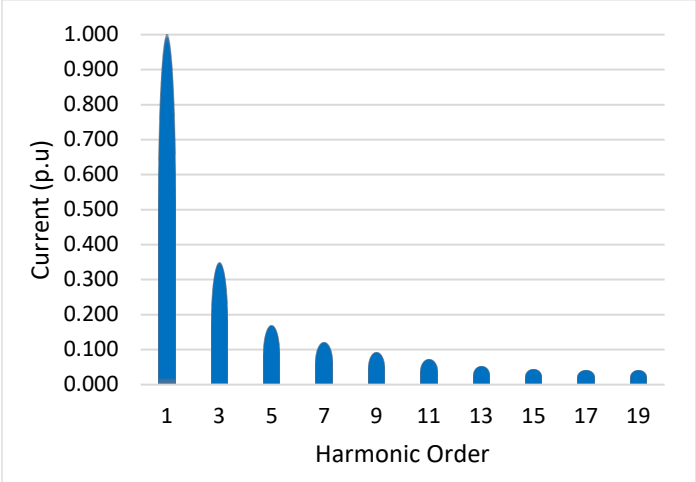
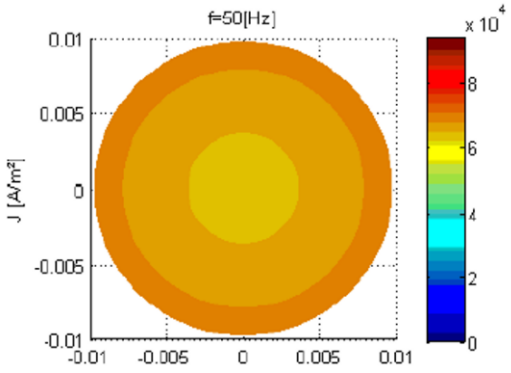
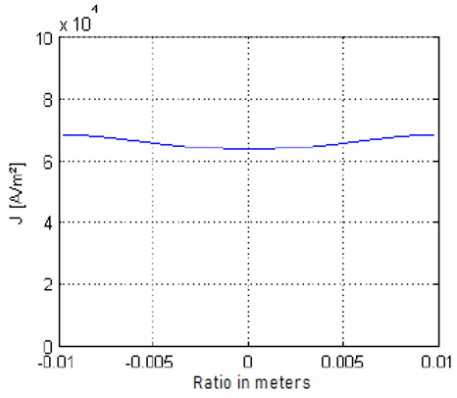


Figure 2: Harmonic spectrum.



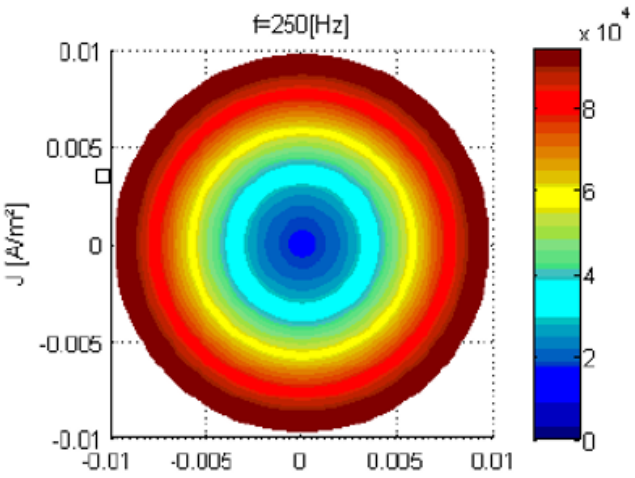
(a)



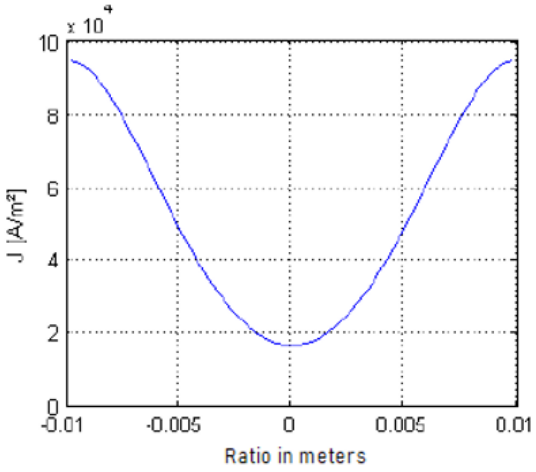
(b)

Figure 3: Winding conductor's skin effect at a fundamental frequency

As anticipated, Figure 3 (a) illustrate a uniform dispersion through the winding conductor. The graphical representation of this effect is shown in Figure 3 (b). The 5th order of the supplied harmonic profile was investigated and the corresponding results in Fig. 4 demonstrate a concentration of the current upon the surface of the winding conductor as demonstrated by Figure 4(a). The graphical representation of this effect is also presented in Figure 4 (b).



(a)



(b)

Figure 4: Winding conductor's skin effect at 5th harmonic order

4. Results

In this section, the results for de-rating a transformer to facilitate the supplied harmonic load current using the analytical method and the proposed 3D FEM procedure.

4.1. Classical Approach: Analytical method

The winding Eddy losses at fundamental and under harmonic conditions are acquired by the difference between the total transformer losses and copper losses as:

$$P_{WEC\_R} = 1375 - 3 \times [1.44^2(121.5) + 79.2^2(0.03)] = 54{,}635W$$

The peak ratio of the winding Eddy losses is evaluated in p.u using eq. (6) given that the studied unit is within the category up to 650kVA.

$$P_{WEC\_R} = \frac{0.8 \times 54.635}{79.2^2 \times 0.03} = 0.232\ p.u$$

In order to attain the Factor- K, the supplied harmonic load current is applied, as shown in Table 2.

Table 2: K-factor estimation

h	I_h(A)	$I^2{}_h$(A)	h^2	$I^2{}_h \times h^2$(A)
1	1,000	1,000	1,000	1,000
3	0,350	0,123	9,000	1,103
5	0,170	0,029	25,000	0,723
7	0,120	0,014	49,000	0,706
9	0,092	0,008	81,000	0,686
11	0,071	0,005	121,000	0,610
13	0,051	0,003	169,000	0,440
15	0,043	0,002	225,000	0,416
17	0,040	0,002	289,000	0,462
19	0,039	0,002	324,000	0,493
Σ		**1,187**		**6,637**

The maximum allowable current of the studied transformer based in eq. (8) is evaluated as follows:

$$I_{max} = \sqrt{\frac{1 + 0.232}{1 + 6.637 \times 0.232}} = 0.696$$

The maximum allowable load current in Ampere (A) is evaluated as:

$$I_{max} = 0.696 \times 79.2 = 55A$$

The continuous kVA rating of the transformer during service is evaluated as:

$$Equivalent\ kVA = 500 \times 0.696 = 347.263kVA$$

During service, this equivalent kVA is the valuation of the continuous power rating the transformer will be operating on under the supplied harmonic spectrum. In the event the utility owner changes the harmonic loading seen by the transformer, the kVA must be re-evaluated

4.2. Finite Element Method: Proposed Procedure

In the development of the proposed procedure, Ansys Maxwell, an electromagnetic field simulation software, is employed. A 3D cross-section of the three-phase, oil-immersed, DPVP transformer is developed for the finite element software. The short circuit test of the transformer is evaluated using the circuit model at fundamental and under harmonic conditions. The superposition property of the winding Eddy losses is employed in the event when the winding conductor dimensions are not more than the skin and proximity effect as well as when the flux degree is lesser than the saturation level. At large, the superposition property is employed to evaluate the transformer load losses under harmonic conditions. This loss is thereby the arithmetic sum of the losses due to different harmonic load current and harmonic orders. For the proposed FEM procedure, the windings' leakage flux and resistance are used to calculate the different loss components.

In Figure 5, the 3D FEM cross-section of the studied transformer is presented. The model takes into account the actual geometries, material properties and the supplied harmonic spectrum. Undoubtedly, the finer meshing of the geometries takes an extended processing time but leads to the most optimised solution. Consequently, in the assessment of the various harmonic

orders, a concession has been formed between the optimised solution and extended processing time.

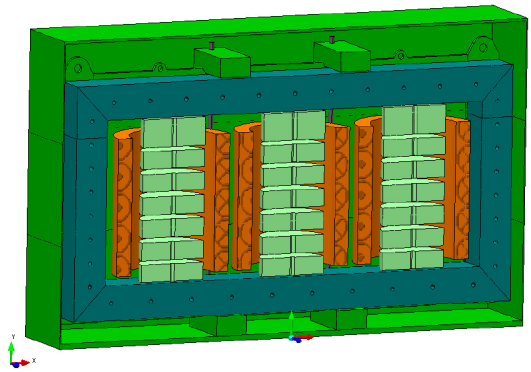


Figure 5: Proposed 2D FEM model

For the purpose of obtaining the results of the proposed 3D FEM procedure, the following steps are carried out on the transformer model:

- Excitation of the windings by the harmonic load current of each harmonic order,
- The corresponding flux density distribution in the active components are then calculated,
- The triggered currents generated by the distributed flux along with the material core, are determined, and
- The resistances of the windings are then presented to the geometries and the winding Eddy losses are then computed premised on the copper.

In the event the transformer under study is operating at a fundamental frequency, the load losses by employing the proposed FEM procedure are tabulated as shown in Table 3.

Table 3: Technical specification

Loss component	Value
HV load loss (Watts)	879,16
LV Load loss (Watts)	455,45
Total Load loss (Watts)	1334,59

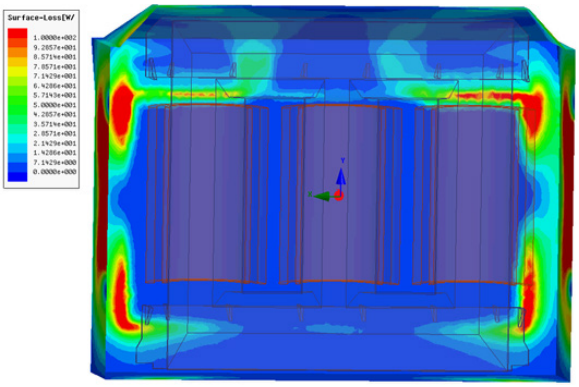


Figure 6: Magnetic flux leakage distribution.

In the magnetic circuit of the studied transformer, the distribution of the magnetic flux density is illustrated in Figure 6. At loading condition during service, the short circuit phenomena of the transformer may be presumed and the amplitude of the magnetisation current and leakage flux are minimal. As a result of the short circuit of the winding phases, significant density seeps through the windings. The latter presents a crucial factor in this condition. Under the supplied harmonic spectrum supplied in Figure 2, the total transformer losses are calculated as illustrated in Table 4 below.

The services load losses for each harmonic order is attained and tabulated as shown in Table 4 as 1410.53W.

Table 4: K-factor estimation

h	I_h(A)	P_H(W)	P_L(W)	P_{TOT}(W)
1	1,000	668,09	346,10	1014,19
3	0,350	167,02	86,53	253,55
5	0,170	58,28	29,68	87,95
7	0,120	20,05	11,49	31,54
9	0,092	6,68	2,87	9,56
11	0,071	5,51	2,83	8,34
13	0,051	1,82	0,74	2,56
15	0,043	0,61	0,49	1,10
17	0,040	0,81	0,39	1,20
19	0,039	0,37	0,18	0,55
Σ		**929,24**	**481,29**	**1410,53**

The winding Eddy losses at fundamental and under harmonic conditions are acquired by the difference between the total transformer losses and copper losses. The overall load current is at both conditions is presumed to be 1 p.u. The winding copper losses at a fundamental frequency and current as evaluated as:

$$3 \times [1.44^2(121.5) + 72^2(0.03)] = 1320.36W$$

The winding Eddy losses at fundamental frequency are evaluated as:

$$1410.53 - 1320.36 = 14.23W$$

The winding Eddy losses under harmonic conditions is evaluated as:

$$1410.53 - 1320.36 = 90.18W$$

By employing eq. (5), the harmonic loss factor is the ratio of the winding Eddy losses under harmonic conditions and at fundamental frequency.

$$F_{HL} = \frac{90.17}{14.23} = 6.337$$

Given that harmonic load current is equivalent to the rated current of the studied transformer (i.e. 1 p.u), the harmonic factor for the proposed FEM procedure is equal to the K-Factor as 6.337.

4.3. Method Comparison

The most significant outcomes of de-rating the transformer under study are presented in this sub-section. To critically evaluate the performance of the analytical method and the proposed FEM procedure, the harmonic loss factors under the supplied harmonic spectrum are compared.

Table 5: Harmonic loss factor comparison

Method	F_{HL}
AM	6,693
PFEM	6,337

Comparatively, the proposed 3D FEM procedure envisage a lower harmonic loss factor than the analytical method as evidenced in Table 5. The assumption by the analytical method on the direct proportion between the winding Eddy losses and the harmonic load current is glitch, which then yield hidebound results. The maximum allowable current of the studied transformer based in eq. (8) is evaluated as follows:

$$\mathrm{I}_{max} = \sqrt{\frac{1 + 0.232}{1 + 6.337 \times 0.232}} = 0{,}71p.u$$

The maximum allowable load current in Ampere (A) is evaluated as:

$$\mathrm{I}_{max} = 0.71 \times 79.2 = 55{,}93A$$

The equivalent power rating (in kVA) of the transformer using the proposed FEM procedure is:

$$Equivalent\ kVA = 500 \times 0.71 = 353{,}10kVA$$

The performance of the analytical procedure recommended by the C57.110-2018 standard and the proposed 3D FEM procedure are compare and tabulated as shown in Table 6. Comparatively, the results indicate that the two methods are close, regardless of the fact that the classical analytical procedure is hidebound.

Table 6: Transformer rating comparison

Method	kVA	I_{max}(A)
AM	347,26	55,01
PFEM	353,10	55,93

A further look into the results indicates that since the analytical method cannot consider parameters such as skin effect, proximity effect, and the magnetic flux leakage on the windings, it can underestimate the equivalent kVA and maximum allowable current. Based on practical perspective during service, this unit may experience issues such as stray gassing. The authors presented the work related to this phenomenon in the publications [14] and [15]. The embedding of FEM into the design philosophy proves to have enhanced results. The authors herein have also presented additional work on this application in [16] and [17].

5. Conclusion

In this work, an extension of previous work [10], the impact of harmonics and distortion upon a DPVP transformer predicated on the classical method has been investigated with the ambition to examine its de-rating capability. A 3D FEM procedure has been proposed to evaluate the equivalent power rating and maximum allowable load current of a 500kVA, three-phase, oil-immersed, DPVP transformer. The following conclusions can be drawn from the case scenario and the practical experience of the transformer under study:

- The most crucial impact of the harmonics load current upon transformers planned for the DPVP application is the service winding Eddy losses and the load losses.
- A surge in the transformer service losses under harmonic conditions carries off premature insulation materials and designed transformer service lifetime. The power rating (kVA) of a DPVP transformer must consequently be de-rated under harmonic conditions.
- The conservative assumption of the C57.110-2018 standard of the direct proportion between the winding Eddy losses and the harmonic load current is a glitch as it does not take into account parameters suchlike skin effect, proximity effect, and the magnetic flux leakage on the windings

The proposed 3D FEM procedure as a highly accurate procedure for evaluating the transformer service losses under harmonic conditions as it takes into account parameters that cannot be captured by the procedure proposed by the C57.110-2018 standard may be employed in the final design stage for de-rating.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] E. Cazacu, L. Petrescu and V. Ionita, "Derating of power distribution transformers serving nonlinear industrial loads," in 2017 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM) & 2017 Intl Aegean Conference on Electrical Machines and Power Electronics (ACEMP), Brasov, 90-95, 2017, doi: 10.1109/OPTIM.2017.7974953.

[2] D. Yildirim and E. F. Fuchs, "Measured transformer derating and comparison with harmonic loss factor (F/sub HL/) approach," in IEEE Transactions on Power Delivery, **15** (1), 186-191, 2000, doi: 10.1109/61.847249.

[3] E. Cherian and G. R. Bindu, "Minimizing Harmonics and Transformer Derating in Low Voltage Distribution Networks by DC Distribution," in 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR), Ernakulam, 2018, 1-6, doi: 10.1109/ICETIETR.2018.8529063.

[4] B. P. Das and Z. Radakovic, "Is Transformer kVA Derating Always Required Under Harmonics? A Manufacturer's Perspective," in IEEE Transactions on Power Delivery, **33** (6), 2693-2699, 2018, doi: 10.1109/TPWRD.2018.2815901.

[5] D. W. Egolf and A. J. Flechsig, "Harmonics-transformer derating," in Proceedings of Industrial and Commercial Power Systems Conference, Irvine, CA, USA, 79-84, 1994, doi: 10.1109/ICPS.1994.303557.

[6] E. Cazacu and L. Petrescu, "On-site derating of in-service power distribution transformers supplying nolinear loads", Revue Roumaine des Sciences Techniques - Serie Électrotechnique et Énergétique **59**(3), 259-268, 2014, doi: 18845326 .

[7] A.Y Shklyarskiy and A I Bardanov , "Test of the method for calculation of derating of workshop transformers on engineering plants", IOP Conf. Series: Materials Science and Engineering, **177**, 012054, 2017, doi:10.1088/1757-899X/177/1/012054.

[8] N. Xie, A. Luo, Fujun Ma, et al., "Harmonic interaction between large-scale photovoltaic power stations and grid", Proceedings of the Chinese Society of Electrical Engineering, **33**(34), 9-16, 2013, doi: 288752772.

[9] IEEE Recommended Practices and Requirements for Harmonic Control in Electrical Power Systems," in IEEE Std 519-1992 ,1-112, 1993, doi: 10.1109/IEEESTD.1993.114370.

[10] B. A. Thango, J. A. Jordaan and A. F. Nnachi, "Contemplation of Harmonic Currents Loading on Large-Scale Photovoltaic Transformers," in 2020 6th IEEE International Energy Conference (ENERGYCon), Gammarth, Tunis, Tunisia, 2020, 479-483, doi: 10.1109/ENERGYCon48941.2020.9236514.

[11] IEEE Recommended Practice for Establishing Liquid-Immersed and Dry-Type Power and Distribution Transformer Capability When Supplying Nonsinusoidal Load Currents," in IEEE Std C57.110™-2018 (Revision of

IEEE Std C57.110-2008), 1-68, 2018, doi: 10.1109/IEEESTD.2018.8511103.
[12] UL 1561, UL Standard for Safety Dry-Type General Purpose and Power Transformer, 4th Edition, 2011, doi: 00096942.
[13] J. Faiz, M.B.B Sharifian, S.A. Fakheri , "Research report on effect of non-linear loads upon distribution transformers and correction factor estimation for optimal operation of transformer-Part I", (in Persian), Azarbaijan Regional Electricity Company, Tabriz, Iran, 2001.
[14] D.B. Nyandeni, M. Phoshoko (Pr. Eng.), R. Murray, B.A. Thango, "Transformer Oil Degradation on PV Plants – A Case Study", 8th South Africa regional conference, 14-17, 2017, doi: 10.1109/MEI.2014.6749569.
[15] B. A. Thango, Jacobus A. Jordaan and Agha F. Nnachi, "Stray Gassing of Transformer Oil in Distributed Solar Photovoltaic (DSPV) Systems" 6th IEEE R8 International ENERGY Conference, 13-16, 2020, doi: 10.1109/ENERGYCon48941.2020.9236522.
[16] B.A Thango, J.A Jordaan, A.F Nnachi, D.B Nyandeni "Solar Power Plant Transformer Loss Calculation under Harmonic Currents using Field Element Method", 9th CIGRE Southern Africa Regional Conference, 1st – 4th October 2019, doi: 10.1109/MEI.2014.6749569.
[17] B. A. Thango, J.A. Jordaan, A. F. Nnachi, "Step-Up Transformers for PV Plants: Load Loss Estimation under Harmonic Conditions", 19th International Conference on Harmonics and Quality of Power (ICHQP), 2020, doi: 10.1109/ICHQP46026.2020.9177938.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 161-169 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Estimation of the Population Mean for Incomplete Data by using Information of Simple Linear Relationship Model in Data Set

Juthaphorn Sinsomboonthong[*,1], Saichon Sinsomboonthong[2]

[1]*Department of Statistics, Faculty of Science, Kasetsart University (KU), Bangkok 10900, Thailand*

[2]*Department of Statistics, School of Science, King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok 10520, Thailand*

ARTICLE INFO

Article history:
Received: 03 June, 2021
Accepted: 13 July, 2021
Online: 20 July, 2021

Keywords:
Bias
Estimator
Mean Square Error
Missing
Population Mean
Simple Linear Relationship

ABSTRACT

The objective of this research is to propose the estimator of the population mean for incomplete data by using information of simple linear relationship model in the data set. In addition, the factorization of the likelihood function is created to derive the maximum likelihood estimator for the population mean. The simulation study was conducted for 630 situations to compare the efficiency of the proposed estimator with the two population mean estimators, namely pairwise deletion and Anderson estimators. In this study, two criteria—bias and mean square error—of the performances for estimators are examined. It is found that all percentage levels of missing data, the mean square error of the proposed estimator tends to be lower than those of pairwise deletion and Anderson estimators for the large correlation levels between two variables in the data set whatever the sample sizes will be, especially for the large percentage level of missing data. However, for the small correlation between two variables in the data set, the three estimators tend to have the same performances in terms of both two criteria for all sample sizes and all percentage levels of missing data.

1. Introduction

Missing data are frequently found in many fields of research [1,2]. For example, some individuals may refuse to express any attitude for some sensitive questions in an opinion survey. In an experimental research, the experimental units may be leave or die before the experiment is completed. In longitudinal study, the monotone missing data pattern usually occurs. These missing data problems lead to increase an inaccuracy of the inference about the parameters in the population if the researchers ignore about the missing value in the data set. In estimation of the population mean for incomplete data set, imputation technique [3,4] is one of the familiar methods that researchers used it to replace the missing values with substituted values before estimate the population mean by using standard methods. However, the variance of estimator for this technique is underestimated and lead to the wrong inference about the population mean [5–7]. Available cases analysis is another technique that sample mean is used for estimation about the population mean and sometimes this is called pairwise deletion method. Moreover, this method will not suitable for the large amount of missing values because it will give the biased estimator and its standard error will increase [5, 8]. Ignoring missing values from the data set for inferential statistical analysis will affect the reliability of the conclusion about parameter in the population as the studied of [9–13]. Therefore, there are several researchers proposed about the estimators of the population mean for incomplete data set by considering only available cases analysis as follows: the maximum likelihood estimators of parameters for a bivariate normal distribution and case of some observations are missing for one variable were studied by [14]. That is, the factorization of likelihood function approach that proposed by [14] has been mostly used to derive the estimators of parameters for incomplete data set such as the studied of [15] and the research of [16]. Furthermore, these studies were found that the estimators derived by using likelihood function approach have a good performance, especially for a small sample size. Therefore, the proposed estimator of the population mean for incomplete dataset was derived based on a factorization of the likelihood function and using information of a simple linear relationship model in the data set. Moreover, a simulation study was conducted 630 situations to compare the efficiency of the proposed estimator with the two estimators, namely pairwise deletion estimator and Anderson

*Corresponding Author: Juthaphorn Sinsomboonthong, Faculty of Science, Kasetsart University, Thailand, E-mail: fscijps@ku.ac.th

estimator. In this study, the efficiency comparison criteria are bias and mean square error (MSE).

2. Materials and Methods

In this paper, the estimation methods of a population mean for incomplete data set are studied for efficiency comparison as follows:

2.1. Anderson Estimator

In 1957, the maximum likelihood estimators of the parameters of a bivariate normal distribution for incomplete data set with one variable was proposed by [14]. Suppose random variables Y_1 and Y_2 have the bivariate normal distribution with mean vector (μ_1, μ_2) and covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$. Suppose r observations of Y_1 and Y_2 are bivariate normally distributed with mean vector (μ_1, μ_2) and covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$. In addition, $n-r$ observations of Y_1 are normally distributed with mean μ_1 and variance σ_1^2. The data are shown in Figure 1.

$y_{11}, \cdots, y_{1r}, y_{1,r+1}, \cdots, y_{1n}$
$y_{21}, \cdots, y_{2r}$

Figure 1: Missing data pattern of the bivariate normal distribution

From data pattern in Figure 1, the likelihood function of vector parameter $\underline{\theta}^* = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12})$ can be written in the formula of equation (1).

$$L(\underline{\theta}^* | Y_{Obs}) = \prod_{j=1}^{n} f_{Y_1}(y_{1j} | \mu_1, \sigma_1^2) \prod_{j=1}^{r} f_{Y_2|Y_1}(y_{2j} | \beta_0 + \beta_1 y_{1j}, \sigma_{2|1}^2) \quad (1)$$

where $\beta_0 = \mu_2 - \beta_1 \mu_1$, $\beta_1 = \rho \frac{\sigma_2}{\sigma_1}$ and $\sigma_{2|1}^2 = (1-\rho^2)\sigma_2^2$.

The maximum likelihood estimators of $\mu_1, \sigma_1^2, \sigma_{2|1}^2, \beta_1$ and β_0 are as follows:

$$\hat{\mu}_1 = \bar{y}_1 = \frac{1}{n}\sum_{j=1}^{n} y_{1j}, \quad \hat{\sigma}_1^2 = \frac{1}{n}\sum_{j=1}^{n}(y_{1j} - \bar{y}_1)^2, \quad \hat{\sigma}_{2|1}^2 = s_2'^2 - \frac{s_{12}'^2}{s_1'^2},$$

$$\hat{\beta}_1 = \frac{\sum_{j=1}^{r}(y_{1j} - \bar{y}_1')(y_{2j} - \bar{y}_2')}{\sum_{j=1}^{r}(y_{1j} - \bar{y}_1')^2} \quad \text{and} \quad \hat{\beta}_0 = \bar{y}_2' - \hat{\beta}_1 \bar{y}_1'$$

where, $s_1'^2 = \frac{1}{r}\sum_{j=1}^{r}(y_{1j} - \bar{y}_1')^2$, $\bar{y}_2' = \frac{1}{r}\sum_{j=1}^{r} y_{2j}$, $\bar{y}_1' = \frac{1}{r}\sum_{j=1}^{r} y_{1j}$

$s_2'^2 = \frac{1}{r}\sum_{j=1}^{r}(y_{2j} - \bar{y}_2')^2$ and $s_{12}' = \frac{1}{r}\sum_{j=1}^{r}(y_{1j} - \bar{y}_1')(y_{2j} - \bar{y}_2')$.

Moreover, the maximum likelihood estimators of μ_2 and σ_2^2 are given by $\hat{\mu}_2 = \bar{y}_2' - \hat{\beta}_1(\bar{y}_1' - \bar{y}_1)$ and $\hat{\sigma}_2^2 = \hat{\sigma}_{2|1}^2 + \hat{\beta}_1^2 \hat{\sigma}_1^2 = \hat{\beta}_1 \frac{\hat{\sigma}_1}{\hat{\sigma}_2}$, respectively.

2.2. Pairwise Deletion Estimator

In this study, pairwise deletion estimator is the estimation of the population mean for incomplete data set based on complete data or available-cases analysis [5], even if the values for the same individual on other variables are missing. Suppose three variables Y_1, Y_2 and Y_3 are trivariate normally distributed in the population and n observations of Y_1 are completely observed for all individuals, but Y_2 and Y_3 are not completely observed for all individuals or they have missing data occurrence. That is, r observations of Y_2 are observed whereas $n-r$ observations of Y_3 are observed. Available cases analysis for the population means μ_1, μ_2 and μ_3 can be written in the forms of equation (2).

$$\hat{\mu}_1 = \frac{1}{n}\sum_{j=1}^{n} y_{1j}, \quad \hat{\mu}_2 = \frac{1}{r}\sum_{j=1}^{r} y_{2j} \quad \text{and} \quad \hat{\mu}_3 = \frac{1}{n-r}\sum_{j=r+1}^{n} y_{3j} \quad (2)$$

Under MCAR [5] of the missing data mechanism, pairwise deletion method will yield consistent and unbiased estimators in a large sample size [5].

2.3. The Proposed Estimator of the Population Mean for Incomplete Data Set

In this section, the estimator of the population mean for incomplete data set is proposed. This proposed estimator is derived using the factorization of the likelihood function [5,14] and a procedure of finding the usual maximum likelihood estimator is applied. Suppose dependent variable Y_1 is assumed to have the linear relationship with independent variable X_1 and its relationship model is given by equation (3).

$$y_{1j} = \delta_0 + \delta_1 x_{1j} + \varepsilon_{1j}, \quad j = 1, 2, ..., n \quad (3)$$

where δ_0 and are random ε_{1j} and , are unknown parameters δ_1 errors that have the normal distribution with mean 0 and variance σ_1^2. Then the mean and variance of Y_1 can be written as $E(Y_1) = \delta_0 + \delta_1 X_1 = \mu_1$ and $V(Y_1) = \sigma_1^2$, respectively. Further, ε_{1j} can be written in the form of equation (4).

$$\varepsilon_{1j} = y_{1j} - \delta_0 - \delta_1 x_{1j}, \quad j = 1, 2, ..., n \quad (4)$$

Let Y_2
μ_2 and variance σ_2^2. In addition, r observations of Y_1 and Y_2

$\underline{\mu} = (\delta_0 + \delta_1 X_1 \ , \ \mu_2)$ and covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$.

The $n-r$ observations of Y_1 are normally distributed with mean $\delta_0 + \delta_1 X_1$ and variance σ_1^2. The study data pattern is shown in Figure 2.

Observations	X_1	Y_1	Y_2
1	x_{11}	y_{11}	y_{21}
2	x_{12}	y_{12}	y_{22}
⋮	⋮	⋮	⋮
r	x_{1r}	y_{1r}	y_{2r}
$r+1$	$x_{1,r+1}$	$y_{1,r+1}$	
⋮	⋮	⋮	
n	x_{1n}	y_{1n}	

Figure 2: Missing data pattern of the proposed study

Let E_1 be a random variable that have the relationship of Y_1 and X_1 in the form of $E_1 = Y_1 - \delta_0 - \delta_1 X_1$. Then two random variables E_1 and Y_2 are bivariate normally distributed with mean vector $\underline{\mu} = (0, \ \mu_2)$ and covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$. Additionally, the missing data pattern of E_1 and Y_2 are shown in Figure 3.

Observations	E_1	Y_2
1	ε_{11}	y_{21}
2	ε_{12}	y_{22}
⋮	⋮	⋮
r	ε_{1r}	y_{2r}
$r+1$	$\varepsilon_{1,r+1}$	
⋮	⋮	
n	ε_{1n}	

Figure 3: Random error and missing data pattern of Y_2

Lemma 1 Let $E_1 = Y_1 - \delta_0 - \delta_1 X_1$, Y_1 and Y_2 be the random variables where δ_0, δ_1 are unknown parameters and X_1 be independent variable. Suppose E_1 and Y_2 are bivariate normally distributed with mean vector $\underline{\mu} = (0, \ \mu_2)$ and covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$. Then, $Y_2 \,|E_1 = \varepsilon_1$ is normally distributed with mean $\mu_{2|1} = \mu_2 + \tau_{12}\varepsilon_1$ and variance $\sigma_{2|1}^2 = (1-\rho^2)\sigma_2^2$ where $\varepsilon_1 = y_1 - \delta_0 - \delta_1 x_1$,. $\underline{\theta}_{2|1} = (\delta_0, \delta_1, \sigma_{2|1}^2, \tau_{12})$ and $\tau_{12} = \dfrac{\rho\sigma_2}{\sigma_1}$

Proof Let $E_1 = Y_1 - \delta_0 - \delta_1 X_1$ and Y_2 be bivariate normally distributed with mean vector $\underline{\mu} = (0, \ \mu_2)$ and covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$. Then, the joint probability density function of E_1 and Y_2 is given by equation (5).

$$f_{12}(\varepsilon_1, y_2; \underline{\theta}_{12}) = \frac{1}{2\pi\sqrt{(1-\rho^2)\sigma_1^2\sigma_2^2}} e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{\varepsilon_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{\varepsilon_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right) + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right\}} \quad (5)$$

where $-\infty < \varepsilon_1 < \infty$, $-\infty < y_2 < \infty$ and $\underline{\theta}_{2|1} = (\delta_0, \delta_1, \sigma_{2|1}^2, \tau_{12})$. Moreover, the probability density function of E_1 is given by equation (6).

$$f_1(\varepsilon_1; \underline{\theta}_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2}\left(\frac{\varepsilon_1}{\sigma_1}\right)^2} \quad (6)$$

where $-\infty < \varepsilon_1 < \infty$ and $\underline{\theta}_1 = (\delta_0, \delta_1, \sigma_1^2)$.

Hence, a conditional probability density function of Y_2 given $E_1 = \varepsilon_1$ can be written as follows:

$$\begin{aligned} f_{2|1}(y_2 \mid \varepsilon_1; \underline{\theta}_{2|1}) &= \frac{f_{12}(\varepsilon_1, y_2; \underline{\theta}_{12})}{f_1(\varepsilon_1; \underline{\theta}_1)} \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_2^2}} e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{y_2-\mu_2}{\sigma_2}\right) - \rho\left(\frac{\varepsilon_1}{\sigma_1}\right)\right\}^2} \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_2^2}} e^{-\frac{1}{2(1-\rho^2)\sigma_2^2}\left\{ y_2 - \mu_2 - \frac{\rho\sigma_2}{\sigma_1}\varepsilon_1 \right\}^2} \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_2^2}} e^{-\frac{1}{2(1-\rho^2)\sigma_2^2}\{ y_2 - \mu_2 - \tau_{12}\varepsilon_1 \}^2} \ ; \ \tau_{12} = \frac{\rho\sigma_2}{\sigma_1} \\ &= \frac{1}{\sqrt{2\pi\sigma_{2|1}^2}} e^{-\frac{1}{2\sigma_{2|1}^2}\{ y_2 - \mu_{2|1} \}^2} \end{aligned} \quad (7)$$

where $\mu_{2|1} = \mu_2 + \tau_{12}\varepsilon_1$ and $\sigma_{2|1}^2 = (1-\rho^2)\sigma_2^2$.

From Equation (7), this is the probability density function of a normal distribution with mean $\mu_{2|1} = \mu_2 + \tau_{12}\varepsilon_1$ and variance $\sigma_{2|1}^2 = (1-\rho^2)\sigma_2^2$. Therefore, a random variable $Y_2 \,|E_1 = \varepsilon_1$ is normally distributed with mean $\mu_{2|1} = \mu_2 + \tau_{12}\varepsilon_1$ and variance $\sigma_{2|1}^2 = (1-\rho^2)\sigma_2^2$ where $\varepsilon_1 = y_1 - \delta_0 - \delta_1 x_1$ and $\tau_{12} = \dfrac{\rho\sigma_2}{\sigma_1}$.

Lemma 2 For $j = 1, 2, ..., r$, the two random variables E_{1j} and Y_{2j} are assumed to have the bivariate normal distribution with a mean vector $\underline{\mu} = (0,\ \mu_2)$ and covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$. For $j = r+1,\ r+2,\ ...,\ n,$ the random variable E_{1j} is assumed to have a normal distribution with a mean 0 and variance σ_1^2 where $E_{1j} = Y_{1j} - \delta_0 - \delta_1 X_{1j}$; δ_0 and δ_1 are unknown parameters and X_{1j} be independent variable. Let $\underline{W} = [E_{11}\ E_{12} ... E_{1n}\ Y_{21}\ Y_{22} ... Y_{2r}]'$ be a random vector. Then, the likelihood function of parameter vector $\underline{\theta} = (\delta_0, \delta_1, \sigma_1^2, \sigma_{2|1}^2, \tau_{12})$ is denoted by equation (8).

$$L(\underline{\theta} \mid \underline{w}) = \left(\left(2\pi\sigma_1^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}\varepsilon_{1j}^2} \right) \left(\left(2\pi\sigma_{2|1}^2\right)^{-\frac{r}{2}} e^{-\frac{1}{2\sigma_{2|1}^2}\sum_{j=1}^{r}\left(y_{2j} - \mu_{2|1}\right)^2} \right) \quad (8)$$

where $\sigma_{2|1}^2 = (1-\rho^2)\sigma_2^2$ and $\tau_{12} = \dfrac{\rho\sigma_2}{\sigma_1}$, $\varepsilon_{1j} = y_{1j} - \delta_0 - \delta_1 x_{1j}$

Proof For $j = 1, 2, ..., r$, the two random variables E_{1j} and Y_{2j} are assumed to have a bivariate normal distribution with mean vector $\underline{\mu} = (0,\ \mu_2)$ and covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$. For $j = r+1,\ r+2,\ ...,\ n,$ the random variable E_{1j} is assumed to have a normal distribution with a mean 0 and variance σ_1^2.

Let $\underline{w} = [\varepsilon_{11}\ \varepsilon_{12}\ ...\ \varepsilon_{1n}\ \ y_{21}\ \ y_{22}\ ...\ y_{2r}]'$ be a vector of value for the random vector $\underline{W} = [E_{11}\ E_{12} ... E_{1n}\ Y_{21}\ Y_{22} ... Y_{2r}]'$. Then, the likelihood function of $\underline{\theta} = (\delta_0, \delta_1, \sigma_1^2, \sigma_{2|1}^2, \tau_{12})$ can be written as follows:

$$\begin{aligned} L(\underline{\theta} \mid \underline{w}) &= \prod_{j=1}^{r} f_{12}(\varepsilon_{1j}, y_{2j}; \underline{\theta}_{12}) \prod_{j=r+1}^{n} f_1(\varepsilon_{1j}; \underline{\theta}_1) \\ &= \left(\prod_{j=1}^{r} f_1(\varepsilon_{1j}; \underline{\theta}_1) \times f_{2|1}(y_{2j} \mid \varepsilon_{1j}; \underline{\theta}_{2|1}) \right) \left(\prod_{j=r+1}^{n} f_1(\varepsilon_{1j}; \underline{\theta}_1) \right) \\ &= \prod_{j=1}^{n} f_1(\varepsilon_{1j}; \underline{\theta}_1) \prod_{j=1}^{r} f_{2|1}(y_{2j} \mid \varepsilon_{1j}; \underline{\theta}_{2|1}) \end{aligned} \quad (9)$$

From Lemma 1, the likelihood function $L(\underline{\theta} \mid \underline{w})$ in equation (9) can be written as

$$\begin{aligned} &L(\underline{\theta} \mid \underline{w}) \\ &= \left(\prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2}\left(\frac{\varepsilon_{1j}}{\sigma_1}\right)^2} \right) \left(\prod_{j=1}^{r} \frac{1}{\sqrt{2\pi\sigma_{2|1}^2}} e^{-\frac{1}{2\sigma_{2|1}^2}\left\{ y_{2j} - \mu_{2|1} \right\}^2} \right) \\ &= \left(\left(2\pi\sigma_1^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma_1^2}\sum_{j=1}^{n}\varepsilon_{1j}^2} \right) \left(\left(2\pi\sigma_{2|1}^2\right)^{-\frac{r}{2}} e^{-\frac{1}{2\sigma_{2|1}^2}\sum_{j=1}^{r}\left(y_{2j} - \mu_{2|1}\right)^2} \right) \end{aligned}$$

Theorem 1 For $j = 1, 2, ..., r$, the two random variables E_{1j} and Y_{2j} are assumed to have a bivariate normal distribution with mean vector $\underline{\mu} = (0,\ \mu_2)$ and covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$. For $j = r+1,\ r+2,\ ...,\ n,$ the random variable E_{1j} is assumed to have a normal distribution with mean 0 and variance σ_1^2 where $E_{1j} = Y_{1j} - \delta_0 - \delta_1 X_{1j}$; δ_0 and δ_1 are unknown parameters and X_{1j} be independent variable. Let $\underline{W} = [E_{11}\ E_{12} ... E_{1n}\ Y_{21}\ Y_{22} ... Y_{2r}]'$ be a random vector. Then, the factorization maximum likelihood estimator of μ_2 is given in equation (10).

$$\hat{\mu}_{2_{\text{Proposed}}} = \overline{y}_2' - \hat{\tau}_{12}\overline{e}_1' \quad (10)$$

where $\hat{\delta}_1 = \dfrac{\sum_{j=1}^{n} x_{1j}y_{1j} - n\overline{x}_1\overline{y}_1}{\sum_{j=1}^{n} x_{1j}^2 - n(\overline{x}_1)^2}$,, $\overline{y}_1 = \dfrac{1}{n}\sum_{j=1}^{n} y_{1j}$, $\overline{x}_1 = \dfrac{1}{n}\sum_{j=1}^{n} x_{1j}$

$e_{1j} = y_{1j} - \hat{\delta}_0 - \hat{\delta}_1 x_{1j}$, $\hat{\delta}_0 = \overline{y}_1 - \hat{\delta}_1\overline{x}_1$ for, $j = 1, 2, ..., r$

$\hat{\tau}_{12} = \dfrac{\sum_{j=1}^{r} e_{1j}y_{2j} - r\overline{e}_1'\overline{y}_2'}{\sum_{j=1}^{r} e_{1j}^2 - r(\overline{e}_1')^2}$, $\overline{y}_2' = \dfrac{1}{r}\sum_{j=1}^{r} y_{2j}$ and $\overline{e}_1' = \dfrac{1}{r}\sum_{j=1}^{r} e_{1j}$.

Proof Let $\underline{W} = [E_{11}\ E_{12} ... E_{1n}\ Y_{21}\ Y_{22} ... Y_{2r}]'$ be a random vector. From Lemma 2, we known that the likelihood function of $\underline{\theta} = (\delta_0, \delta_1, \sigma_1^2, \sigma_{2|1}^2, \tau_{12})$ is denoted by equation (8). Then, the log-likelihood function can be written in the form of equation (11).

$$\begin{aligned} \ln L(\underline{\theta} \mid \underline{w}) &= -\frac{n}{2}\ln\left(2\pi\sigma_1^2\right) - \frac{1}{2\sigma_1^2}\sum_{j=1}^{n}\varepsilon_{1j}^2 - \frac{r}{2}\ln\left(2\pi\sigma_{2|1}^2\right) \\ &\quad - \frac{1}{2\sigma_{2|1}^2}\sum_{j=1}^{r}\left(y_{2j} - \mu_{2|1}\right)^2 \end{aligned} \quad (11)$$

From Lemma 1, the random variable $Y_2 \mid E_1 = \varepsilon_1$ is normally distributed with mean $\mu_{2|1} = \mu_2 + \tau_{12}\varepsilon_1$ and variance $\sigma_{2|1}^2 = (1-\rho^2)\sigma_2^2$ where $\varepsilon_1 = y_1 - \delta_0 - \delta_1 x_1$ and $\tau_{12} = \dfrac{\rho\sigma_2}{\sigma_1}$.

Then, the log-likelihood function as shown in equation (11) need to maximize and achieve the maximum likelihood estimators of μ_2, δ_0, δ_1 and are as follows: τ_{12}

$$\frac{\partial}{\partial\delta_0}\ln L(\underline{\theta}\,|\,\underline{w}) = \frac{\partial}{\partial\delta_0}\left[-\frac{1}{2\sigma_1^2}\sum_{j=1}^{n}\varepsilon_{1j}^2\right] = 0$$

$$= \frac{\partial}{\partial\delta_0}\left[-\frac{1}{2\sigma_1^2}\sum_{j=1}^{n}\left(y_{1j}-\delta_0-\delta_1 x_{1j}\right)^2\right] = 0$$

$$= \sum_{j=1}^{n} y_{1j} - n\delta_0 - \delta_1\sum_{j=1}^{n} x_{1j} = 0 \qquad (12)$$

$$\frac{\partial}{\partial\delta_1}\ln L(\underline{\theta}\,|\,\underline{w}) = \frac{\partial}{\partial\delta_1}\left[-\frac{1}{2\sigma_1^2}\sum_{j=1}^{n}\varepsilon_{1j}^2\right] = 0$$

$$= \frac{\partial}{\partial\delta_1}\left[-\frac{1}{2\sigma_1^2}\sum_{j=1}^{n}\left(y_{1j}-\delta_0-\delta_1 x_{1j}\right)^2\right] = 0$$

$$= \sum_{j=1}^{n} x_{1j}y_{1j} - \delta_0\sum_{j=1}^{n} x_{1j} - \delta_1\sum_{j=1}^{n} x_{1j}^2 = 0 \qquad (13)$$

Equation (12) is multiplied by $\sum_{j=1}^{n} x_{1j}$, then it will give the form in equation (14).

$$\sum_{j=1}^{n} x_{1j}\sum_{j=1}^{n} y_{1j} - n\delta_0\sum_{j=1}^{n} x_{1j} - \delta_1\left(\sum_{j=1}^{n} x_{1j}\right)^2 = 0 \qquad (14)$$

Equation (13) is multiplied by *n*, then it will give the form in equation (15).

$$n\sum_{j=1}^{n} x_{1j}y_{1j} - n\delta_0\sum_{j=1}^{n} x_{1j} - n\delta_1\sum_{j=1}^{n} x_{1j}^2 = 0 \qquad (15)$$

Subtraction equation (14) from equation (15), then it will give the form in equation (16).

$$n\sum_{j=1}^{n} x_{1j}y_{1j} - n\delta_1\sum_{j=1}^{n} x_{1j}^2 - \sum_{j=1}^{n} x_{1j}\sum_{j=1}^{n} y_{1j} + \delta_1\left(\sum_{j=1}^{n} x_{1j}\right)^2 = 0 \qquad (16)$$

That is, $= 0\ n\sum_{j=1}^{n} x_{1j}y_{1j} - \sum_{j=1}^{n} x_{1j}\sum_{j=1}^{n} y_{1j} - \left[n\sum_{j=1}^{n} x_{1j}^2 - \left(\sum_{j=1}^{n} x_{1j}\right)^2\right]\delta_1$

Additionally, the value of δ_1 that maximize the log-likelihood function is denoted by

$$\delta_1 = \frac{n\sum_{j=1}^{n} x_{1j}y_{1j} - \sum_{j=1}^{n} x_{1j}\sum_{j=1}^{n} y_{1j}}{n\sum_{j=1}^{n} x_{1j}^2 - \left(\sum_{j=1}^{n} x_{1j}\right)^2} \quad \text{or} \quad \delta_1 = \frac{\sum_{j=1}^{n} x_{1j}y_{1j} - n\bar{x}_1\bar{y}_1}{\sum_{j=1}^{n} x_{1j}^2 - n(\bar{x}_1)^2}$$

Therefore, the maximum likelihood estimator of δ_1 is given by

$$\hat{\delta}_1 = \frac{\sum_{j=1}^{n} x_{1j}y_{1j} - n\bar{x}_1\bar{y}_1}{\sum_{j=1}^{n} x_{1j}^2 - n(\bar{x}_1)^2} \quad \text{for } \bar{x}_1 = \frac{1}{n}\sum_{j=1}^{n} x_{1j} \text{ and } \bar{y}_1 = \frac{1}{n}\sum_{j=1}^{n} y_{1j}.$$

From equation (12), the form of this equation can be written as $n\delta_0 = \sum_{j=1}^{n} y_{1j} - \delta_1\sum_{j=1}^{n} x_{1j}$ or Then, the maximum . $\delta_0 = \bar{y}_1 - \delta_1\bar{x}_1$

likelihood estimator of δ_0 is given by $\hat{\delta}_0 = \bar{y}_1 - \hat{\delta}_1\bar{x}_1$.

From Lemma 1, we know that $\mu_{2|1} = \mu_2 + \tau_{12}\varepsilon_1$ then the maximum likelihood estimator of parameter τ_{12} can be derived as follows:

$$\frac{\partial}{\partial\tau_{12}}\ln L(\underline{\theta}\,|\,\underline{w}) = \frac{\partial}{\partial\tau_{12}}\left[-\frac{1}{2\sigma_{2|1}^2}\sum_{j=1}^{r}\left(y_{2j}-\mu_{2|1}\right)^2\right] = 0$$

$$= \frac{\partial}{\partial\tau_{12}}\left[-\frac{1}{2\sigma_{2|1}^2}\sum_{j=1}^{r}\left(y_{2j}-\mu_2-\tau_{12}\varepsilon_{1j}\right)^2\right] = 0$$

$$= \sum_{j=1}^{r}\varepsilon_{1j}y_{2j} - \mu_2\sum_{j=1}^{r}\varepsilon_{1j} - \tau_{12}\sum_{j=1}^{r}\varepsilon_{1j}^2 = 0 \qquad (17)$$

$$\frac{\partial}{\partial\mu_2}\ln L(\underline{\theta}\,|\,\underline{w}) = \frac{\partial}{\partial\mu_2}\left[-\frac{1}{2\sigma_{2|1}^2}\sum_{j=1}^{r}\left(y_{2j}-\mu_{2|1}\right)^2\right] = 0$$

$$= \frac{\partial}{\partial\mu_2}\left[-\frac{1}{2\sigma_{2|1}^2}\sum_{j=1}^{r}\left(y_{2j}-\mu_2-\tau_{12}\varepsilon_{1j}\right)^2\right] = 0$$

$$= \sum_{j=1}^{r} y_{2j} - r\mu_2 - \tau_{12}\sum_{j=1}^{r}\varepsilon_{1j} = 0 \qquad (18)$$

Equation (18) is multiplied by $\sum_{j=1}^{r}\varepsilon_{1j}$, then it will give the form in equation (19).

$$\sum_{j=1}^{r}\varepsilon_{1j}\sum_{j=1}^{r} y_{2j} - r\mu_2\sum_{j=1}^{r}\varepsilon_{1j} - \tau_{12}\left(\sum_{j=1}^{r}\varepsilon_{1j}\right)^2 = 0 \qquad (19)$$

Equation (17) is multiplied by *r*, then it will give the form in equation (20).

$$r\sum_{j=1}^{r}\varepsilon_{1j}y_{2j} - r\mu_2\sum_{j=1}^{r}\varepsilon_{1j} - r\tau_{12}\sum_{j=1}^{r}\varepsilon_{1j}^2 = 0 \qquad (20)$$

Subtraction equation (19) from equation (20), then it will give the form in equation (21).

$$r\sum_{j=1}^{r}\varepsilon_{1j}y_{2j} - r\tau_{12}\sum_{j=1}^{r}\varepsilon_{1j}^2 - \sum_{j=1}^{r}\varepsilon_{1j}\sum_{j=1}^{r} y_{2j} + \tau_{12}\left(\sum_{j=1}^{r}\varepsilon_{1j}\right)^2 = 0 \qquad (21)$$

Furthermore, the value of τ_{12} that maximize the log-likelihood function is denoted by

$$\tau_{12} = \frac{r\sum_{j=1}^{r}\varepsilon_{1j}y_{2j} - \sum_{j=1}^{r}\varepsilon_{1j}\sum_{j=1}^{r} y_{2j}}{r\sum_{j=1}^{r}\varepsilon_{1j}^2 - \left(\sum_{j=1}^{r}\varepsilon_{1j}\right)^2} \quad \text{or} \quad \tau_{12} = \frac{\sum_{j=1}^{r}\varepsilon_{1j}y_{2j} - r\bar{\varepsilon}_1'\bar{y}_2'}{\sum_{j=1}^{r}\varepsilon_{1j}^2 - r(\bar{\varepsilon}_1')^2}.$$

Therefore, the maximum likelihood estimator of τ_{12} is given by

$$\hat{\tau}_{12} = \frac{\sum_{j=1}^{r} e_{1j}y_{2j} - r\bar{e}_1'\bar{y}_2'}{\sum_{j=1}^{r} e_{1j}^2 - r(\bar{e}_1')^2} \qquad \bar{y}_2' = \frac{1}{r}\sum_{j=1}^{r} y_{2j} \text{ and } \bar{e}_1' = \frac{1}{r}\sum_{j=1}^{r} e_{1j}$$

$$e_{1j} = y_{1j} - \hat{\delta}_0 - \hat{\delta}_1 x_{1j} \qquad j = 1, 2, ..., r\ ;$$

, thTherefore $\mu_2 = \bar{y}_2' - \tau_{12}\bar{\varepsilon}_1'$. or $r\mu_2 = \sum_{j=1}^{r} y_{2j} - \tau_{12}\sum_{j=1}^{r}\varepsilon_{1j}$ maximum likelihood estimator of parameter μ_2 is given by $\hat{\mu}_{2_{\text{Proposed}}} = \bar{y}_2' - \hat{\tau}_{12}\bar{e}_1'$.

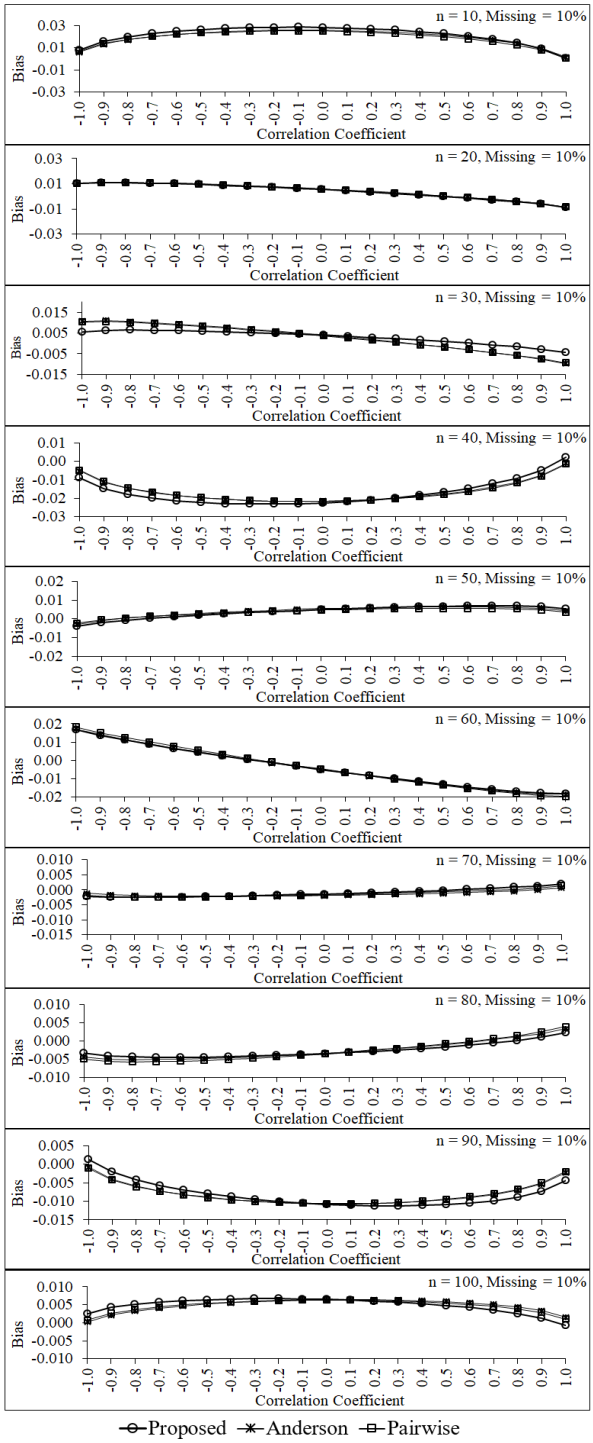


Figure 4: Biases of the three estimators for percentage of missing data equals 10 of each sample size

3. Results of a Simulation Study

The efficiency investigation of the proposed estimator and comparison of its efficiency with the two estimators—Anderson and pairwise deletion estimators—are studied via the simulation data. Moreover, these data are generated 630 situations and repeated 50,000 times for each situation. In this section, the criteria in terms of bias and mean square error are used for efficiency comparison. The population data of random variables Y_1 and Y_2 are generated in the form of bivariate normal distribution with mean vector $\underline{\mu} = (\delta_0 - \delta_1 X_1,\ \mu_2)$ and covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$.

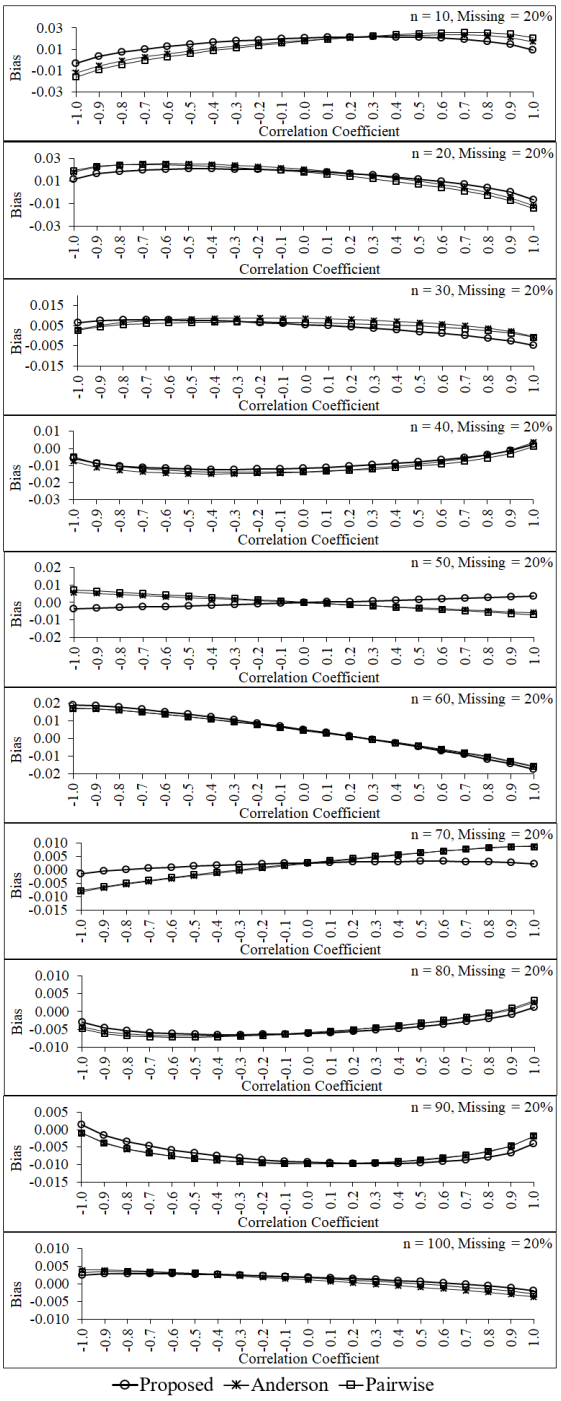


Figure 5: Biases of the three estimators for percentage of missing data equals 20 of each sample size

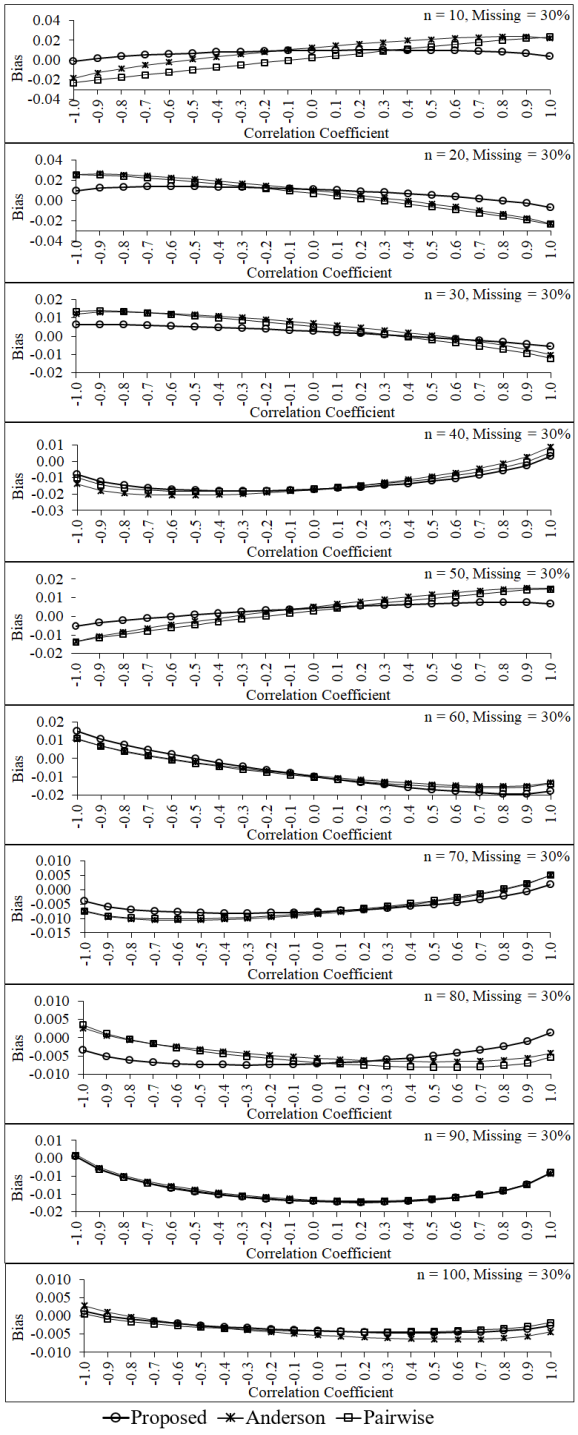


Figure 6: Biases of the three estimators for percentage of missing data equals 30 of each sample size

In this study, the values of parameters are defined as follows: $\delta_0 = 2$, $\delta_1 = 3$, $\mu_2 = 5$, $\sigma_2^2 = 9$ and the correlations between Y_1 and Y_2 are given by $\rho = -1.0, -0.9, ..., 0, ..., 0.9, 1.0$. Then, the samples of size n = 10, 20, 30, ..., 100 are randomly taken from these populations. Missing data mechanism in the form of MCAR [5] for three levels—10%, 20% and 30%—are constructed from each sample. The simulation results are shown in Figure 4 to Figure 9. Figure 4 to Figure 6 show that when percentages of missing data equal 10, 20 and 30 of each sample size, bias of the proposed estimator tends to be no difference from those of pairwise deletion and Anderson estimators for almost all sample sizes and all levels of the correlation between two variables in the data set. Moreover, some situations (e.g., n = 20, 30 and percentage of missing data in the data set equals 30) and negative high correlation between two variables, its bias tends to be smaller than the bias of pairwise deletion and Anderson estimators.

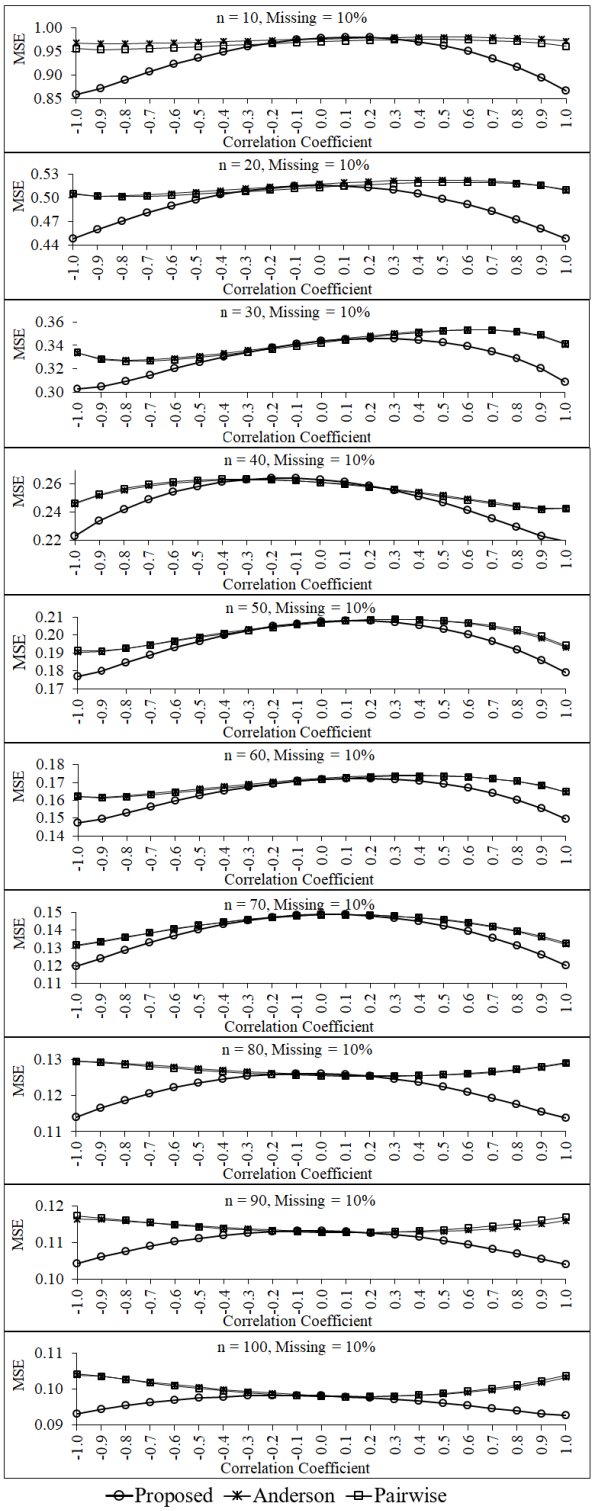


Figure 7: Mean square errors of the three estimators for percentage of missing data equals 10 of each sample size

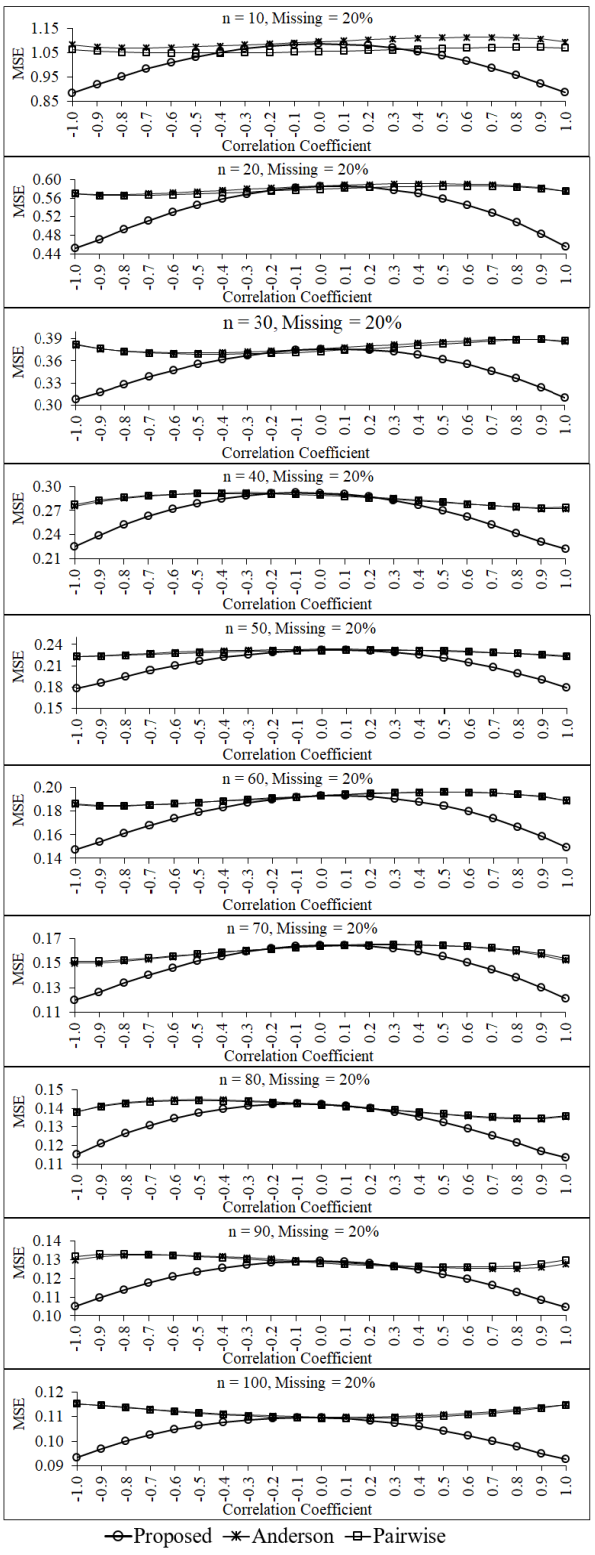


Figure 8: Mean square errors of the three estimators for percentage of missing data equals 20 of each sample size

When considering the performance of the proposed estimator in term of mean square error in Figure 7, it is found that the mean square error of the proposed estimator tends to be lower than those of pairwise deletion and Anderson estimators for the large correlation levels between two variables in the data set and all sample sizes when the data have 10 % of missing data.

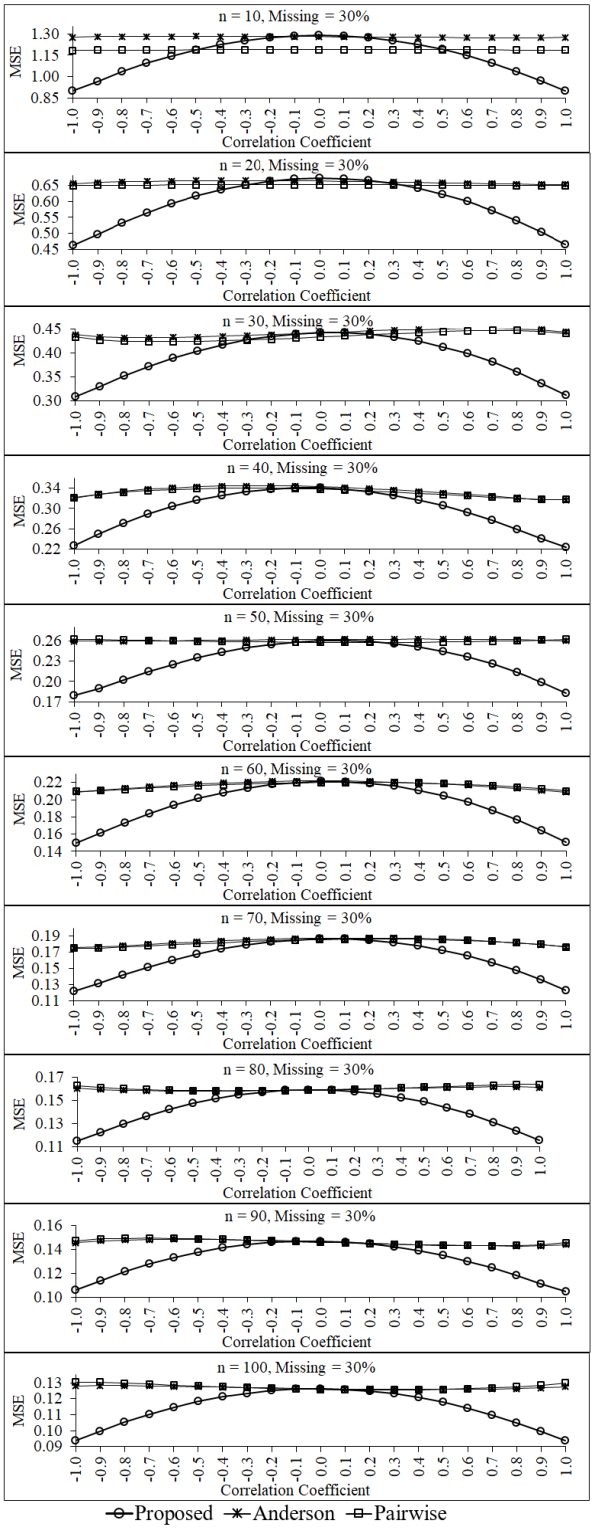


Figure 9: Mean square errors of the three estimators for percentage of missing data equals 30 of each sample size

For higher percentages of missing data of each sample sizes as show in Figure 8 and Figure 9, the performance of the proposed estimator in term of mean square error are similar to the case of the small percentages of missing data as mention above. Additionally, the mean square error of the proposed estimator tends to be obviously lower than those of pairwise deletion and Anderson estimators for the large correlation levels between two variables in

the data set whatever the sample sizes will be. However, for the small correlation levels between two variables in the data set, the three estimators tend to have the same performances in terms of both two criteria—bias and mean square error—for all sample sizes and all percentage levels of missing data. This simulation study is found that the mean square errors of three estimators tend to be decrease when the sample size increases for all levels of the correlations between two variables in the data set and all levels of the percentages of missing data. In addition, the mean square error of the proposed estimator tends to be lower than those of the two estimators—pairwise deletion and Anderson estimators—for the small sample sizes (e.g., n = 10, 20, 30) and high correlations (e.g., ρ = -0.1, -0.9, -0.8, 0.8, 0.9, 1.0) between two variables in the data set, especially the percentage of missing data is equal to 30. However, the mean square errors of three estimators tend to have a similar performances for the low correlations between two variables in the data set and all levels of the percentages of missing data.

4. Discussion

In this study, the simulation results show that pairwise deletion estimator tends to be a biased estimator for the small sample sizes as mention by [5,9]. Moreover, the maximum likelihood estimator of the population average for incomplete data set is derived by using factorization of the likelihood function approach [14] tends to have a good performance for the large correlation levels between two variables in the data set and small sample sizes. This conforms to the studies of [14,16]. In addition, the maximum likelihood estimation of the population mean for incomplete data set tends to have a good efficiency for small sample sizes as the study of [7]. This discovery of the proposed estimator will benefit for some applications in the real life data, especially nowadays it is the era of big data analysis which has the large number of variables in data set. Therefore, we should find the relationships of some attributes in data set before estimating the average of the interested variables for incomplete data analysis. Further, this proposed estimator will lead to correct estimate as possible.

5. Conclusion

The proposed estimator of the population mean for incomplete dataset was derived by using the linear relationship between some variables in the data set and the factorization of likelihood function [14] was created to derive the proposed maximum likelihood estimator. Additionally, the investigation of this proposed estimator was studied via the simulation data for 630 situations to compare the efficiency in terms of bias and mean square error with two estimators, namely pairwise deletion and Anderson estimators. It is found that the efficiency of the proposed estimator tends to be better than those of two above mention estimators, especially for case of the high percentages of missing data and the strong linear correlation between two variables (e.g., the degree of ρ close to -1 or 1) whatever the sample size will be. However, for the small correlation between two variables (e.g., the degree of ρ close to zero), the three estimators tend to have the similar efficiencies for all sample sizes and all percentage levels of missing data.

Acknowledgment

The authors would like to express our special thanks of gratitude to head of Kasetsart University Research and Development Institute (KURDI) for financial support of this research.

References

[1] S. Gaucher, O. Klopp, G. Robin, "Outlier detection in networks with missing links," Computational Statistics & Data Analysis, **164**, 107308, 2021, doi:10.1016/j.csda.2021.107308.

[2] L.A. Vale-Silva, K. Rohr, "Long-term cancer survival prediction using multimodal deep learning," Scientific Reports, **11**(1), 1–12, 2021, doi:10.1038/s41598-021-92799-4.

[3] J.A. Smith, J.H. Morgan, J. Moody, "Network sampling coverage III: Imputation of missing network data under different network and missing data conditions," Social Networks, **68**(June 2021), 148–178, 2022, doi:10.1016/j.socnet.2021.05.002.

[4] N. Kumar, M.A. Hoque, M. Sugimoto, "Kernel weighted least square approach for imputing missing values of metabolomics data," Scientific Reports, **11**(1), 1–12, 2021, doi:10.1038/s41598-021-90654-0.

[5] R.J.A. Little, D.B. Rubin, Statistical Analysis with Missing Data, John Wiley&Son, 2002.

[6] M.N. Norazian, Y.A. Shukri, R.N. Azam, A.M.M. Al Bakri, "Estimation of missing values in air pollution data using single imputation techniques," ScienceAsia, **34**(3), 341–345, 2008, doi:10.2306/scienceasia1513-1874.2008.34.341.

[7] P.T. Von Hippel, "The Bias and Efficiency of Incomplete-Data Estimators in Small Univariate Normal Samples," Sociological Methods and Research, **42**(4), 531–558, 2013, doi:10.1177/0049124113494582.

[8] A.F. C.R. Rao, H. Toutenburg, Linear models: least squares and alternatives, 2nd ed., Springer Verlag, 1999.

[9] A.C. Acock, "Working With Missing Values," Journal of Marriage and Family, **67**(November), 1012–1028, 2005.

[10] D.W. A. Rotnitzky, "A Note on the biased of estimators with missing data," Biometrics, **50**, 1163–1170, 1994.

[11] M.H. Gorelick, "Bias arising from missing data in predictive models," Journal of Clinical Epidemiology, **59**(10), 1115–1123, 2006, doi:10.1016/j.jclinepi.2004.11.029.

[12] P.L. Roth, J.E. Campion, S.D. Jones, "The impact of four missing data techniques on validity estimates in Human Resource Management," Journal of Business and Psychology, **11**(1), 101–112, 1996, doi:10.1007/BF02278259.

[13] G. Fitzmaurice, "Missing data: implications for analysis," Nutrition, **24**, 200–202, 2008.

[14] T.W. Anderson, "Maximum likelihood estimates for the multivariate normal distribution when some observations are missing," Journal of the American Statistical Association, **52**, 200–203, 1957.

[15] A.M. C. Gourieroux, "On the problem of missing data in linear models," Review of Economic Studies, **48**(4), 579–586, 1981.

[16] J. Sinsomboonthong, "Jackknife maximum likelihood estimates for a bivariate normal distribution with missing data," Journal of Thai Statistical Association, **9**(2), 151–169, 2011, doi:10.1214/aos/1176345020.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 170-175 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Study on Deformation Behavior of Sediments and Applicability of Sealants in Seabed Mining

Takashi Sasaoka[1], Hiroto Hashikawa[1], Akihiro Hamanaka[1,*], Hideki Shimada[1], Keisuke Takahashi[2]

[1]*Department of Earth Resources Engineering, Faculty of Engineering, Kyushu University, Fukuoka 819-0395, Japan*

[2]*Ube Ind. Ltd, Yamaguchi 755-8633, Japan*

ARTICLE INFO

Article history:
Received: 09 June, 2021
Accepted: 12 July, 2021
Online: 20 July, 2021

Keywords:
Rare-Earth
Sealing material
Submarine Mining
Suction mining
Surface coverage

ABSTRACT

The importance of rare earth resources is increasing and a lot of investigations are conducting all over the world. As a result, it was discovered that abundant deep-sea mud contained rare-earth elements on the deep-sea floor. The suction mining method can be one of the effective seabed mining methods to recover these seafloor sediments. However, it is required to evaluate the deformation behavior of the sediments in seabed mining in terms of environmental evaluation. For this reason, this study investigates the deformation behavior of sediments with different water contents by a laboratory suction test. In the test, the sediments filled in a box whose size is 155 mm× 50 mm×180 mm are vacuumed with a suction pump. The suction pressure of the pump is adjusted to 4.0 kPa, the diameter of the suction pump is 10 mm, and the duration of suction is 8 seconds. The results show that suction volume increases with an increase of water content/liquid limit ratio. In addition, the deformation behavior can be categorized as three shapes based on water content/liquid limit ratio; sharp, cone-shaped, and gentle circular arc when the ratio of water content/liquid limit is under 1.3, from 1.3 to 1.6, and over 1.6, respectively. Furthermore, the application of sealants on the sediment surface is effective to reduce the environmental disturbance although its density has to be the same level as the density of sediments to inhibit sinking the sealants.

1. Introduction

The demand for rare mineral resources is growing day by day. Rare earth elements have an important role in modern industry and are used in recent technologies, e.g., electric vehicles, wind turbines, solar panels, rechargeable batteries, mobile phones, a light-emitting diode (LED), and laser system. A lot of investigations are conducted to discover new ore deposits all over the world. The results of various surveys have confirmed the presence of abundant deep-sea mud which contains the rare-earth elements as the seafloor sediments [1, 2], meaning that it could be a promising option as a source of supply of rare-earth resources for the future if these untouched resources on the deep seafloor can be recovered. Some heavy machines/systems are introduced for the exploitation of resources of the seabed for digging, collecting, and transportation [3]. Regarding the deep-sea sediment contained the rare-earth elements, suction mining can be preferred as one of the effective methods because the sediments are the aggregation of fine particles and can be suctioned. In addition, the suction mining method has higher applicability than other mining methods using heavy machines when ore deposits exist on and under the deep seafloor. On the other hand, the development of these untouched seabed resources must be carried out considering the environmental impacts on the ecosystem in the sea to protect biodiversity. The development of deep-sea minerals does not still incorporate into society, and local communities care about the environmental impact on the ocean ecosystems due to the lack of investigations [4].

Mining developments in land-based mining such as surface and underground mining causes adverse environmental impacts historically: water pollutions, the impact on groundwater hydraulics and surface topography, the destruction of habitat/ecosystems, and the loss of diversity. These adverse impacts are often the results of poor planning, the inadequate process of environmental remediation, and insufficient inspection after mine closure due to the lack of standard regulations. The regulation of seabed mining is not established yet because there are insufficient baseline data of environment in the deep-sea is lacking [5]. Therefore, various researches have been reported recently in

*Corresponding Author: Akihiro Hamanaka E-mail: hamanaka@mine.kyushu-u.ac.jp

terms of the regulations known as 'Mining Code' [5-9]. Several approaches suggested to establish the regulation of deep seabed mining such as adaptive management and utilizing the legal framework in the surface mining reclamation [10, 11].

Regarding the suction mining, the hybrid system of air-lift pump combined with the device of dredging using the jet pump is expected to adopt. The air-lift pump is an expected technology to transport the seabed minerals from the deep seafloor, but the extraction of the seafloor sediments with this system may be not efficient. Therefore, another mechanism to extract the sediments requires for the efficient mining system. Some devices have been developed for dredging using a jet pump in the civil engineering field [12-14]. The technology of dredging can be adopted for the extraction of seabed resources. However, it is required to evaluate the deformation behavior of the sediments in seabed mining because it causes suspension and topographic variation. For this reason, it is important to estimate and control the deformation behavior of the sediments in seabed mining. In addition, the impacts on the surrounding environment include the diffusion and redeposition of suspended particles caused by mining. In fact, sediments containing toxic metals such as arsenic have been found near seabed hydrothermal deposits [15, 16]. Thus, sealants are considered as one of the potential methods to reduce the environmental impact in seabed mining [17]. This study uses cement-based materials as sealants based on ground improvement technology in the civil engineering field. Ground improvement is often used to improve the stability of buildings by solidifying soft ground and making it strong by adding improvement materials. It is also widely used for additives and technologies to immobilization of toxic substances in contaminated soil/ground. Cement-based materials are also adopted to construct offshore structures using underwater concreting. Several kinds of research investigated the properties of cement-based materials in deep-sea conditions and showed the results of the deterioration mechanism due to the microstructural changes with pressurization [18, 19]

The authors invented an environment-friendly underwater mining method that aims to control the dispersion of seabed surface sediments and fill the extracted areas by sealing a submerged mine site with anti-washout cement-based sealants. That is to say, sealants can restrain suspension by covering the mining area in advance. Figure 1 illustrates the idea of the underwater mining method. This paper discussed the deformation behavior of the sediments which contain the rare-earth elements on the deep sea, and the application of sealants in the suction mining method by means of several laboratory tests.

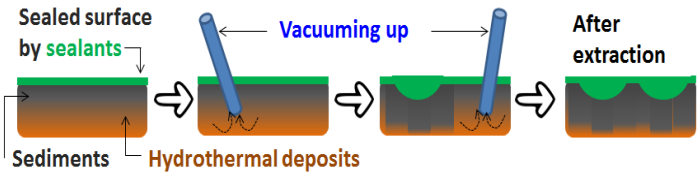


Figure 1: Environment-friendly Submarine Mining

2. Preparation of Sediment and Suction Test

2.1. Sediment Samples

The sediments contained the rare-earth elements on the deep sea are classified as clay on soil classification according to the previous investigation results by Ministry of Economy, Trade and Industry. The state of sediments is saturated in the seafloor, indicating that the liquid limit and water content can be the related parameters to show their properties. However, the liquid limit and water content of the sediments are different depending on the place as shown in Table 1, indicating that it is expected that these differences affect the deformation behavior of the sediments on suction mining. Therefore, some sediment samples which had a different liquid limit and water content were prepared for the suction test. Those samples were prepared by blending two types of bentonite (see Table 2) in different mixing ratios. Liquid limit of each soil sample is 50%, 60%, 70%, 80%, 90%, 100%, respectively. Each mixing ratio is shown in Table 3.

Table 1: Property of Sample of Deep-sea Mud
(Ministry of Economy, Trade and Industry in Japan, 2016)

	Density (g/cm^3)	Liquid limit (%)	Water Content (%)
A-1	2.850	116.3	124.1
A-2	2.831	111.1	138.9
A-3	2.792	98.7	156.3
A-4	2.792	105.4	140.3
B	2.833	117.1	128.8
C	2.821	109.8	146.1

Table 2: Property of Bentonite

	Density (g/cm^3)	Liquid limit (%)
Bentonite I	2.507	48.84
Bentonite II	2.661	110.01

Table 3: Mixing Ratio and Liquid Limit of Sediments

	Liquid limit (%)	Mixing ratio (%)	
		I	II
W_L=50	50	99.06	0.94
W_L=60	60	80.50	19.50
W_L=70	70	61.94	38.06
W_L=80	80	43.38	56.62
W_L=90	90	24.82	75.18
W_L=100	100	6.26	93.74

2.2. Suction Test

A conceptual diagram of the suction test is shown in Figure 2. The sediment samples which arranged the water contents from 60~160% were filled into the plastic box. In the suction test, the ratio of water content and liquid limit (W_C/W_L) was defined to evaluate the fluidity of sediment samples quantitatively. The water content/liquid limit ratio of the sediment samples is from 1.00~1.85. The length, width, and height of the sediments are 155 mm, 50 mm, and 130 mm, respectively. The sediments are vacuumed with a suction pump from the center of the surface of the sediments. The suction pressure of the pump is adjusted to 4.0 kPa (maximum suction pressure of the pump is 21.4 kPa). The

diameter of the suction pump is 10 mm and the duration of suction is 8 seconds. The suction target is 50 mm depth from the simulated soil surface, meaning that the pump moves downward in 6.25 mm/sec from the surface of the sediments. At the end of the test, the suction volume and deformed shape were measured to evaluate the deformation behavior of the sediment samples. The deformed shape of sediment samples is identified by measuring the maximum vertical length (V) and horizontal length (H) of the deformation area. Besides, the span of influence was defined as H/V as a normalized parameter to show the deformation behavior.

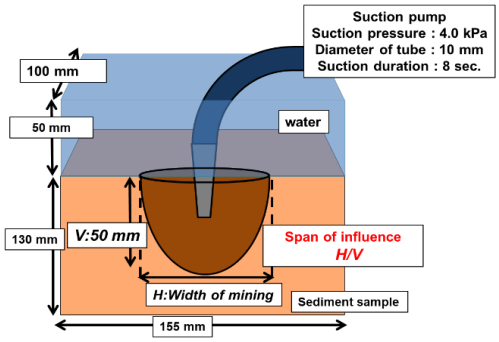


Figure 2: Conceptual Diagram of Suction Test

2.3. Results and Discussion

Figure 3 shows the relation of the ratio of water content/liquid limit and suction volume. Suction volume increases exponentially with an increase of the water content/liquid limit ratio: its volume is 18-70 mL when W_C/W_L is 1.0 while its volume increases to 194-533 mL when W_C/W_L is 1.7. Figure 4 shows the relation of the water content/liquid limit ratio and span of influence. The span of influence also increases with an increase of water content/liquid limit ratio. It is seemingly categorized into 3 parts: H/L is around 1.0 when the W_C/W_L is less than 1.3, H/L is around 2.0 when the W_C/W_L is from 1.3-1.6, and H/L is more than 2.0 when the W_C/W_L is more than 1.6. From the aspect of deformation behavior of sediment samples, the behavior can be classified into 3 types; sharp deformation, cone-shaped deformation, gentle circular arc deformation as shown in Figure 5. In comparison with the water content/liquid limit ratio and the span of influence, the deformation behavior can be defined with the water content/liquid limit ratio. It is respectively categorized as sharp, cone-shaped, and gentle circular arc when the ratio of water content/liquid limit is under 1.3, from 1.3 to 1.6, and over 1.6. In comparison with deformation behavior and suction volume, the suction volume is small when the deformation behavior is sharp. This result indicates that less sediments can be recovered in the limited area at one time, meaning that the efficiency of seabed suction mining is expected to be small although the environmental impact is also expected to be small because the span of influence is small. On the other hand, the suction quantity is large when the deformation behavior is a gentle circular arc. This result indicates that many sediments can be recovered in a wide range at one time, indicating that the efficiency of seabed suction mining is expected to be large whereas the environmental impact is also large.

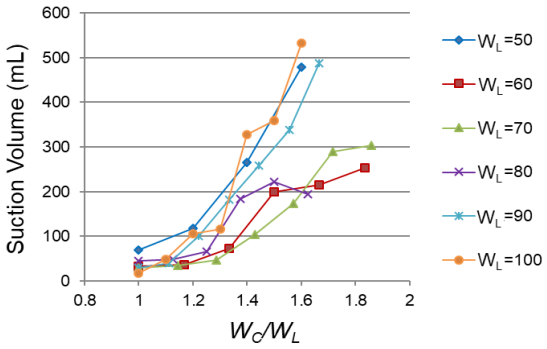


Figure 3: Relation of W_C/W_L and Suction Volume

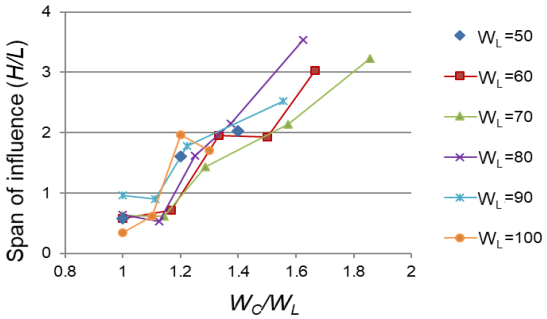
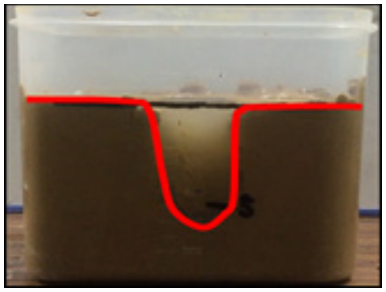
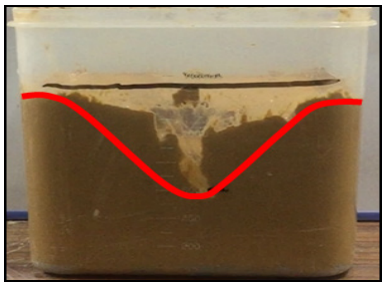


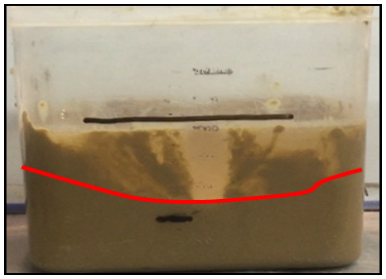
Figure 4: Relation of W_C/W_L and Span of Influence



(a) W_C/W_L<1.3 (sharp)



(b) W_C/W_L=1.3~1.6 (cone-shaped)



(c) W_C/W_L>1.6 (gentle circular arc)

Figure 5: Classified of Deformation Behavior of Sediment Sample

3. Application of Sealants as Surface Cover

3.1. Sealants

It is important to evaluate and consider the environmental impact properly. At present, deformation of seafloor and diffusion of the suspended particle is concerned while the environmental standard has not been established yet in seabed resources mining. Sealants are considered as one of the methods to reduce environmental impact in seabed mining. The sealants are based on ground improvement techniques and enhanced anti-washout ability under the water. The technique is commonly used to improve ground stability by injecting cement [20]. In addition, these materials are also utilized for restraining the elution of detrimental substances by covering mining areas as stated previously [21]. Figure 6 shows the conceptual diagram of seabed suction mining covered with sealants. These materials are, additionally, able to react along the line of deformation of landforms and cracks because of their viscosity. From these considerations, it is expected to prevent the diffusion of toxic materials on the seafloor and protect environmental effects and collapse of the ground by applying sealants as a surface cover in seabed mining [22]. In this study, slag-type sealants which contain slag, polycarboxylic ether, superplasticizer, and hydroxyethyl cellulose are used [23]. The water/powder ratio is 30%.

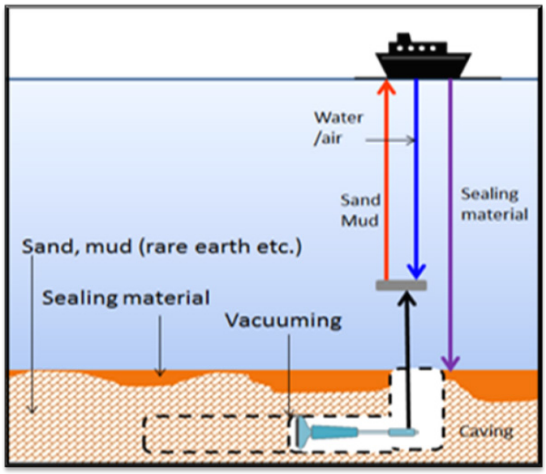


Figure 6: Conceptual Diagram of Seabed Suction Mining

3.2. Covering Test

The sealants have to be covered on the surface of the sediments to reduce environmental disturbance. However, it is suspected to sink the sealants into the sediments due to their weight. Therefore, the covering test using sediment samples and sealants was conducted to investigate the effectiveness of surface covering with sealants. In this test, the sealants are poured into the surface of the sediment samples with a funnel. The thickness covered by the sealants was 10 mm. The sediment sample is prepared with a lower density; the density is 1.23 g/cm^3 and W_C/W_L is 1.7. The conceptual diagram of the covering test is shown in Figure 7. The sealants which have the different density (1.25 ~ 1.60 g/cm^3) were prepared in this test. The density of sealants was arranged with the modification of aggregate weight. The surface area of the container was 28.8 cm^2. Additionally, the surface coverage ratio is calculated with the binarization of image analysis. This study uses Image J for the image analysis [24]. Figures 8 show the relation of coverage ratio and the density of sealant. The images were taken from the top of sealants, meaning that the coverage ratio is 100% if the sealants remain above the sediment sample without sinking. It is clear that the coverage ratio can be improved by decreasing the density of sealants. The surface coverage is approximately 100% if the density of the sealants is below 1.4 g/cm^3 while the surface coverage is decreased if the density of the sealants is more than 1.5 g/cm^3, meaning that the sinking of the sealants occurs. Considering the density of the simulated sediment is 1.23 g/cm^3, it is possible to suppress the sinking of the sealant by applying the sealant of the same level as the density of sediment samples.

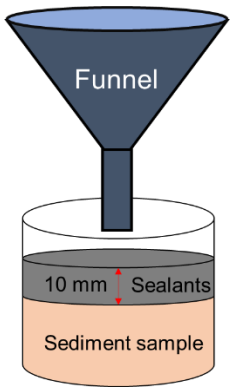


Figure 7: Conceptual Diagram of Covering Test

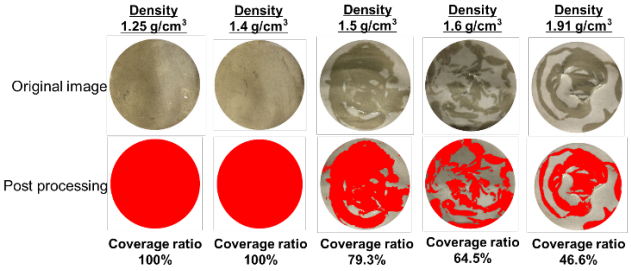


Figure 8: Relation of Density and Coverage Ratio

3.3. Suction Test with Sealants

The suction test with sealants was conducted to evaluate the deformation behavior of sediment samples when sealants were applied on the surface. The suction volume and span of influence were measured in this test. The thickness of sealants is 10 mm. Other experimental conditions are the same as the test in the case of sediment samples only. Figure 9 shows the relation of the ratio of water content/liquid limit and suction volume. In either case, suction volume increases with an increase of water content/liquid limit ratio. The suction volume may be less affected with/without sealants while some cases show the decreasing trend: the suction volume shows from 44.76-337.63 mL without sealants and from 61.19-270.14 mL with sealants. Figure 10 shows the relation of the

ratio of water content/liquid limit and span of influence. The span of influence also increases with an increase of water content/liquid limit ratio. Additionally, the span of influence decreases when the surface is covered by sealants: it shows from 1.44-3.54 without sealants and from 0.33-2.27 with sealants. This might be suspected that the adhesive force between sealants and sediments restraint the movement of sediments surface. As a result, it is indicated that the application of sealants can reduce the environmental impact like suspended particles and topographic variation.

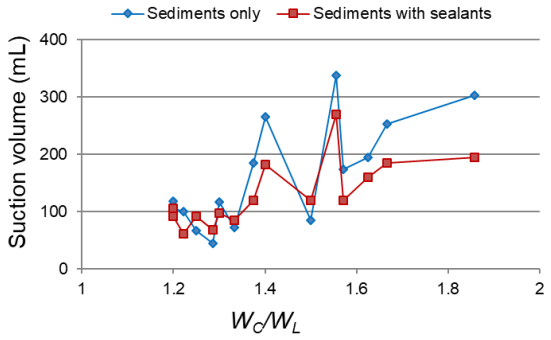


Figure 9: Relation of W_C/W_L and Suction Volume

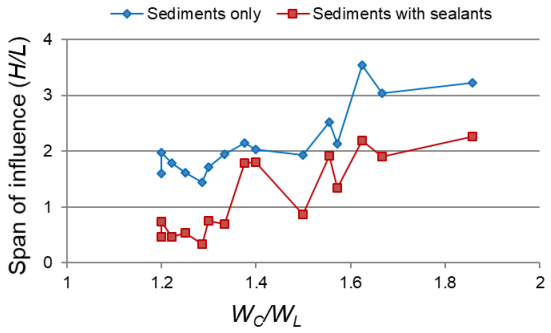


Figure 10: Relation of W_C/W_L and Span of Influence

4. Conclusion

The abundant deep-sea mud which contains rare-earth elements has been found as the seafloor sediments. The development of these untouched resources contributes to a diversification of the supply sources of the rare-earth resources. In this study, the recovery of sediments using seabed suction mining is discussed on a laboratory scale. In the results, the deformation behavior of the sediments with suction mining can be assessed by the ratio of water content/liquid limit. In addition, it is shown that the suction volume increases with the ratio of water content/liquid limit. In addition, the deformation behavior changes to sharp deformation, cone-shaped deformation, and gentle circular arc deformation based on the water content/liquid limit ratio. It is clarified that the coverage decreases when the density of sealants is larger than that of the sediments because the sealants sink into the sediments due to their weight. Therefore, the density of sealants should be arranged to the same level as the density of sediment samples by selecting the proper aggregate. Furthermore, it is indicated that sealants can reduce the environmental impact like suspended particles and topographic variation because adhesive force between sealants and simulated soil restraint the movement of sediments surface.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP 19K05351.

Reference

[1] K. Hirai, "Trends in Metal Resource Development and Strategies for Resource Security", Surface Science for Resource. **35**(2), 114-115, 2014, doi: 10.1380/jsssj.35.114.

[2] Y. Kato, K. Fujinaga, K. Nakamura, Y. Takaya, K. Kitamura, J. Ohta, R. Toda, T. Nakashima, H. Iwamori, "Deep-sea mud in the Pacific Ocean as a potential resource for rare-earth elements", Nature Geoscience, **4**(8), 535–539, 2011, doi: 10.1038/ngeo1185.

[3] N. Toro, P. Robles, R.I. Jeldres, "Seabed mineral resources, an alternative for the future of renewable energy: A critical review", Ore Geology Reviews, **126**, 2020, doi: 10.1016/j.oregeorev.2020.103699.

[4] R. Motoori, B.C. McLellan, "Resource security strategies and preferences for deep ocean mining from a community survey in Japan", Marine Policy, **128**, 2021, doi: 10.1016/j.marpol.2021.104511.

[5] A. Hallgren, A. Hansson, "Conflicting narratives of deep sea mining", Sustainability (Switzerland), **13**(9), 2021, doi: 10.3390/su13095261.

[6] W. Leal Filho, I.R. Abubakar, C. Nunes, J. Platje, P.G. Ozuyar, M. Will, G.J. Nagy, A.Q. Al-Amin, J.D. Hunt, C. Li, "Deep seabed mining: A note on some potentials and risks to the sustainable mineral extraction from the oceans", Journal of Marine Science and Engineering, **9**(5), 2021, doi: 10.3390/jmse9050521.

[7] A. Jaeckel, "Strategic environmental planning for deep seabed mining in the area", Marine Policy, **114**, 2020, doi: 10.1016/j.marpol.2019.01.012.

[8] L.J. Gerber, R.L. Grogan, "Challenges of operationalising good industry practice and best environmental practice in deep seabed mining regulation", Marine Policy, **114**, 2020, doi: 10.1016/j.marpol.2018.09.002.

[9] V. Tunnicliffe, A. Metaxas, J. Le, E. Ramirez-Llodra, L.A. Levin, "Strategic Environmental Goals and Objectives: Setting the basis for environmental regulation of deep seabed mining", Marine Policy, **114**, 2020, doi: 10.1016/j.marpol.2018.11.010.

[10] J. Hyman, R.A. Stewart, O. Sahin, "Adaptive Management of Deep-Seabed Mining Projects: A Systems Approach", Integrated Environmental Assessment and Management, 2021, doi: 10.1002/ieam.4395.

[11] M. Squillace , "Best regulatory practices for deep seabed mining: Lessons learned from the U.S. Surface Mining Control and Reclamation Act", Marine Policy, **125**, 2021, doi: 10.1016/j.marpol.2020.104327.

[12] W. Kong, "Design of dredging device for immersed gravel bed and analysis of water jet", in 2021 7th International Symposium on Mechatronics and Industrial Informatics, ISMII 2021, 52–56, 2021, doi: 10.1109/ISMII52409.2021.00018.

[13] Y. Zhang, J. Song, "Study on fluidization process in jet flow dredging device and the effects of device structure to sand collecting performance", in IOP Conference Series: Materials Science and Engineering, 2020, doi: 10.1088/1757-899X/892/1/012117.

[14] M.K. Sarkar, S. Sarkar, "Assisting pumps for dredging", Lecture Notes in Civil Engineering, **23**, 571–579, 2007, doi: 10.1007/978-981-13-3134-3_43.

[15] T. Yamanaka, K. Maeto, H. Akashi, J. Ishibashi, Y. Miyoshi, K. Okamura, T. Noguchi, Y. Kuwahara, T. Toki, U. Tsunogai, T. Ura, T. Nakatani, T. Maki, K. Kubokawa, H. Chiba, "Shallow submarine hydrothermal activity with significant contribution of magmatic water producing talc chimneys in the Wakamiko Crater of Kagoshima Bay, southern Kyushu, Japan", Journal of Volcanology and Geothermal Research, **258**, 74–84, 2013, doi: 10.1016/j.jvolgeores.2013.04.007.

[16] H. Sakamoto, "The Distribution of Mercury, Arsenic, and Antimony in Sediments of Kagoshima Bay", Bulletin of the Chemical Society of Japan, **58**, 580–587, 1985, doi: 10.1246/bcsj.58.580.

[17] K. Takahashi, T. Yamanaka, H. Shimada, T. Sasaoka, A. Hamanaka, "Application of Cement-based Sealants for Prevention and Remediation of Environmental Impact of Submarine Resource Mining", in 26th International Symposium on Mine Planning and Equipment Selection, 363–

367, 2017.
[18] M. Kobayashi, K. Takahashi, Y. Kawabata, "Physicochemical properties of the Portland cement-based mortar exposed to deep seafloor conditions at a depth of 1680 m", Cement and Concrete Research, 2021, doi: 10.1016/j.cemconres.2020.106335.
[19] K. Takahashi, Y. Kawabata, M. Kobayashi, S. Gotoh, S. Nomura, T. Kasaya, M. Iwanami, "Action of Hydraulic Pressure on Portland Cement Mortars – Current Understanding and Related Progress of the First-Ever In-Situ Deep Sea Tests at a 3515 m Depth", Journal of Advanced Concrete Technology, **19**, 226-239, 2021, doi: 10.3151/jact.19.226.
[20] H. Shimada, A. Hamanaka, T. Sasaoka, K. Matsui, "Development of Injection Material for Offshore Structures Using Flyash-Surfactant Mixtures", Development and Applications of Oceanic Engineering , **2**(3), 70–76, 2013.
[21] R.A, Wuana, F.E. Okieimen, Heavy metals in contaminated soils: A review of sources, chemistry, risks, and best available strategies for remediation, Apple Academic Press, 2014.
[22] S. Sakamoto, S. Matsumoto, T. Sasaoka, H. Shimada, S. Fujita, K. Takahashi, "Fundamental Study on Deformation Behavior of Seafloor Covered with Sealing Materials in Seabed Mining", in International Symposium on Earth Science and Technology 2016, 74–78, 2016.
[23] H. Shimada, T. Sasaoka, S. Fujita, S. Wahyudi, Y. Yoshida, "Construction of seabed structures by new development covering material using fly ash", Inzynieria Mineralna, **17**(1), 143–151, 2016, doi: 10.29227/IM-2016-01-22.
[24] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D.J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, "Fiji: An open-source platform for biological-image analysis", Nature Methods, **9**(7), 676–682, 2012, doi: 10.1038/nmeth.2019.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 176-183 (2021)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Multi-Robot System Architecture Design in SysML and BPMN

Ahmed R. Sadik*, Christian Goerick

Honda Research Institute Europe, 63073, Germany

ARTICLE INFO

Article history:
Received: 30 November, 2020
Accepted: 08 July, 2021
Online: 20 July, 2021

Keywords:
Multi-robot System Model
Model-based System Engineering
Multi-agent Simulation
Systems Modeling Language
Process Model and Notation
Java Agent Development

ABSTRACT

Multi-Robot System (MRS) is a complex system that contains many different software and hardware components. This main problem addressed in this article is the MRS design complexity. The proposed solution provides a modular modeling and simulation technique that is based on formal system engineering method, therefore the MRS design complexity is decomposed and reduced. Modeling the MRS has been achieved via two formal Architecture Description Languages (ADLs), which are Systems Modeling Language (SysML) and Business Process Model and Notation (BPMN), to design the system blueprints. By using those abstract design ADLs, the implementation of the project becomes technology agnostic. This allows to transfer the design concept from on programming language to another. During the simulation phase, a multi-agent environment is used to simulate the MRS blueprints. The simulation has been implemented in Java Agent Development (JADE) middleware. Therefore, its results can be used to analysis and verify the proposed MRS model in form of performance evaluation matrix.

1. Introduction

This paper extends the work presented at the 2019 International Conference on Mechatronics, Robotics, and System Engineering (MoRSE) [1]. Related work can be also seen in [2].

Multi-Robot System (MRS) is a cyber-physical system that contains more than one robot, each of them owns a unique set of capabilities. The idea of an MRS is to solve a complex problem by collectively using the current capabilities of existing robots [3]. Therefore, the MRS must match the given problem with the existing robots' capabilities, to plan the solution steps. Many MRS applications can be seen in swarm robotics, cooperative automated transportation, unmanned aerial vehicles, and cooperative manufacturing [4]. The advantages of an MRS is increasing the performance by saving the time and the effort to solve the problem. Moreover, distributing the solution among different robots provides more computational processing power, this means faster and higher capacity to solve many problems simultaneously [5].

Implementing an MRS without a proper system architecture design is a crucial mistake that is often done by the system developers. Because the system requirements and functionalities are lost in a non-human readable machine code. Therefore, in this article we purpose a model driven development approach that uses the system model as the main software artifacts [6]. The proposed design approach in this article is based on the V-Model, which is a de facto solution for complex systems such as MRS.

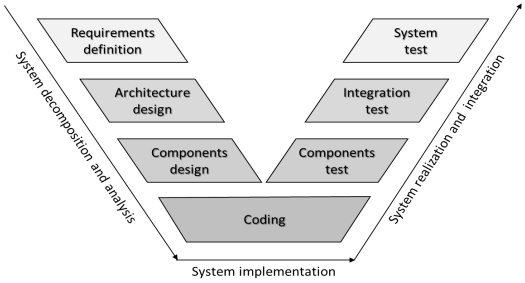


Figure 1: The V-Model simplified version – adapted from [6]

The V-Model shown in Figure 1 describes the required stages to build an MRS. In the first stage of the V-Model, the system is decomposed. In this stage, the system components and architecture are designed based on the system requirements. In the second stage, the implementation of the MRS is carried out. The implementation of an MRS often involves the coding the individual components. In the final stage, the MRS individual components are tested through unit tests, then integration tests are carried out over sub-systems and eventually the overall integrated system. This article focusses on the first stage of the V-Model to build an MRS. As the design stage is the most curtail stage of an

*Corresponding Author: Ahmed Rabee Sadik, ahmed.sadik@honda-ri.de

https://dx.doi.org/10.25046/aj060421

MRS system building, because all the following stages are depending on this design.

Section 2 of the article describes the problem and the use case of concern. Section 3 introduces the background that that is needed to model and simulate the use case. Section 4 discusses the system requirements that are used to build and evaluate the system performance Modeling the use case is explained in detail in section 5, while its simulation is shown in section 6. Therefore, the performance analysis is explained in section 7. Ultimately, the last section concludes the work and the future research.

2. Problem and use case

The main article objective is to design an MRS architecture model that can be simulated and evaluated due to a predefined evaluation criterion. An MRS architecture is an overall system description that abstracts its functionalities, logic, and constrains [7]. Accordingly, it provides an analysis tool to grasp and improve system characteristics, and a conceptual model that can be used as the system blueprints [8]. In this work, SysML block definition diagram is used to describe the proposed MRS architecture and components as shown in Figure 2 and Figure 3. SysML diagrams will be explained in the next section as many of them are used in constructing the proposed MRS design.

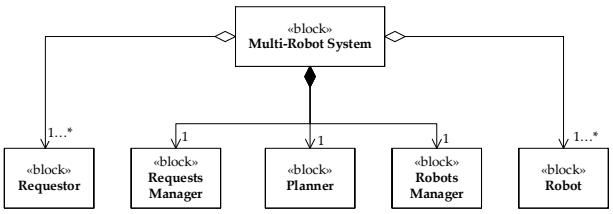


Figure 2: SysML block definition diagram for the proposed architecture

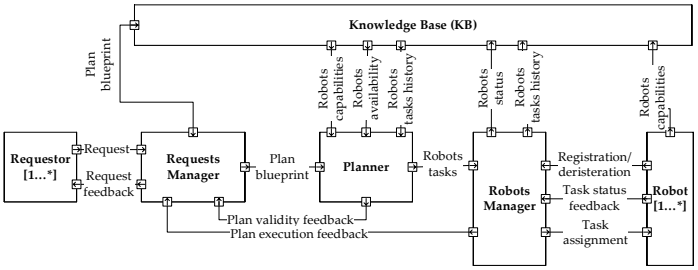


Figure 3: SysML internal block diagram for the proposed MRS architecture

Figure 2 is the proposed MRS block definition diagram. The block definition diagram defines the main components of the architecture, which are the Requests Manager (RqM), the planner (PLN), and the Robots Manager (RbM). Figure 3 shows the proposed MRS internal block diagram that describes the connections among the components as illustrated in Figure 2. When the RqM receives a request (Rq), it checks if there is a plan-blueprint (Pb) in the Knowledge Base (KB) to fulfill this Rq. A Pb is a sequence of tasks (T), i.e., $Pb_i = \{T_1,...,T_n\}$, where n is the number of tasks and could be different from one blueprint to another. A task is a function of the capabilities (C) of the robot (R), i.e., $T_i = f(C_x, C_y, \ldots)$, where each robot owns different capabilities set. If the RqM finds a match between Pb to and a Rq, it forwards the Pb to the PLN. The PLN checks the robots' availability, and their capabilities to achieve the tasks in the Pb. If more than a robot owns the capabilities to fulfill the task, the PLN compares the number of tasks that have been achieved by these robots in the past. Based on this comparison, the PLN selects a robot to assign for the task. If the PLN complete the matching of all the tasks with the robots, it sends a verified plan (Pv) to the RbM. The RbM sends the tasks to the robots and waits their response.

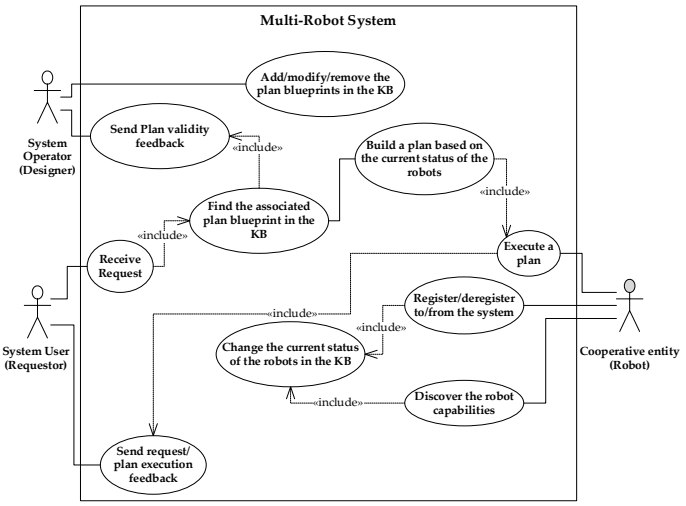


Figure 4: UML/SysML use case diagram for the proposed architecture

The use case diagram in Figure 4 shows three variation types that are considered during the simulation. First is the Pbs variation, by adding, editing, or omitting a Pb. Second is variation in the number of the available robots. The maximum number of robots that can exist is constrained to three. The robots are constrained to register or deregister through the RbM. Third is the variation in the robots' capabilities, by updating or editing the capabilities of a robot. the robot is constrained to deregister to be able to update its capabilities, then register again through the RbM, which automatically updates the robot new capabilities in the KB.

3. Solution preliminaries

3.1. Systems Modeling Language

SysML is a general-purpose modeling language that is derived from Unified Modeling Language (UML) [9]. SysML and UML belong are both developed by Object Management Group (OMG). UML is is a visual modeling language that is particularly used to construct, design, and document the software systems in fields such as web-development, telecommunication, banking, and enterprise services [10]. While SysML is extending and modifying UML diagrams to fit complex industrial systems that involve variety of hardware, software, information, and processes (e.g., Aviation, Space, Automotive) [11].

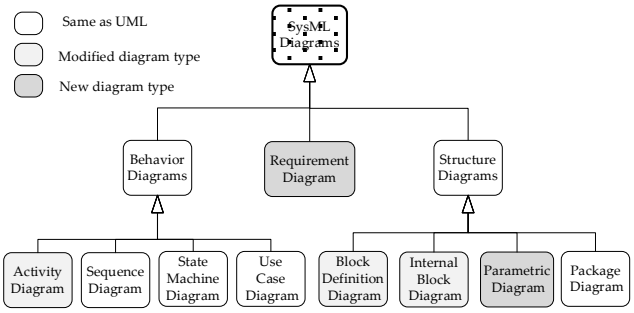


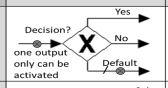
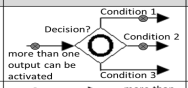
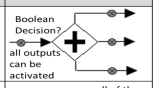
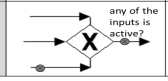
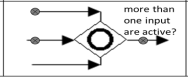
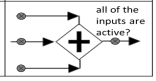
Figure 5: SysML Taxonomy and comparison to UML

Figure 5 shows the relation between SysML and UML graphs [12]. Requirement and parametric diagram are two new diagrams that distinguish SysML [13]. Section 4 of this article used the requirement diagram to define the system performance criteria requirements and their relations. Block definition diagram and internal block diagram are used to describe the main components of the system architecture and the connection among the components as illustrated in section 2, while the use case diagrams describe the interaction of the system as a black box with the external world or actors. The activity diagram is used in section 5 to represent the MRS components logic. The more detailed logic is represented in the activity model, the easier to automatically generate a low-level code from this activity model. For this reason, BPMN is used to build the MRS logic, as it extends the notations, semantics, and syntax of SysML and UML activity diagram. The state machine diagram is used in section 4 to model the internal states of the robot. While the sequence diagram is used in section 6 to represent the interaction and communication among the components during a simulation scenario.

3.2. Business Process Model and Notation

Since UML activity diagram provides an abstract high-level process description, BPMN extends the UML activity diagram to fulfill the following two drawbacks. First, UML activity diagram lakes the syntax and the logical execution among the actions. Second, the poverty in UML notations and semantics in comparison with BPMN [14].

Table 1: BPMN control gateways

Rule	Gateways		
	Exclusive OR (XOR)	Inclusive OR	Parallel AND
Split			
Merge			

Flow control gateways is the best example to demonstrate how BPMN is improving UML activity diagram. Flow control gateways are all equivalent to only one notation in UML, which is the decision notation. Table 1 shows the notations, semantics, and syntax of the basic gateways of BPMN. Three different notations are demonstrated in Table 1, which are exclusive-OR, inclusive-OR, and parallel-AND. The three mentioned gates operate either as spilt or merge context. In spilt context, exclusive-OR splits one input to only one output based on the conditions on the output branches. Inclusive-OR splits one input to more than one output simultaneously based on the conditions on the output branches. Parallel-AND splits one input to all the output simultaneously when the input branch is triggered. In merge context, exclusive-OR merges any of the input branches to only one output, when any of the input branches is triggered. Inclusive-OR merges more than one input branches to only one output, when these inputs are simultaneously triggered. Parallel-AND merges all the input branches to only one output, when all the inputs are simultaneously triggered [15].

3.3. Java Agent Development

JADE is a Multi-Agent System (MAS) middleware [16] that has been used in this research to deploy the proposed solution as shown in Figure 6-a. Each entity in the proposed SysML internal block diagram is implemented as a JADE agent. JADE Agent Management System (AMS) address each agent with a unique Identifier (AID) to facilitate the communication among the agents. While JADE directory Facilitator (DF) announces the services that every agent afford. JADE applies the Foundation for Intelligent Physical Agent (FIPA) specifications, to enable agent communication through FIPA-Agent Communication Language (FIPA-ACL) [17].

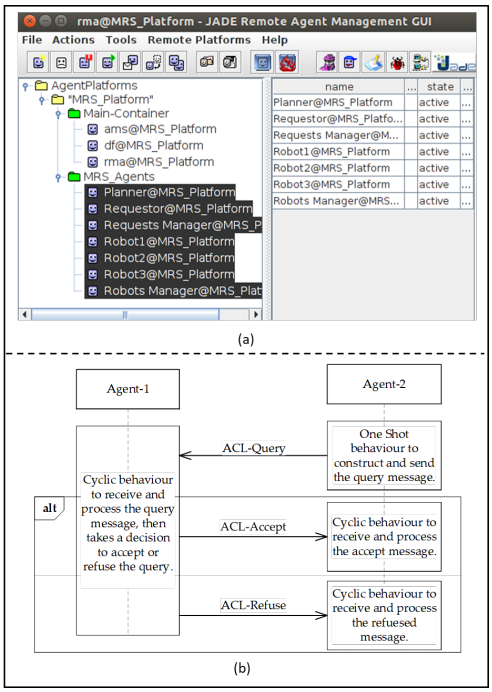


Figure 6 (a) JADE framework – (b) JADE sequence diagram example

Each JADE agent has a complex individual behaviour that can be seen as a composite of two simple behaviours. First is one-shot behaviour that is executed only once when it is triggered. Second is a cyclic behaviour that continuously executed when it is triggered. An example of JADE agent communication and decision making based on their behaviours can be seen in Figure 6-b. JADE is a suitable tool to build an agent simulation based on the MRS SysML/BPMN model. As the MRS logic and architecture can be easily translated to JADE implementation concepts [18].

4. Performance requirements

To evaluate the MRS design, it is necessary to measure the system performance during the simulation. Qualitative criteria such as reusability, scalability, extensibility, and interoperability have been proposed in [19]. However, these criteria are often relatively vague without quantitative performance measurements. Therefore, this research defines the quantitative indicators that are shown in Figure 7. The research assumes that the MRS is a black box that receives different Rq, that can either success or fail during the execution. The following measurements can be used to express the system performance:

- Throughput: the number of requests that are processed.
- Latency: the time needed from the request arrival till the request execution.

- Success rate: the number of request that success to be executed per the overall received requests number.
- Failure rate: the number of request that fail to be executed per the overall received requests number.
- Efficiency: the ration between the success rate and the failure rate.

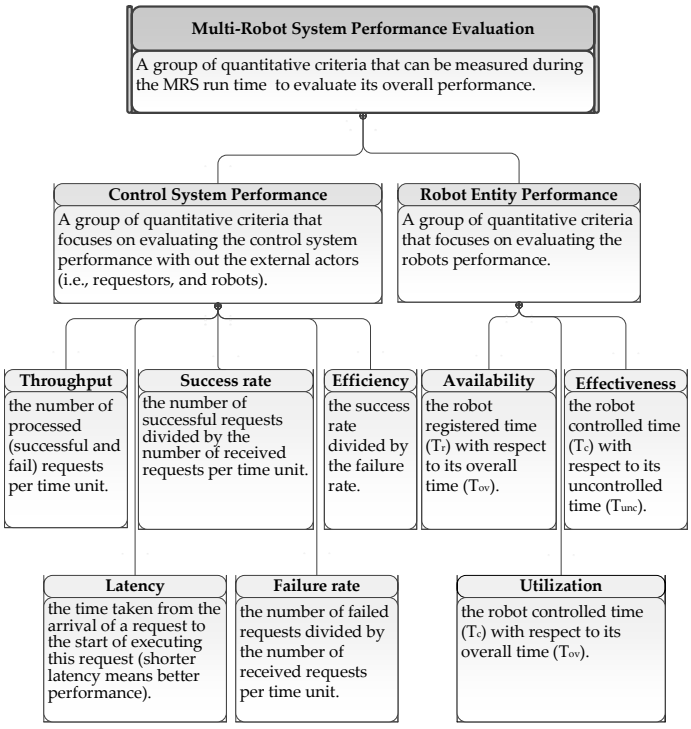


Figure 7: Requirement diagram of the MRS performance evaluation criteria

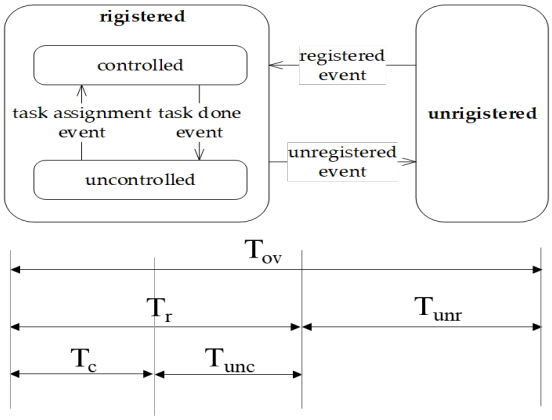


Figure 8: State machine diagram of the robot entity

The robot performance is also considered in this research as another measurement of the MRS performance [20]. The robot performance is fundamentally derived from its state machine diagram that is shown in Figure 8. The the robot state machine is built upon measuring the following times:

- Controlled time (T_c): the time that the robot needs to perform an assigned task.
- Uncontrolled time (T_{unc}): the robot waiting time to be assigned to a task after registration.
- Registered time (T_r): the sum of the controlled and the uncontrolled time of the robot.
- Unregistered time (T_{unr}): the accumulation of the robot unregistered time.
- Overall time (T_{ov}): the sum of the registered and the unregistered time of the robot.

Accordingly, the robot performance criteria are calculated as follows:

- Availability: the ration between the robot registered time (T_r) and the overall time (T_{ov}).
- Utilization: the ratio between the robot controlled time (T_c) and the overall time (T_{ov}).
- Effectiveness: the ration between the robot controlled time (T_c) and the uncontrolled time (T_{unc}).

5. System model

5.1. Requests manager

The RqM receives requests from various requestors, then it looks for an associated Pb within the KB. If the RqM finds the associated Pb, it forwards it to the PLN. The RqM decision making model is shown in Figure 9 via the BPMN activity diagram.

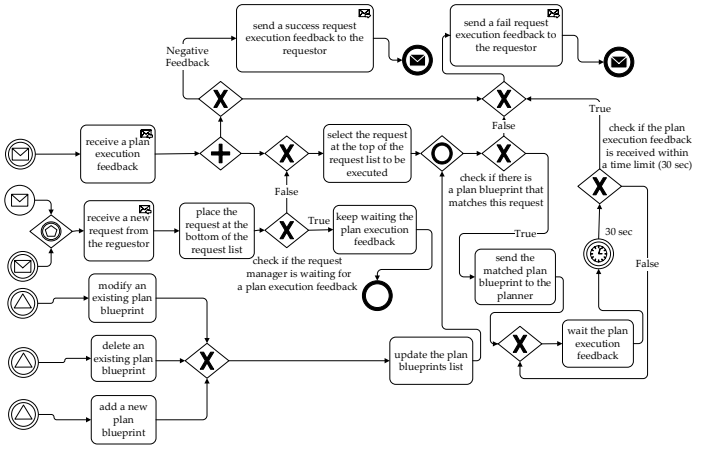


Figure 9: Requests-manager BPMN activity diagram

The RqM uses First Come First Serve (FCFS) technique to schedule the received requests. The RqM checks in the associated Pb for every received request. If there is no associated Pb with the request, the RqM directly sends a negative feedback to the requestor. If the RqM finds an associated Pb to the request, it forwards this Pb to the PLN, and waits for the feedback. If this feedback exceeds predefined limits, the RqM considers this request as a failure one. If not, it waits the execution feedback to forward it to the requestor.

5.2. Planner

The PLN receives the Pb and makes sure that it is visible to build a Pv instance according to the current system status. The PLN decision making model is shown in Figure 10 via the BPMN activity diagram.

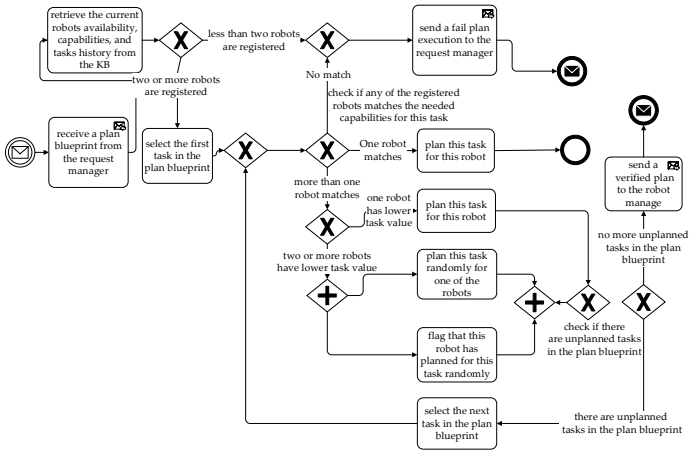


Figure 10: Planner BPMN activity diagram

To construct a Pv instance from a Pb, The PLN checks the available registered robots, the robots' capabilities, and the robots' tasks history. In case that there is only one available robot, the PLN directly considers a plan failure, as it is known in advance that a plan requires at least two robots to get executed. If at least two robots are available, the PLN compares the tasks in the Pb to the available robots' capabilities. If the robots' capabilities do not match the required tasks in the Pb, the PLN considers a plan failure. If there are two robots or more that can perform the same task, the PLN checks their tasks history, and assign the task to the robot that performed less tasks. This is to balance the task assignment among the available robots within the MRS. If all the tasks in the Pb could be assigned to robots, the PLN creates a Pv instance and sends it is the RbM to be executed.

5.3. Robots Manager

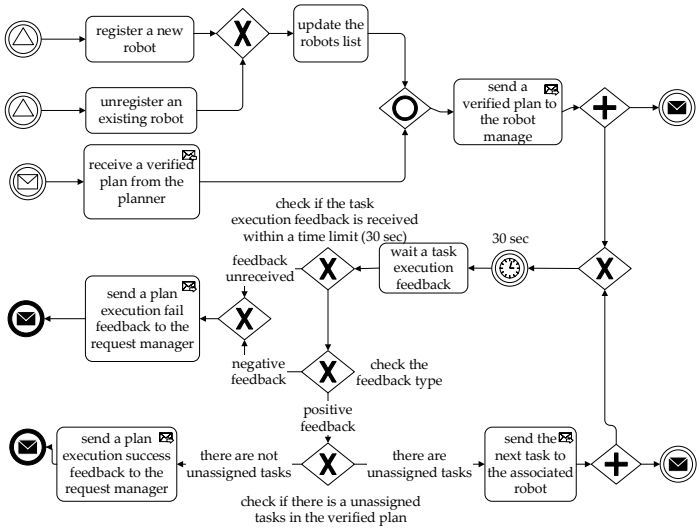


Figure 11: Robots-manager BPMN activity diagram

The RbM receives the Pv, then it assigns the tasks in this Pv to the available robot. Additionally, the RbM is also responsible for registering/unregister the robots from the MRS. this way it monitors the robots' availability. The RbM decision making model is shown in Figure 11 via the BPMN activity diagram.

When the RbM assigns a task to a robot, it waits the robot feedback within a time limit. If the robot feedback did not arrive within the predefined limits, the RbM sends a negative feedback to the RqM. This feedback means that the whole plan is failed to be executed. If the RbM received a positive feedback from the robot within the predefined time limits, it assigns the next task due to the Pv. If all the tasks in the Pv are executed, the RqM sends a positive feedback to the RqM, otherwise it sends a negative feedback.

6. Simulation

The activity diagrams that have been illustrated in the previous section are used as the MRS blueprints. JADE has been used in this research to deploy these blueprints, and hence enables the MRS simulation during the design phase. The Graphical User Interface (GUI) shown in Figure 12 has been created to achieve interact with every entity in the proposed architecture. The RqM GUI in Figure 12-a can be used to add/edit/remove the Pb. The PLN GUI in Figure 12-b is used to monitor the Pv execution, the robots' availability, the robots' status, the robots' capabilities, and the robots' tasks history. The RbM GUI in Figure 12-b is used to show the assigned tasks status.

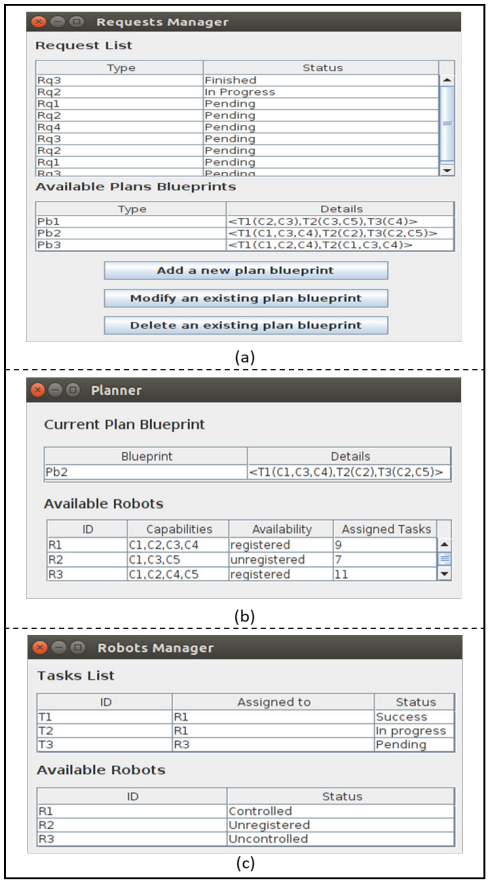


Figure 12: (a) Requests Manager GUI – (b) Planner GUI – (c) Robots Manager GUI

To illustrate the simulation scenario, an interaction example among the MRS entities is show in Figure 13. In this example, The RqM receives Rq_2. Therefore, the RqM sends the Pb in a form of the ACL-message shown in Figure 14-a to the PLN. Accordingly, the PLN constructs a Pv by matching the available robots' capabilities and tasks history with the received Pb. In this case, R_1 and R_3 were registered into the MRS as shown in Figure 14-b. As

T_1 needs (C_1, C_3, C_4) to be executed, T_1 was assigned to R_1, because (C_1, C_3, C_4) are unique capabilities of R_1. Similarly, T_3 was assigned to R_3, as T_3 needs (C_2, C_5) which is unique capability of R_3. However, in case of T_2, both R_1 and R_3 own the capability C_2 which is needed to execute this task. Therefore, the PLN checks both robots' task history to be able to assign T_2. The PLN finds out that R_1 task history is 9 while R_3 task history is 11. Accordingly, the PLN assigns T_2 to R_1, to balance the robots' tasks distribution.

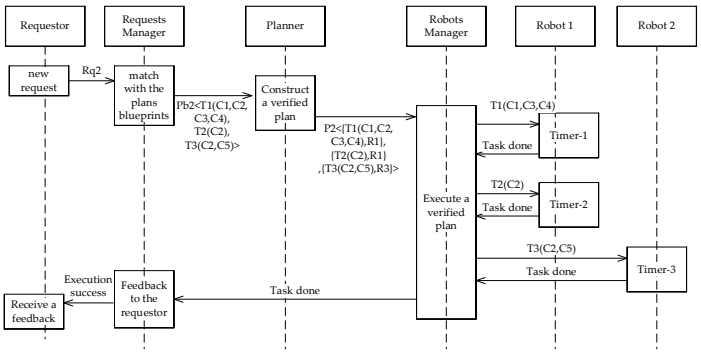


Figure 13: Simulation scenario plan execution sequence diagram – an example

Ultimately, the PLN sends the Pv in form of the ACL-message shown in Figure 14-b to the RbM. The RbM assigns the tasks to the associated robots according to the Pv. The task assignment is sent as an ACL-message as shown in Figure 14-c. The RbM waits the robots' feedback within a timeframe window. If all the RbM received success feedbacks for all the assigned tasks, it sends a plan success feedback to the RqM.

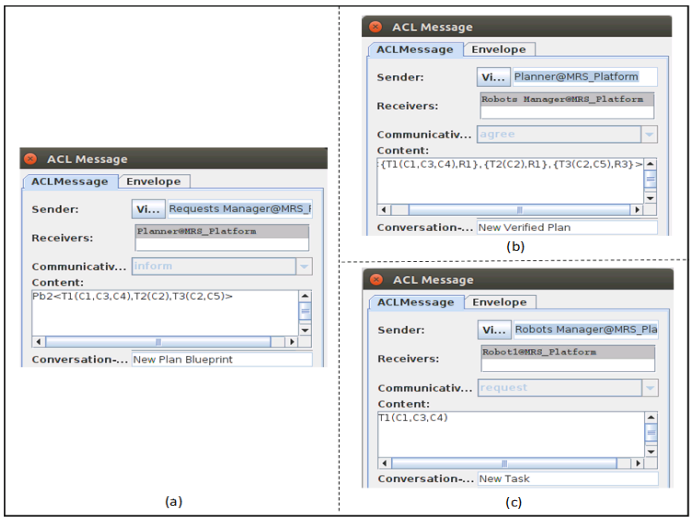


Figure 14: (a) Plan blueprint message – (b) Verified Plan message – (c) Assigned task message

7. Simulation results analysis

As it has been demonstrated in the previous section, the robots' availability, the robot's capabilities, and the the plan blueprints are the variables that can be used to build different simulation scenarios. Accordingly, to measure the system performance, the robots' availability was randomly altered during the run time. Thus, analyzing the simulation results has been done by running JADE MAS for 30 minutes as shown in Figure 15, then measuring the system performance indicators that are concluded in section 4. Each one minute, a new request is generated, one robot randomly unregister from JADE MAS, and one random robot register to JADE MAS. the robot's capabilities and the the plan blueprints do not change during the simulation scenario.

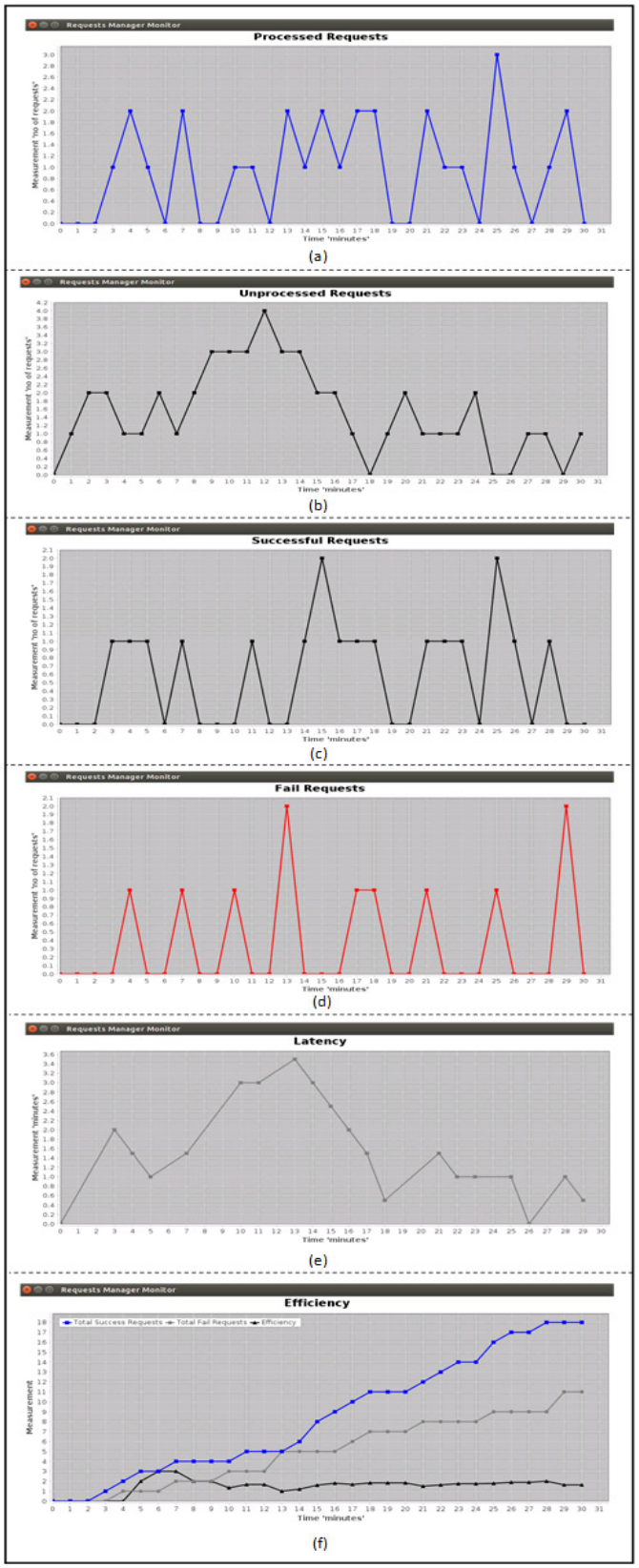


Figure 15: Simulation graphs (a) Processed requests – (b) Unprocessed request – (c) Successful requests – (d) Failed requests – (e) Latency – (d) Efficiency

One of the RqM responsibilities is to monitor the requests status. The number of processed requests by the RqM is shown in the graph in Figure 15-a. Accordingly, the MRS throughput can be directly calculated from this chart. On the one hand, MRS throughput expresses how fast the system, therefore it is a relative value. Thus, to understand the MRS throughput, Figure 15-c and Figure 15-d should be considered as well. For instance, the number of requests at minute 4 is two requests as can be seen in Figure 15-a. But, if we look closely into Figure 15-c and Figure 15-d, we will find out that one request is success and another fail. This means that, it is not important if the system is so fast, but most of the requests are failed to be executed. On the other hand, MRS latency expresses how much delay in the system as it can be seen in Figure 15-e. If the system delay value is equal to zero as can be seen in the 26th minutes of Figure 15-e, this means that the number of unprocessed requests is equal to zero as well, as can be seen in the 26th minutes of Figure 15-b.

The MRS efficiency graph shown in Figure 15-f is derived from dividing the data in Figure 15-c (successful requests) by the data Figure 15-d (fail requests). The MRS efficiency value is absolute. When the MRS efficiency is higher than one, this means that the number of success requests is higher than the number of fail request. Figure 15-f shows that the simulated MRS efficiency is higher than or equal to one during the simulation runtime.

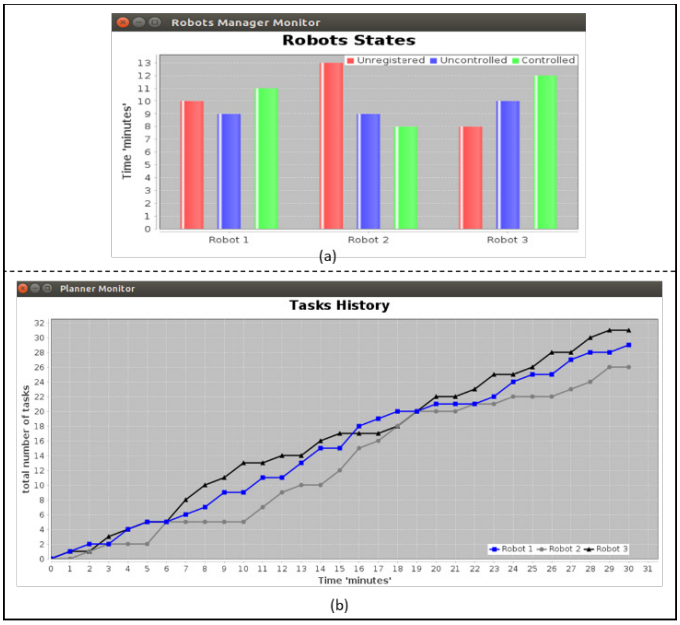


Figure 16: (a) Robots States – (b) Robots tasks history

Table 2: Robots availability, utilization, and effectiveness

	Robot 1 (R1)	Robot 2 (R2)	Robot 3 (R3)
Availability $\frac{T_r}{T_{ov}}$	$\frac{20\ min}{30\ min} = 0.67$	$\frac{17\ min}{30\ min} = 0.57$	$\frac{22\ min}{30\ min} = 0.73$
Utilization $\frac{T_c}{T_{ov}}$	$\frac{11\ min}{30\ min} = 0.37$	$\frac{8\ min}{30\ min} = 0.27$	$\frac{12\ min}{30\ min} = 0.4$
Effectiveness $\frac{T_c}{T_{unc}}$	$\frac{11\ min}{10\ min} = 1.1$	$\frac{8\ min}{9\ min} = 0.89$	$\frac{12\ min}{10\ min} = 1.2$

One of the PLN responsibilities is to monitor the balance the tasks among the available robots. Figure 16-a shows that the robots' available is changing over the simulation runtime. Simultaneously, Figure 16-b shows that the PLN compensates this variation by balancing the MRS. For instance, the task distribution among the available robot is converging to be 6 tasks per robot at the 6th minute of the simulation. Then, the robots' tasks distribution is diverging till it balanced again to be 20 tasks per robot at the 19th minute of the simulation. Table 2 can be also concluded from Figure 16-a. In this table, R_3 is the most utilized and available robot during the simulation runtime, and hence R_3 is the most effective in comparison to R_1 and R_2. Accordingly, the PLN compensates this variation by maximizing R_1 and R_2 task assignment, to balance them with R_3.

8. Summary and Discussion

This article has highlighted new dimensions of the MRS design problem, which are the formalization, simulation, and evaluation of the solution architecture. The proposed modeling approach is based on a formal generic ADLs, that can be used to transfer the solution concept over different system case studies, regardless the implementation technology. Furthermore, the illustrated simulation method can be used to verify different architecture design patterns, based on the concluded system performance measurements.

The fundamental SysML diagrams have been implemented to design the proposed MRS system model. Moreover, BPMN language has been used to implement the activity diagram as it extends UML/SysML notations, semantics, and syntax. The collection of these standard models is used as the MRS blueprints. Those blueprints can be easily coded in any programming environment that supports distributed system implementation. For instance, JADE has been used in this research to implement these blueprints, however Robot Operation System (ROS) or Web Service (WS) are very suitable candidates to deploy the system.

A group of MRS performance requirements have been defined during this article, to quantify the system performance during the simulation runtime. Those criteria can are technology agonistic as well, which means that they can be used to compare between the system performance when it is implemented with different technologies. Furthermore, the system simulation is not only used during the design phased, but it can be reused in a form of a real time digital twin during the implementation phase. For instance, to check in advance different planning and scheduling algorithms before executing them on the real system.

Using a formal description language such as SysML or BPMN enables separating the model from the code, which is a common domain specific programming method. Therefore, in the future work, we will write a code generator that can be used to automatically generate the implementation code. Therefore, the model that has been developed in this article will turn to be executable and will be used as the main software artifact of the project. This can dramatically reduce the coding time and effort and improve the system readability and maintainability. Additionally, in the future work, the same performance measurements that have been used in this article can be used in the implementation phase, as a part of the system visualization.

References

[1] A.R. Sadik, C. Goerick, M. Muehlig, "Modeling and Simulation of a Multi-Robot System Architecture," in Proceedings of the 2019 International

Conference on Mechatronics, Robotics and Systems Engineering, MoRSE 2019, 2019, doi:10.1109/MoRSE48060.2019.8998662.
[2] B. Sendhoff, H. Wersing, "Cooperative Intelligence-A Humane Perspective," in 2020 IEEE International Conference on Human-Machine Systems (ICHMS), 1–6, 2020.
[3] L.E. Parker, "Current research in multirobot systems," Artificial Life and Robotics, **7**(1), 1–5, 2003.
[4] R. Alami, "Multi-robot Cooperation : Architectures and Paradigms," in Journées nationales de la recherche en robotique, 2005.
[5] I. Jawhar, N. Mohamed, J. Wu, J. Al-Jaroodi, "Networking of multi-robot systems: Architectures and requirements," Journal of Sensor and Actuator Networks, **7**(4), 2018, doi:10.3390/jsan7040052.
[6] D.D. Walden, G.J. Roedler, K. Forsberg, "INCOSE Systems Engineering Handbook Version 4: Updating the Reference for Practitioners," INCOSE International Symposium, **25**(1), 2015, doi:10.1002/j.2334-5837.2015.00089.x.
[7] L. Bass, P. Clements, R. Kazman, "Software Architecture in Practice (3rd Edition)," Architecture, 2012.
[8] R. Tesoriero Tvedt, P. Costa, M. Lindvall, Evaluating Software Architectures, Advances in Computers, **61**(C), 2004, doi:10.1016/S0065-2458(03)61001-6.
[9] J. Holt, S. Perry, SysML for systems engineering: 2nd edition: A model-based approach, 2013, doi:10.1049/PBPC010E.
[10] J.E. Perez-Martinez, A. Sierra-Alonso, "UML 1.4 versus UML 2.0 as languages to describe software architectures," Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 3047, 2004, doi:10.1007/978-3-540-24769-2_7.
[11] S. Friedenthal, A. Moore, R. Steiner, A Practical Guide to SysML, 2012, doi:10.1016/C2010-0-66331-0.
[12] I. Malavolta, P. Lago, H. Muccini, P. Pelliccione, A. Tang, "What industry needs from architectural languages: A survey," IEEE Transactions on Software Engineering, 39(6), 2013, doi:10.1109/TSE.2012.74.
[13] R. Eshuis, "Symbolic model checking of UML activity diagrams," ACM Transactions on Software Engineering and Methodology, 15(1), 2006, doi:10.1145/1125808.1125809.
[14] R. Petrasch, R. Hentschke, "Process modeling for industry 4.0 applications: Towards an industry 4.0 process modeling language and method," in 2016 13th International Joint Conference on Computer Science and Software Engineering, JCSSE 2016, 2016, doi:10.1109/JCSSE.2016.7748885.
[15] S. Zor, D. Schumm, F. Leymann, "A Proposal of BPMN Extensions for the Manufacturing Domain," Proceedings of the 44th CIRP Conference on Manufacturing Systems (ICMS 2011); Madison, Wisconsin, June 1-3, 2011., 2011.
[16] N.R. Jennings, M.J. Wooldridge, Agent Technology: Foundations, Applications and Markets, 1998.
[17] S. Kumar, U. Kumar, Java Agent Development Framework, International Journal of Research, **1**(9), 2014.
[18] A.R. Sadik, A. Taramov, B. Urban, "Optimization of tasks scheduling in cooperative robotics manufacturing via Johnson's algorithm case-study: One collaborative robot in cooperation with two workers," in Proceedings - 2017 IEEE Conference on Systems, Process and Control, ICSPC 2017, 2017, doi:10.1109/SPC.2017.8313018.
[19] M. Hoffmann, "Analysis of the current state of enterprise architecture evaluation methods and practices," in ECIME 2007: European Conference on Information Management and Evaluation, 2007.
[20] I. Sommerville, Software engineering (10th edition), 2016.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 184-188 (2021)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Comparison of Learning Style for Engineering and Non-Engineering Students

Mimi Mohaffyza[1,*], Jailani Md Yunos[1], Yee Mei Heong[1], Junita[1], Fahmi Rizal[2], Badaruddin Ibrahim[1]

[1]*Faculty of Technical and Vocational Education, Universiti Tun Hussein Onn Malaysia, Parit Raja Batu Pahat, 86400, Malaysia*

[2]*Fakultas Teknik, Universitas Negeri Paandg, Jln Prof Hamka Air Tawar Paandg, Sumatera Barat Paandg, 25131, Indonesia*

ARTICLE INFO

Article history:
Received: 31 December, 2020
Accepted: 05 April, 2021
Online: 20 July, 2021

Keywords:
Learning styles
Accommodator
Converger
Diverger

ABSTRACT

Educators should be considered the learning style of students so that the best practice approach can be applied in learning activities. As students understand their learning style, they will be able to integrate it into their learning process. Kolb Learning Style was the learning style that was widely used based on the theory of learning experiences. Therefore, this study aimed to describe engineering and non-engineering students' learning style. The survey research design with a quantitative approach was applied in this study. A total of 300 respondents were selected randomly from all faculties in Universiti Tun Hussein Onn Malaysia. The survey questionnaire consisted of two main sections representing Learning Goals, Learning Style, and Learning Activities. The result explains that both engineering and non-engineering students are more dominant to adopt the Accommodator learning style, followed by the Converger learning style, and then Assimilator learning style and Diverger learning style. It is concluded that the engineering and non-engineering students are more incline to be a kinesthetic learner. These learning preferences and learning styles will contribute to their engagement in the concept of learning and for educators to plan teaching strategies.

1. Introduction

Learning about students' learning styles can be very beneficial for both teachers and students. Involving students in the active learning phase necessitates recognizing and comprehending learners' learning styles and teachers' teaching styles. Types of learning play a considerable role in learners' lives. Students may incorporate their learning style into their learning process as they become more aware of it. Students learn in various ways, and teachers must design their courses according to different types of learning. Learning skills, creativity and life and career skills are evidence that students master the process of capability and development, integration and knowledge assessment from different subjects and sources of understanding [1]. Identifying students' learning styles will help educators plan their teaching methods and activities effectively to achieve their learning outcome [2]. The learning style of students is f the supporting forms of active learning [3]. The style of learning plays an important role in ensuring that the learning process is performed effectively.

*Corresponding Author: Mimi IMohaffyza, mimi@uthm.edu.my

Students should have 21st-century skills, especially soft skills, to enhance their employability and values. [4]. Universities must make vital elements of education to conduct learning by introducing effective student learning processes in the growth of international education in the formulation of skills in the twenty-first century. [5]. To ensure that all students receive knowledge from the learning process, educators must observe and consider the discrepancies and similarities between students and use the knowledge to prepare for the learning process [6] to design learning regardless of the learning style of the students [7]. To compare learning style preferences between engineering and non-engineering students in Malaysia, this study used a measurement method called the Kolb Learning Style Inventory (LSI) since LSI is able to provide a simple validation of the Experiential Learning Theory.

1.1. Kolb Learning Style

The Experiential Learning Theory of Kolb forms the basis of the paradigm of Kolb's learning style. Experiential learning, which is distinct from other cognitive learning theories, notes the increase in learning process interactions [8]. The Kolb Learning Style Inventory (LSI) is one method for measuring learners'

preferential teaching style. Kolb's learning style, or more generally known as Experiential Learning Theory (ELT), describes learning as a process in which information is created by transforming experience [9]. Learning is a process, according to Kolb, and knowledge is the transformation of experience [10]. Kolb also indicates that, to have a complete learning experience, students must go through all four phases of the learning cycle, as depicted in Figure 1. These four stages not only allow students to explore a subject through various activities and viewpoints thoroughly, but they also accommodate different learning styles.

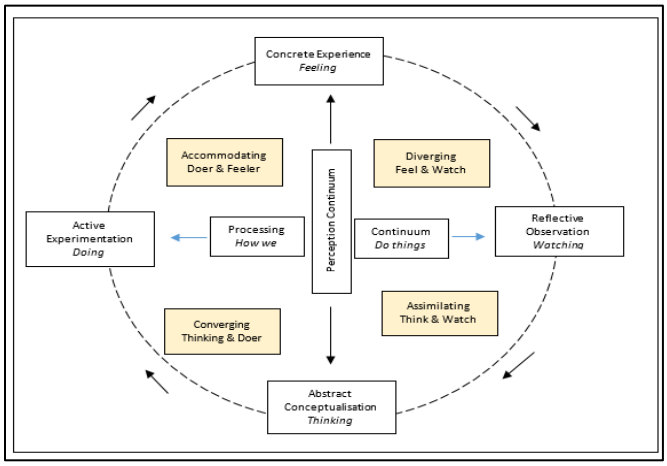


Figure 1: The Learning Model of Experiential Learning at Kolb's Learning Styles

In the Kolb view, learning styles refer to processes in which the person organizes the ideas, rules and principles that address them in dealing with new situations. In practice, one of the most powerful methods in the learning analysis of the individual is the theory of the Kolb learning style. The learning styles as a collection of values, interests, and habits that people attempt to learn about a particular situation by using it [11]. Previous research put it another way: the learner first conducts an action (concrete experience), then tries to think about it (reflective observation), then develops a hypothesis (abstract conceptualization), and finally attempts to exempt it (active experimentation) [12]. According to Kolb, experiential learning can be used in both engineering and non-engineering educational environments [13]. It enables students to participate actively in the learning process to develop awareness, skills, values, and attitudes through direct experience. The learning stages will promote knowledge transfer by providing direct practice tailored to student expertise's scope [14]. This learning method enables students to create their awareness and experience and the acquisition of new skills and knowledge. It stresses that learners learn to use their expertise and experience to solve their problems. This study aimed to compare the preferential learning styles of engineering and non-engineering students in Malaysia by using a Kolb Learning Style Inventory (LSI) model as a reference model because it can provide a basic foundation for validating the theory of experiential learning.

2. Learning Style in Technical and Vocational Education

Technical and Vocational Education is an important road to vocational education and the growth of skills. To meet Malaysia's economic needs, the country's TVET enrollment must be increased by 2.5 times by 2025. Transformation Programme [15]. The human resources to meet this demand, however are inadequate. Right at the time. Moreover, TVET is regarded to be less appealing than traditional university education. This has led to a shortage of, especially highly skilled, TVET students. Malaysia must therefore move from the commonly accepted assumption that the only career path for Malaysian youth is traditional university education, and also emphasize TVET as a valid higher education choice.

Technical and Vocational education students are exposed to an educational system aimed at getting a job. (1) A component of an educational activity aimed at providing the necessary knowledge and skills to carry out a specific job, occupation or professional activity in the labour market can be technical and vocational education. At the same time, other types of education, by training people not only as workers but also as citizens, act as an additional form; (2) an activity associated with the technology transition, innovation, and growth processes Knowledge and skills must be transferred since they form the foundation of technical progress and growth [16]. In technical and vocational teaching, as in many fields of knowledge, it is important to identify and understand students differences to adopt the institute's needs to best suit the students' learning conditions and skills. A fact in the classroom, which can be seen in actual scenarios or in virtual techniques, is the need to adapt teaching methods to student learning styles and interests.

If learning styles are not identified, they may influence the teaching and learning process [17]. A lack of knowledge of the modes of learning can also be problematic. In the implementation among students of the acceptable and successful learning styles [18]. Academic success will be impaired as a result [19]. Unfortunately, teacher-centred learning sessions are held by most educators, allowing fewer students to engage in the process and activities of learning [17]. Therefore, for the performance of students, learning style is an important matter. The style of learning will ensure that a learner learns well [19]. Students need to identify their learning styles to build on their learning skills and expand their learning skills. By posing a challenge or using various education methods, educators are also expected to encourage their students to identify their learning style. [20].

3. Material and Method

The survey research design with a quantitative approach was applied in this research. A set of questions was designed based on the collected learning style and activities found in literature based on the Kolb Learning Style Inventory. A total of 300 respondents were randomly selected from all faculties in Universiti Tun Hussein Onn Malaysia, UTHM (i.e. Faculty of Civil Engineering and Built Environment, Faculty of Technology Management and Business, Faculty of Technical and Vocational Education, Faculty of Electrical and Electronic Engineering, Faculty of Computer Science and Information Technology, Faculty of Applied Sciences and Technology and Faculty of Engineering Technology). The survey questionnaire consisted of two main sections representing the Learning Goals, Learning Style and Learning Activities. This questionnaire was deployed online from the university's online forum and platform. Respondents

were able to complete the questionnaire in approximately 10-15 minutes.

4. Finding and Discussion

The findings discussed are based on the data of the Learning Goals, Learning Style and Learning Activities items that were constructed. Data that had been collected were used to analyze in the context of Learning Style characteristics, and T-test was conducted to determine whether there are any variations between the two groups of fields, as well as descriptive statistics such as frequency and percentage, to evaluate and interpret the results in this report. The interpretation in the research instrument was used to explain the frequencies and percentages. The agreement level was used to assess the students' perceptions in both areas, which were either Yes or No.

4.1. The Learning Style Between Engineering and Non-Engineering Students (Descriptive Results)

Based on a survey conducted, the different learning styles of engineering and non-engineering students were gathered and divided into four forms of learning style defined by Kolb, following the learning style Diverger, Assimilator, Converger and Accommodator. To better understand both of these learning styles, it should be understood that the Assimilator learning style (think and watch) is a variation of Reflective Observation (RO) and Abstract Conceptualization (AC).Converger learning (think and do) is a synthesis of Abstract Conceptualization (AC) and Active Exploration (AE) (AE). Accommodation learning style (feel and do) is a combination of Active Experimentation (AE) and Concrete Experience (CE) and Diverger learning style (feel and watch) is a combination of Concrete Experience (CE) and Reflective Observation (RO) [21]. The percentage of students' data distribution on each Kolb learning style determined by The Kolb Learning Style Inventory is shown in Table 1 and illustrated in Figure 2 below.

Table 1: Results of Learning Style in Vocational Education between Malaysian engineering and non-engineering students

Field	Kolb Learning Style									
	Converger		Assimilator		Accomodator		Diverger		Total	
	f	%	*f*	%	*f*	%	*f*	%	*f*	%
Engineering	42	28	29	19.3	51	34	28	18.7	150	100
Non-Engineering	51	34	21	14	60	40	18	12	150	100

The results show that Accommodator learning style in engineering students is the highest percentage than others learning style with value (*f* = 51, 34%) followed by Converger (*f* = 42, 28%) and Assimilator (*f* = 29, 19.3%). While Diverger learning styles shows the lowest percentage within engineering students with values (*f* = 28, 18.7%). Other than that, a similar condition was shown by non-engineering students where the Accommodator learning style shows the highest worth percentage (*f* = 60, 40%) followed by Converger (*f* = 51, 34%) and Assimilator (*f* = 21, 14%). While the lowest value of percentage is Diverger which is (*f* = 18, 12%).

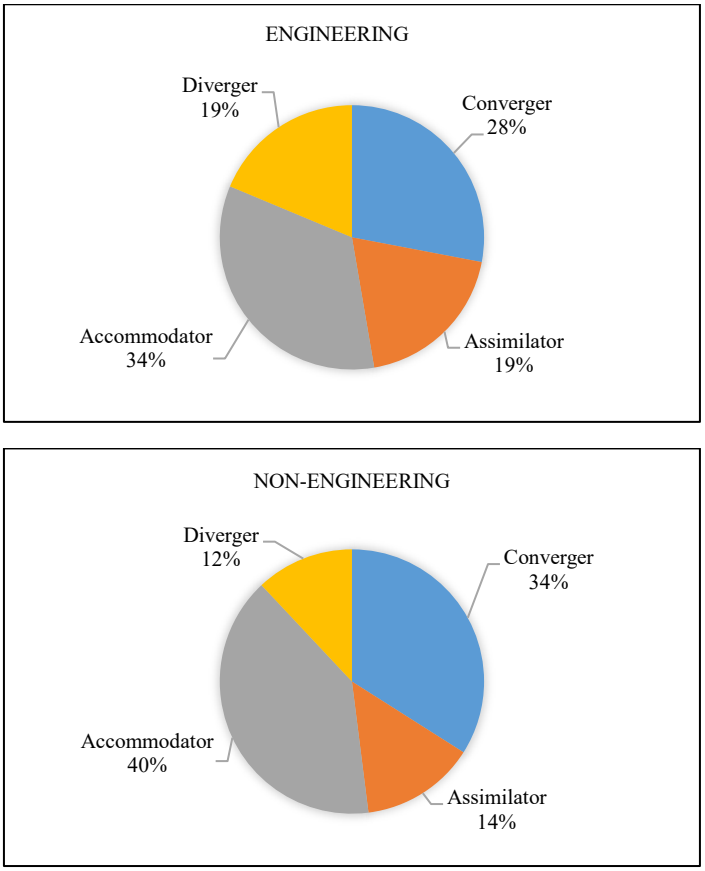


Figure 2: Results of Learning Style in Vocational Education between Malaysian engineering and non-engineering students

The findings of the study can be seen more clearly by referring to figure 3 below where you can see the significant differences between the four learning styles. Although there is a percentage difference between the two fields, it shows that most of the engineering and non-engineering students can be described as an accommodator, which indicates they are most potent in Concrete Experience and Active Experimentation. Instead of logic, they rely on intuition which prefers learning from personal experience, relies on given knowledge rather than carrying out his/her research, requires a clear explanation before starting work [22]. It also shows that both engineering and non-engineering students have strengths that lie in their desire to execute plans and tasks to take part in new events [23]. This result is in line with the Kolb Learning Styles trend, which states that students who use the Doer and Feeler learning styles are best suited for teaching, technician, and engineering jobs and have a background in education, technical studies, and engineering [24].

Other than that, the overall result shows Converger is the second-highest percentage for both fields. In contrast to engineering students, non-engineering students prefer technical tasks and better interpret complex concepts and hypotheses. They also enjoy experimenting. This type of learning style's strengths

lie in their ability to set goals, solve problems, and make decisions [23].

Apart from that, Assimilator shows the third-highest percentage for both engineering and non-engineering fields. The results show that engineering students have a higher percentage than non-engineering students. This means that engineering student who learns in this style has a wide range of knowledge and arrange it in the most logical way [22] compared to non-engineering students. It also indicates that these students prefer rational, factual, and well-thought-through knowledge [24]. The strengths of this learning style lie in their ability to schedule, coordinate, evaluate and engage in inductive reasoning systematically. The results of this study are confirmed by a study conducted by [25] in which engineering students need more diverse knowledge gathered from different sources since they must observe how to execute the task before beginning to perform it. The knowledge is presented from different angles and concluded in a logical, simple, and concise manner. Finally, the type of learning that shows the fourth-highest percentage for both engineering and non-engineering is Diverger. The findings showed that there was a higher proportion of engineering students compare to non-engineering students. It indicates that engineering students with a particular style of learning observe a situation and then look at the situation later from multiple viewpoints, learning from each one [22]. Besides, it also shows that engineering students have more effective at seeing a particular situation from different perspectives than non-engineering students [26].

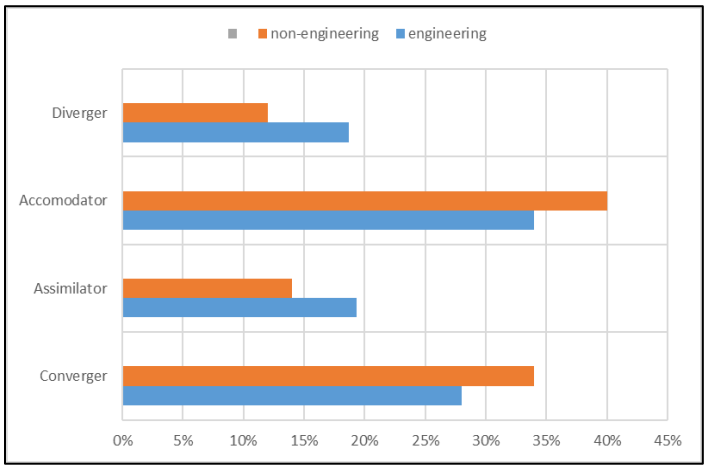


Figure 3:Comparison of Learning Style in Vocational Education between Malaysian engineering and non-engineering students

Because of the interest in designing the learning process, educators have to consider the style of learning of different students. Other than that, to maximize learning effectiveness, the learning method that relates to each learning style is more important because the learning method has a learning style related to it [27]. The difference between the way information was obtained and interpreted was more related to the style of learning that students had. One of the key reasons for gathering learning efficacy is the type of learning [21]. The suggest that knowledge of the learning styles of learners can be important for curriculum and teaching improvement. Similarly, [28] state that If the learning styles of students are evaluated, it is possible to systematically plan learning activities that further strengthen strengths or develop weaker phases to maximize thinking and problem-solving skills."

4.2. Comparison of Learning Style Between Engineering and Non-Engineering Students (Inferential Results)

As for the comparison between Engineering and Non-Engineering students, the inference analysis had shown a non-significant value between both groups in practising learning style in their learning process with mean and significant value (Engineering = 0.538, Non-Engineering = 0.562, p=0.543). Although there is a difference in percentage and frequency values, the inference value indicates no significant value for engineering and non-engineering students learning style. This shows that both groups of students have approximately the same learning style between engineering and non-engineering students.

Table 2: The Differences of Learning Style Between Engineering and Non-Engineering Students

Field	Mean	Std Deviation	Significant
Engineering	0.538	0.133	0.543
Non-Engineering	0.562	0.130	

5. Conclusion

The study showed that learning styles are necessary for a course to achieve total value from learning. While sharing certain characteristics, may show major differences in other aspects that affect learning. Educators who are mindful of these differences and can articulate these characteristics have a better chance of creating good instruction for a wide range of learners. In knowing their strengths and interests and utilizing the learning cycle, all learning styles will become stronger for students exposed to learning style models and Kolb's Learning Style Inventory (LSI), which will enable them to become more active learners. This research can be very beneficial for educators who want to increase the effectiveness of the learning process. Recognizing and reacting to individual learning styles may improve students' ability to accept and retrain content and help to avoid possible learning difficulties by selecting the appropriate teaching method. This may also aid in selecting the most suitable materials and activities for the individual students.

Acknowledgement

We would like to thank the team of Project Matching Grant K135 who participated in this study consisting of experts from two universities, Tun Hussein Onn University, namely Maizam Alias, Tee Tze Kiong, Lee Ming Foong and Faizal Amin Nur Yunus and Universitas Negeri Paandg, namely Ganefri, Nizwardi Jalinus, Syahril, Sukardi, Risfendra and Rahmat Azis Nabawi for their contribution. Finally, we would like to thank the Tun Hussein Onn University of Malaysia for the financial support under UTHM Grant Vot U940.

References

[1] M.C. Sahin, "Instructional design principles for 21 st century learning skills," **1**(1), 1464–1468, 2009, https://doi.org/10.1016/j.sbspro.2009.01.258

[2] S.Kalbasi, M.Naseri, G.H.Sharifzadeh, A.Poursafar, "Medical students' learning styles in Birjand University of medical sciences strides in development of medical education," Journal of Medical Education Development Center of Kerman University of Medical Sciences, 2008, **1**(5): 10-16. doi: 10.5681/rdme.2013.017

[3] Kolb, Y,Alice, & D.A. Kolb, "The Kolb Learning Style Inventory — Version 3," 2005, Technical Specifications. Experience Based Learning Systems, Inc.

[4] N. Azid, R. Rawian, Shaik-Abdullah & T.K. Tee, " The Development of Interactive Case-Based Smart Thinking and Industrial Problem-Solving Stimulator to Enhance TVET students' Thinking Skills," Journal of Engineering Science and Technology, 2019, **14**(5), 2643-2656.

[5] A. Halstead & L. Martin, "Learning styles: A tool for selecting students for group work," International Journal of Electrical Engineering Education, **39**(3), 245–2522, 2002. https://doi.org/10.7227/ijeee.39.3.8

[6] R. Laura, "What is differentiated instruction? New York: Rowan University,"2007. https://doi.org/10.1093/ww/9780199540884.013.u289492

[7] C. Weselby, "What is differentiated instruction? Examples of how differentiated instruction in the classroom," Oregon: Concodia University, 2017. https://doi.org/10.4135/9781483346243.n103

[8] S. Cassidy, "Learning styles: An overview of theories, models, and measures," Educational Psychology, 2004. https://doi.org/10.1080/0144341042000228834

[9] Y.M. Yousafzai, N.Baseer, S.Fatima, A.Ali & I.Shah, I. "Investigating the Relationship between Learning Styles and ESP Reading Strategies in Academic Setting," International Journal of Applied Linguistics & English Literature, **7**(3). 11-115, 2018. https://doi.org/10.7575/aiac.ijalel.v.7n.3p.156

[10] C.Babadoğan, C. "Öğrenme Stilleriyle İlgili Araştırmaların Taranması (Survey of Researches Related to Learning Styles) Eğitim Bilimleri Dergisi," Ankara Üniversitesi, **24**(2), 603-606, 1991.

[11] T. De Jong, "Cognitive load theory, educational research, and instructional design: some food for thought," Instructional science. **38**(2): 34105, 2010. https://doi.org/10.1007/s11251-009-9110-0

[12] A. Z. Moghadam, M.M. Fard "Surveying the Effect of Metacognitive Education on the on the Mathematics Achievement of 1st Grade High Junior School Female Students in Educational District 5, Tehran City, 2009-10 Educational Year, Procedia-Social and Behavioral Sciences, **29**(15):31-40, 2011. https://doi.org/10.1016/j.sbspro.2011.11.394

[13] C. Tseng, "Connecting Self-directed Learning with Entrepreneurial Learning to Entrepreneurial Performance," International Journal of Entrepreneurial Behavior and Research, **19**(4), 425- 446, 2013. https://doi.org/10.1108/ijebr-08-2011-0086

[14] Dumiyati, "Pendekatan Experiential learning dalam Perkuliahan Kewirausahaan di Perguruan Tinggi untuk Menghadapi Asean Economic Community [Suatu Kajian Teoretis]," Prosiding Seminar Nasional Profesionalisme Pendidik dalam Dinamika Kurikulum Pendidikan di Indonesia pada Era MEA, 87- 97, 2015

[15] Ministry of Education Executive Summary Malaysia Education Blueprint 2015-2025, 2015.(Higher Education).

[16] O.O.Oskay, E.Erdem, B.Akkoyunlu, A.Yilmaz, "Prospective chemistry teachers' learning styles and learning preferences," Procedia Social and Behavioral Sciences, **2**, 1362-1367, 2010. https://doi.org/10.1016/j.sbspro.2010.03.201

[17] M.H. Yee, J. Yunos, W.Othman, R. Hassan, T.K.Tee, & M. Mohamad, "Disparity of Learning Styles and Higher Order Thinking Skills among Technical Students," Procedia - Social and Behavioral Sciences, **204**(2014), 143–152, 2015. https://doi.org/10.1016/j.sbspro.2015.08.127

[18] H.Pashler, M.McAndiel, D.Rohrer, & R.Bjork, "Learning styles: Concepts and evidence," Psychological Science in the Public Interest, 2009.https://doi.org/10.1111/j.1539-6053.2009.01038.x

[19] I.J. McCoog, "21st Century teaching and learning. Education Resource Center," 2018.

[20] E. Marin, " Experiential learning: empowering students to take control of their learning by engaging them in an interactive course simulation environment," Procedia - Social and Behavioral Sciences **180**, 854 – 859, 2015. https://doi.org/10.1016/j.sbspro.2015.02.224

[21] A. Y. Kolb & D.A. Kolb, " Experimental Learning Theory Bibliography," Cleveland: OH : Experience Based Learning System Inc, 2003 https://doi.org/10.1007/978-1-4419-1428-6_227

[22] M. MiMuro, A, Terry, "A matter of style: Applying Kolb's learning style model to college mathematics teaching practices," Journal of College Reading and Learning, **38** (1), 53-60, 2007. https://doi.org/10.1080/10790195.2007.10850204

[23] E.F. Turesky & D. Gallagher, " thyself: Coaching for leadership using Kolb's experiential learning theory," Coaching Psychologist, **7**(1), 5–14, 2015.

[24] D.A. Kolb, "Learning Style Inventory: Self Scoring Inventory and Interpretation Booklet," McBer and Company, Boston, 2005. https://doi.org/10.1007/978-3-8350-9212-9_4

[25] M.N. Ghufron & R. Risnawita, "Gaya Belajar: Kajian Teoritik," Yogyakarta: Pustaka Pelajar.

[26] S. Nasution, "Berbagai Pendekatan dalam Proses Belajar and Mengajar," Jakarta: Bumi Aksara. 2009.

[27] Ö. Şimşek, "Marmara Öğrenme Stilleri Ölçeği"nin Geliştirilmesi Ve 9–11 Yaş Çocuklarının Öğrenme Stillerinin İncelenmesi," Marmara Üniversitesi Eğitim Bilimler Enstitüsü. İlköğretim Anabilim Dalı. Yayınlanmamış Doktora Tezi. 2019. İstanbul

[28] J.A.Gyeong & S.Y. Myung, " Critical thinking and learning styles of nursing students at the baccalaureate nursing program in Korea," Contemporary Nurse, **29**(1), 100–109, 2008. https://doi.org/10.5172/conu.673.29.1.100

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 189-195 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Mitigation of Nitrous Oxide Emission for Green Growth: An Empirical Approach using ARDL

Hanan Naser*, Fatema Alaali

College of Business and Management, American University of Bahrain, Riffa, 959, Kingdom of Bahrain

ARTICLE INFO

Article history:
Received: 02 March, 2021
Accepted: 01 July, 2021
Online: 20 July, 2021

Keywords:
N_2O emissions
Environmental Kuznets Curve
FDI
Economic Growth
Financial Development
Electric Power Consumption

ABSTRACT

Although the perception of Environmental Kuznets Curve (EKC) has been thoroughly investigated, but there is inconsistency in the results. The relationship between nitrous oxide (N_2O) emissions, financial development, economic growth, foreign direct investment, and electric power consumption in Bahrain over the period 1980 – 2012 is examined in this paper. The autoregressive distributed lags (ARDL) technique is employed to test for the cointegration in the long run. The results reveal a reversed U-shape long run relationship between N_2O emissions and economic growth for Bahrain. Moreover, electric power consumption affects N_2O emissions positively in the short and long run. Whereas foreign direct investments and financial development affects the emissions of N_2O negatively. Therefore, Bahrain should assist households in installing solar cells to generate clean energy and enhance its financial sector.

1. Introduction

One of the most considerable problems that the whole world is experiencing is the environmental degradation, which leads to irreparable damages to the natural world and human society. Greenhouse gas emissions are the main pollutants of the environment, and the major source of these pollutants is burning fossil fuels. When greenhouse gas emissions are mentioned, it is usually referring to carbon dioxide (CO_2). However, greenhouse gas story involves other types of gases. Nitrous oxide (N_2O) is partially responsible for the greenhouse gas diffusions and its effect on the atmosphere from the point of global warming is greater than that of CO_2. One ton of N_2O corresponds to almost 298 tons of CO_2.[1] Moreover, the procedure followed to eliminate the N_2O from the atmosphere contributes into depleting ozone layer. Therefore, N_2O is considered as an ozone eradicator as well as being greenhouse gas.

Bahrain that is an archipelago made up of 33 islands has a total area of 780 km^2. The current population in the year 2019 is about 1.6 million with a population density of 2155 per km^2 and a population growth of 4.9% in year 2018.There is an increase in total energy consumption by 2% per year on average over the period 2010 to 2017 with a record of 14.5 mega ton of oil equivalent (Mtoe) in 2017. The fundamental factor behind this increase in energy consumption is the increase in population. Rise in population means greater local demand for energy, which increases the greenhouse gas emissions including the N_2O emissions.

One of the most substantial origins of N_2O emissions is burning fossil fuels for energy and transport. Compared to the quantity of CO_2 in the air, the amount of N_2O is much less, but the warming impact of each molecule of N_2O is nearly 300 times that of Co_2.[2] Accordingly, a number of papers have examined the soundness of the Environmental Kuznets curve (EKC) hypothesis using N_2O diffusions as a measure for the environmental deterioration. In [1], the authors scrutinized the theoretical and empirical basis of the EKC using a panel of 156 countries and three different pollutants including N_2O emissions. The results of their paper indicate the presence of a reversed U-curve for all the three pollutants but at different levels of income. In [2], the researchers reviewed the available literature to explain the EKC phenomenon for atmospheric variation and decide the association with the economic evolution of Bangladesh in regard to the EKC.

*Corresponding Author: Hanan Naser, E-mail: hanan.naser@aubh.edu.bh

[1] http://theconversation.com/meet-n2o-the-greenhouse-gas-300-times-worse-than-co2-35204

[2] https://www.carbonbrief.org/nitrous-oxide-emissions-could-double-by-2050-study-finds

Their results show that the EKC is valid only for developed countries which have low-income turning point. The paper of [3] confirm the presence of the EKC assumption for Germany by employing time series data between 1970 – 2012. The study of [4] employed a panel of 36 developing and developed nations between 1995-2013 and the EKC assumption is confirmed for N_2O and CO emissions. Therefore, the contribution of this study is to examine the short- and long-term effects of GDP growth, foreign direct investment, financial development, and electric power consumption on Bahrain N_2O diffusions over the period 1980 – 2012.

The remainder of the paper is formatted as follows: section 2 highlights the review of literature; the following section describes the data and models employed in this study followed by the discussion of results in section 4. Finally, the last section covers the conclusions and policy implications.

2. Literature Review

EKC affirms that environmental pollutants increase as the national output increases just before a certain threshold of output after which emissions decrease as output increases [5]. Particularly, EKC assumption indicates a reversed U-shape relationship between national output and environmental deterioration. This aspect was first called EKC by [6].

A substantial consideration is given to the relationship between economic development and environmental diffusions in the last decades. Various practical studies have scrutinized this hypothesis in distinct countries all over the world. In general, the results of these studies are mixed. One part validates the EKC hypothesis and approves the presence of a reversed U-shape (for example, [7]-[10] among others). The other part of these studies did not succeed to prove the EKC hypothesis but found either a linear or N-shaped relationship (such as [11]-[14] among others).

The presence of the difference in the results of empirical studies is apparent. This difference in the results can be illustrated by the following components. The first component is explained by the various measures of air pollutants used. For example, one part of the studies utilized air contamination index such as greenhouse gas emissions, CO_2, CH_4, SO_2 and NO_x emissions (such as [11]; [15]; [16]). However, the other group of research has assessed the EKC hypothesis by employing other environmental measures such as ecological footprint [17], deforestation [18] and [19], and hazardous waste [20]. The second component is pertained to the different models and methodologies utilized. To illustrate, the studies that employ a panel data for a set of countries examined the EKC theory using panel cointegration [21] and [22] or fixed effects regression [23]. On the contrary, studies that employ an individual country apply time series techniques such as Vector Autoregressive (VAR) models [24] and [25] cointegration approaches of Granger and Johanson and the ARDL bounds techniques (such as [26]-[30], among others). The third component is associated with the different variables incorporated in the model. Some studies examined the EKC hypothesis by incorporating the GDP per capita and its square only. Other set of research papers have expanded the primary model with other explanatory variables like energy consumption, foreign direct investment, financial development, urbanization and trade openness. The last component is related to the various countries included in the study and the chosen period.

The amount of research that examines the EKC theory in the Gulf Cooperation Council GCC area is limited. For example, in [30], the author uses the data of Saudi Arabia to investigate the long-run relationship between the emissions of CO_2, economic development, energy consumption and urbanization and found that the EKC does not exist. This conclusion is supported by the study of [31], which scrutinized the relationship between CO_2 emissions from transports, energy consumption by road transports and economic development in Saudi Arabia. On the contrary, the study of [32] proved the presence of EKC assumption by studying the effect of trade and income level on CO_2 emissions. In the UAE, the presence of EKC relationship was approved [33] and [34]. The empirical results of [35] indicate that inverted U-shaped relationship is held when using the ecological footprint as a measure of environmental deterioration but not for CO_2 emissions. This paper adds to the existing literature by studying the relationship between N_2O emissions, income level, electricity use, FDI and financial development in Bahrain.

3. Material and Method

3.1. Data

To attain the target of this research paper, yearly N_2O emissions data (thousand metric tons of CO_2 equivalent) are used as the environmental pollutant. To check the reliability of EKC theory, GDP per capita at constant 2010 US$ and its square are employed.[3] As asserted by the study of [17] that one of the factors that leads to the divergence in the results of the EKC testing studies is the variables included in the estimated equation. One of the sources of N_2O emissions is fossil fuel combustion that is used to generate electricity. Therefore, this paper augments its model with electric power consumption (KWh per capita) to examine its impact of N_2O emissions.

In [36] and [37], the authors argue that foreign investors and international corporations prefer investing in countries that have loose environmental policies and standards. Most of these investments sponsor forms of production that are environmentally inefficient [38]. Accordingly, this paper augments the basic model with additional variables such as the foreign direct investment, net inflows (% of GDP). Furthermore, domestic credit provided by financial sector (% of GDP) is utilized as a measure of financial development.

[3] GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 2010 U.S. dollars.

The World Development Indicators (WDI) database is utilized to collect annual data for all the variables between 1980 – 2012.[4] In order to reduce heteroscedasticity and stabilize variances all the variables are transformed using the natural logarithm except the financial development indicator and FDI as they are measured as a percentage of GDP. Table 1 summarizes the descriptive statistics of all the variables under concern.

Table 1: Descriptive Statistics

Variables Description	**N_2O emissions** (Thousand metric tons of CO_2 equivalent)	**GDP per capita** (constant 2010 US$)	**Elec Electric Power Consumption** (KWh per capita)	**Fin Domestic credit provided by Financial Sector** (% of GDP)	**FDI Foreign direct investment, net inflows** (% of GDP)
Mean	86.39	20487.51	16972.75	30.42	4.57
Std. deviation	26.90	1995.28	5003.57	21.68	7.79
Minimum	39.684	16571.4	4612.55	-5.82	-13.61
Maximum	131.257	22955.1	21644.38	72.22	33.57
Skewness	0.142	-0.61	-1.53	0.36	1.39
Kurtosis	1.928	1.89	4.32	2.41	7.79
obs	33	33	33	33	33

3.2. Technical Tool and Econometric Model

On the basis of the pioneering work by [39] that is followed by the study of [40], this paper uses a single equation model that allows to examine the emission –growth relationship for Bahrain. Basically, it is suggested that the degradation of environment in Bahrain can be briefly written as follows:

$$N_2O = f\ (GDP, GDP^2, Elec, Fin, FDI) \quad (1)$$

where N_2O represents the nitrous oxide emissions in Bahrain. The above function shows that the explanatory variables for nitrogen oxide emissions are the economic growth (GDP), square of economic growth (GDP^2), electricity consumption (Elec), financial developments (Fin) and foreign direct investments (FDI).

Looking at the main objective of this study, it is suggested to derive a natural log form equation from the above linear equation, so it allows for testing the hypothesis of the EKC. Having natural log model ensure the production of reliable and effective results. Furthermore, in [41], the authors have reported that a natural log model will help in satisfying the stationary condition of the variance-covariance matrix. Accordingly, re-writing the model above can be represented by the following equation:

$$ln(N_2O)_t = \beta_0 + \beta_1\ ln\, GDP_t + \beta_2\ (ln\, GDP_t)^2 + \beta_3\ ln\, Elec_t + \beta_4\ ln\, Fin_t + \beta_5\ ln\, FDI_t + \varepsilon_t \quad (2)$$

where ε_t is the error term associated with the estimation while β_0 is the intercept. The coefficients $\beta_1, \beta_2\ \beta_3, \beta_4\ and \beta_5$ represent, respectively, the impacts of GDP per capita and its quadratic term, electricity consumption per capita, financial developments, and the percentage of foreign direct investments. Based on the theory of EKC, it is likely that β_1 has positive sign while β_2 has a negative sign.

Having a glance at (2) shown above, it is revealed that none of GDP term or its square term can be excluded. The aim of adding both terms in the same equation is to investigate the authenticity of the EKC, which can be verified only if the relationship among environmental emissions and economic growth is expressed by an inverted U-shape curve. In another words, if the relationship among the environmental emissions and economic growth is an inverted U-shape, EKC hypothesis cannot be rejected. Therefore, the environmental emissions will tend to increase as much as there is an increase in output per capita, until it reaches to a specific level (peak). Accordingly, the quality of the environment will be better.

In a study conducted by [6], it is reported that if there is no significant change in technology in an economy, it is expected that an inverted U-shape curve represents the linkage between GDP per capita and the emissions of the environment. Precisely, an expansion of an economy causes negative impacts on environment at early stages. However, as much as it grows, structural change is experienced by the economy due to many factors including information intensive industries and services, advances in technology, increase in environmental consciousness and implementation of environmental protocols, which may have positive influence on reducing environmental emissions.

3.3. ARDL Approach

The purpose of this study is to investigate the hypothesis of having an environmental Kuznets curve, that be represents a curve of an inverted U-shape to express the linkage between the emissions of environment and the growth of an economy. Accordingly, long- and short-term dynamics of the model can be investigated using a cointegration approach named Autoregressive Distributed Lag Model (ARDL). Among others, the authors of [42], [43], [44], and [45] have used ARDL, which starts as a general model and then moves into more specific one to capture the characteristics of the data included in the regression using a sufficient number of lags. As literature has discussed many advantages for using ARDL in cointegration models, it is worth to shed the light on some key advantages that have affected the selection of the model for this study. First, in [42], the author has reported that the ARDL approach is suitable for variables with different integration orders that could be either fractional or at I (0) or I (1). In addition, the equilibrium features of both short- and long-term dynamics are captured by the error correction model (ECM) which is obtained from a transformation that is done linearly for the ARDL model.

Considering the size of the sample, the authors of [44] have claimed that when the investigation is done on a small sample, then the approach of ARDL is more appropriate in comparison to that of [46] cointegration. They have also admitted that ARDL has minimal residual correlation and thus considered as a superior approach against serial correlation problems attributed to endogeneity issue. Lastly, being able to proceed with the

[4] The most available data for N_2O were till the year 2012.

estimation even if the explanatory variables are endogenous is another key advantage of ARDL model [42], [45].

3.4. Unit Root Testing

As time series data exhibit trending behavior and thus the null hypothesis of non-stationary cannot be rejected, unit root testing is applied using three different approaches called Augmented Dickey & Fuller (ADF) [47], Phillips and Perron (PP) [48] and KPSS [49] to test the order of integration of each variable. In fact, both ADF and PP tests suggest that the null hypothesis is non-stationary where KPSS claims a stationary null hypothesis. Since the original Dickey & Fuller test [47] cannot accommodate complex models, the ADF approach is a customized version of the test that can satisfy the need of having an appropriate stationary test for complex models. The ADF for models with unknown orders can be represented as shown below:

$$\Delta y_t = \theta_0 + \alpha_0 t + \alpha_1 y_{t-1} + \sum_{i=0}^{p} \theta_i \Delta y_{t-1} + \mu_t \quad (3)$$

where y_t is the variable in period t; Δy_{t-1} is the y_{t-1}-y_{t-2}; the disturbance term represented by μ_t is i.i.d and has zero mean where the variance is 1; t the linear time trend and p is the lag order. It is worth to note that the ADF test has been developed to examine univariate time series for the presence of unit root. In another word, any time series with presence of unit root should be treated to ensure that the variable is stationary (if it is required for modelling).

Since $\boldsymbol{\alpha_0}$ and $\boldsymbol{\alpha_1}$ are the coefficients of the time trend term (t) and the variable $\boldsymbol{y_{t-1}}$ in previous period (t-1), it is important to investigate the order of integration of the variable y_t. This is done by examining whether or not $\alpha_1 = 0$ in (3). Accordingly, if α_1 is not significantly less than zero, the null hypothesis of a unit root cannot be rejected. Otherwise, both level and first differenced variables are tested against unit root. Since in [49], the KPSS stationary test was introduced which assumes that stationary is the null hypothesis, this study has also applied this test to make sure that the results are superior.

3.5. Bound Testing Approach

The next step after getting the order of integration for each variable under concern, the ARDL bound testing [45] is implemented to investigate the existence of long run association between the variables. The ARDL bound testing depends on assessing the joint significance of the lagged variables coefficients using F-Wald test, which has a null hypothesis of H0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$, whereas the alternative is that at least one coefficient (β) is not equal to zero. There are upper and lower critical values that should be compared with the results obtained from the F-statistics, where all are tabulated by [45]. There will be a proof of cointegration with a presence of long run association between the variables if the estimated F-statistic is more than the upper limit. If the computed F-statistic falls between the two critical limits, no conclusion can be provided by the test. However, the null hypothesis cannot be rejected if the values of F-statistic is less than the lower limit. In addition, this applies that there is no existence of cointegration among the tested variables.

If the estimated F-statistic shows the presence of long run relationship between the variables, the next stage of the ARDL specification is estimating the long run coefficients in (2) [50]. The long run impact on the N_2O emissions is measured by the estimated values of β_i . The Akaike Information Criteria (AIC) [51] is utilized to decide on the optimal lag length for each variable.

The residuals obtained from estimating (2) are used to approximate the error correction term (ECT), which shows the speed that the variables restore to their equilibrium levels in the long run from the short run. Therefore, the value of the coefficient of ECT should be negative and less than or equal to one besides being highly significant. Hence, the specifications of the error correction term of the ARDL approach can be estimated as follows:

$$\Delta ln\,(N_2O)_t = \delta_0 + \sum_{i=0}^{n} \delta_{1i}\,\Delta ln\,N_2O_{t-i} + \sum_{k=0}^{q} \delta_{1k}\,\Delta ln\,GDP_{t-k} + \sum_{j=0}^{d} \delta_{2j}\,\Delta\,(ln\,GDP_{t-j})^2 + \sum_{l=0}^{b} \delta_{3l}\,\Delta\,ln\,Elec_{t-l} + \sum_{w=0}^{y} \delta_{4w}\,\Delta ln\,Fin_{t-w} + \sum_{m=0}^{r} \delta_{5m}\,\Delta\,ln\,FDI_{t-m} + \theta ECT_{t-1} + \varepsilon_t \quad (4)$$

where Δ represents the growth or changes in N_2O emissions ($\boldsymbol{N_2O}$), GDP per capita ($\boldsymbol{GDP}$) and its quadratic term, electricity consumption per capita ($\boldsymbol{Elec}$), financial developments ($\boldsymbol{Fin}$), and the percentage of foreign direct investments ($\boldsymbol{FDI}$). The term of $\boldsymbol{ECT}$ reflects the speed of adjustment when a deviation take place. The value of the $\boldsymbol{ECT}$ is negative.

4. Empirical Results and Discussion

Before estimating any time series model, it is essential to examine the integration order of the variables and identify their order of integration. This study employs ADF test, KPSS test and PP test. The results of the three tests are reported in Table 2 which reveals that all the variables except the FDI are having unit root at level however a first difference convert them to stationary variables. Hence, the integration order is 1; i.e. I (1).

Table 2: Unit root results

Variable	ADF		KPSS		PPerron	
	Constant	Constant and Trend	Constant	Constant and Trend	Constant	Constant and Trend
lnN_2O	-2.33	-3.93**	0.48**	0.14*	-3.01**	-4.10*
lnGDPpc	-1.07	-2.33	0.31	0.11	-1.29	-2.55
lnElec	-3.23**	-2.55	0.37	0.15**	-3.65**	-2.57
Fin	-0.11	-3.04	0.46**	0.15**	0.22	-2.99
FDI	-5.51***	-5.48***	0.29	0.15**	-5.61***	-5.55***
ΔlnN2O	-7.49***	-8.35***	0.32	0.11	-7.37***	-8.45***
ΔlnGDpc	-4.74***	-4.63***	0.14	0.11	-4.76***	-4.62***
ΔlnElec	-5.24***	-5.69***	0.35	0.15**	-5.23***	-5.73***
ΔFin	-5.83***	-5.74***	0.29	0.13	-6.02***	-5.92***
ΔFDI	-8.05***	-7.94***	0.31	0.17**	-11.2***	-11.1***

Notes: ADF is the Augmented Dickey Fuller Unit root test, KPSS is the Kwiatkowski, Phillips, Schmidt & Shin stationarity test. PP is Phillips & Perron unit root test. lnN_2O is the natural logarithm of nitrous oxide emissions, lnGDPpc is the natural logarithm of GDP per capita, lnElec is the natural logarithm of electric power consumption, Fin is the Domestic credit provided by Financial Sector (% of GDP) and FDI is Foreign direct investment, net inflows (% of GDP). Δ is the first difference. *, ** and *** show 10%, 5% and 1% level of significance, respectively.

The findings obtained from stationary tests serve as the basics to implement the following steps of estimation. In order to explore the presence of long run relationship among the N_2O emissions and its determinants, the bound testing method of ARDL that aims for exploring cointegration is employed in (4) and the results are shown in Table 3. AIC is utilized in selecting the optimal lag structure for all the variables and the results are as (1,0,0,0,2,0) for the function $N_2O = f\ (GDP, GDP^2, Elec, Fin, FDI)$. The estimated F statistic is 10.052 which is greater than the value of the upper critical limit developed by [52]. Therefore, it is possible that the null hypothesis of no cointegration is rejected.

Table 3: Results of ARDL bound testing to cointegration

Model	Optimal lag structure	F - value	t - statistics
$N_2O = f\ (GDP, GDP^2, Elec, Fin, FDI)$	(1,0,0,0,2,0)	10.052	-6.77***

Table 4: Estimated Coefficients from ARDL (1,0,0,0,2,0) for Model $N_2O = f\ (GDP, GDP^2, Elec, Fin, FDI)$

	Variable	Coefficients	t-statistics
Long run estimates: lnN_2O as **dependent variable**	GDP_t	4.5	2.18**
	${GDP^2}_t$	-2.3	-2.20**
	$Elec_t$	0.66	2.74**
	Fin_t	-0.20	-4.64***
	FDI_t	-0.40	-1.33*
Short run estimates: ΔlnN_2O as **dependent variable**	GDP_{t-1}	2.30	2.42**
	${GDP^2}_{t-1}$	-4.6	-2.44**
	$Elec_{t-1}$	0.72	2.36**
	Fin_{t-1}	-0.10	-1.29
	Fin_{t-2}	0.40	3.60**
	FDI_{t-1}	-0.20	-0.33
	ECT_{t-1}	-0.76	-6.77***
	Constant	14.43	1.54
	trend	-0.44	-6.97***

Table 5: Diagnostic Tests

Test	Coefficient
R^2	0.76
Adjusted R^2	0.66
F- statistics	10.052(0.023)
Jarque-Bera normality test	1.062(0.334)
Heteroscedasticity Test: ARCH	2.83(0.984)
Breusch-Godfrey Serial Correlation LM Test	0.718(0.315)
Ramsey RESET test	0.723(0.413)

Table 4 presents the findings of ARDL estimation. The upper part of Table 4 illustrates the coefficients of the long run relationship. All the estimated coefficients appear to have significant impacts on N_2O emission at 1% or 5% level of significance except the coefficient of FDI, which is only significant at 10% level. Specifically, the GDP per capita elasticity is statistically significant and has a positive sign in both long and short run relationships. Moreover, the coefficient of the squared GDP per capita turned out to be negative and significant in both timeframes. The negative coefficient on the squared value of GDP per capita confirms the presence of the Environmental Kuznets Curve (EKC), which indicates the relationship between Bahrain economic growth and the level of N_2O emissions follows the inverted U-shape curve in the long run.

The findings of Table 4 demonstrate the coefficients of the other variables under concern. A 1% increase in electricity consumption causes 0.66% surge in N_2O diffusion. However, foreign direct investments and financial development variables are negatively related to the N_2O emissions as they cause a decrease in N_2O emissions of 0.4% and 0.2%, respectively. This is a good sign for policy makers to consider the improvement in the financial sector and draw the attention of foreign direct investments, which may help boosting air quality in Bahrain. Furthermore, the electricity consumption variable occurs to have a negative significant impact on N_2O emissions level.

The lower part of Table 4 reveals the findings of the short run estimations of the dynamic effects on N_2O diffusions by its determinants. The Δ sign implies that the variables are in their first differences. Briefly, for the error correction term (ECT_{t-1}), the coefficient is not only negative (as expected) but also significant at 5% level. This reinforces the hypothesis of cointegration and gives a measure for the how fast is the adjustment to equilibrium in the short run, which is around 76% in a year.

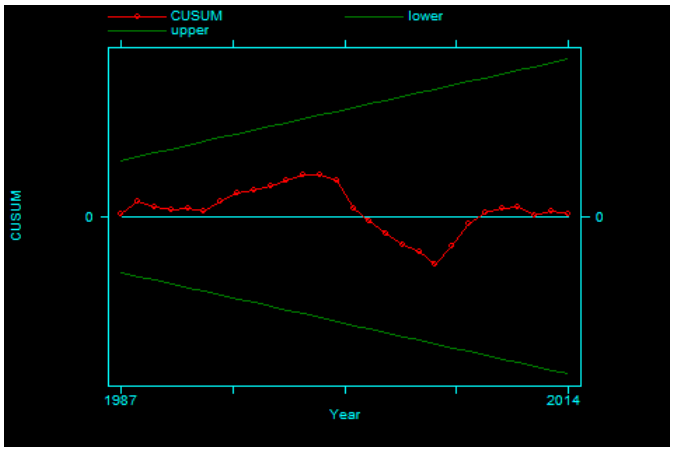


Figure 1: Plot of Cumulative Sum of Recursive Residuals

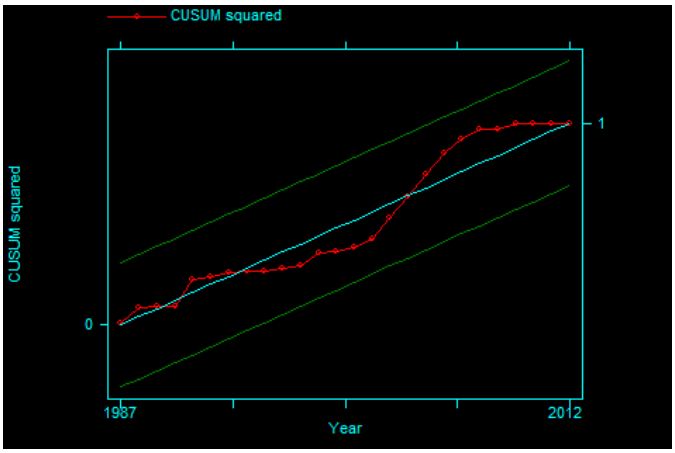


Figure 2: Plot of Cumulative Squared Sum of Recursive Residuals

To reassure the stability of the obtained findings, Table 5 reports the results obtained from some diagnostic tests that are applied for ARDL estimations. The findings prove that there is no serial correlation in the residuals, and the distribution is normal. Moreover, the variance of the errors is constant (homoscedastic).

Furthermore, the cumulative sum of recursive residuals (CUSUM) and the cumulative squared sum of recursive residuals (CUSUMSQ) established by [53] are plotted in Figures 1 and 2 to examine the model's parameters stability. Figures 1 and 2 illustrate the stability of the estimated model as both plots occur within the limits of the 5% confidence interval. It is important to check the stability of the model for it to be strong enough for forecasting issues and accordingly applicable for policies implementations.

5. Conclusion and Recommendations

This paper examines the validity of the EKC hypothesis in Bahrain over the period 1980 – 2012 with the implication of nitrous oxide (N_2O) emissions as an environmental pollutant. The estimated equation is augmented with electric power consumption, foreign direct investment and financial development indicator. ARDL approach is employed to examine the short and long run impact of the variables under interest on N_2O emissions. The obtained results from the ARDL estimation indicate the validity of the EKC hypothesis. Moreover, electric power consumption has a positive impact on N_2O emissions in the short and long run. However, foreign direct investments and financial development have negative impact on the emissions of N_2O.

In order to decrease the impact of electric power consumption on N_2O emissions, Bahrain should continue its efforts in decreasing the environmental emissions such as assisting households to install solar cells on the top of their houses and use solar energy in generating their own need for electricity, which may help in reducing the electricity production using fossil fuel combustion.

Although the authors of [36] and [37] argue that foreign investors and international corporations prefer investing in countries that have loose environmental policies and standards as most of these investments sponsor forms of production that are environmentally inefficient [38], this paper found that FDI reduces N_2O emissions. Our results are in line with that of [54] who found that FDI can be good for the environment. They explain this relationship by referring to the possibility of transferring the foreign firms' green technologies to their domestic counterparts which may have low environmental-friendly technologies. As the FDI helps decreasing N_2O emissions, Bahrain should increase its focus on the quality of FDI to be technology oriented FDI. Moreover, the financial development indicator reduces the N_2O emissions. Therefore, Bahrain should enhance its financial sector by developing bond and securities markets. This will improve the financial services and provide more funds to be invested in research and development on new and advanced techniques to generate clean energy.

Acknowledgment

Authors would like to thank the referees for their useful comments and support.

References

[1] R.J. Hill, E. Magnani, "An exploration of the conceptual and empirical basis of the environmental Kuznets curve," Australian Economic Papers, **41**(2), 239–254, 2002. doi: 10.1111/1467-8454.00162

[2] M.D. Miah, M.F.H. Masum, M. Koike, "Global observation of EKC hypothesis for CO2, SOx and NOx emission: A policy understanding for climate change mitigation in Bangladesh," Energy Policy, **38**(8), 4643–4651, 2010. doi: 10.1016/j.enpol.2010.04.022

[3] M.A. Zambrano-Monserrate, M.A. Fernandez, "An environmental Kuznets curve for N2O emissions in Germany: an ARDL approach," in Natural resources forum, Wiley Online Library: 119–127, 2017. doi: 10.1111/1477-8947.12122

[4] A.M. Rasli, M.I. Qureshi, A. Isah-Chikaji, K. Zaman, M. Ahmad, "New toxics, race to the bottom and revised environmental Kuznets curve: The case of local and global pollutants," Renewable and Sustainable Energy Reviews, **81**, 3120–3130, 2018. doi: 10.1016/j.rser.2017.08.092

[5] G.M. Grossman, A.B. Krueger, Environmental impacts of a North American free trade agreement, National Bureau of economic research Cambridge, Mass., USA, 1991.

[6] T. Panayotou, Empirical tests and policy analysis of environmental degradation at different stages of economic development, International Labour Organization, 1993.

[7] S. Farhani, M. Shahbaz, R. Sbia, A. Chaibi, "What does MENA region initially need: grow output or mitigate CO2 emissions?," Economic Modelling, **38**, 270–281, 2014. doi: 10.1016/j.econmod.2014.01.001

[8] L.-S. Lau, C.-K. Choong, Y.-K. Eng, "Investigation of the environmental Kuznets curve for carbon emissions in Malaysia: do foreign direct investment and trade matter?," Energy Policy, **68**, 490–497, 2014. doi: 10.1016/j.enpol.2014.01.002

[9] S. Oshin, A. Ogundipe, "An empirical examination of environmental Kuznets curve (EKC) in West Africa," Euro-Asia Journal of Economics and Finance, **3**(1), 2014. https://ssrn.com/abstract=2512136

[10] N. Apergis, "Environmental Kuznets curves: New evidence on both panel and country-level CO2 emissions," Energy Economics, **54**, 263–271, 2016. doi: 10.1016/j.eneco.2015.12.007

[11] S. Park, Y. Lee, "Regional model of EKC for air pollution: Evidence from the Republic of Korea," Energy Policy, **39**(10), 5840–5849, 2011. doi: 10.1016/j.enpol.2011.06.028

[12] O.A. Onafowora, O. Owoye, "Bounds testing approach to analysis of the environment Kuznets curve hypothesis," Energy Economics, **44**, 47–62, 2014. doi: 10.1016/j.eneco.2014.03.025

[13] S. Özokcu, Ö. Özdemir, "Economic growth, energy, and environmental Kuznets curve," Renewable and Sustainable Energy Reviews, **72**, 639–647, 2017. doi: 10.1016/j.rser.2017.01.059

[14] E.L. Effiong, A.O. Iriabije, "Let the data speak: semiparametric evidence on the environmental Kuznets curve in Africa," Quality & Quantity, **52**(2), 771–782, 2018. doi: 10.1007/s11135-017-0487-6

[15] M. Fodha, O. Zaghdoud, "Economic growth and pollutant emissions in Tunisia: an empirical analysis of the environmental Kuznets curve," Energy Policy, **38**(2), 1150–1156, 2010. doi: 10.1016/j.enpol.2009.11.002

[16] C.-H. Cho, Y.-P. Chu, H.-Y. Yang, "An environment Kuznets curve for GHG emissions: a panel cointegration analysis," Energy Sources, Part B: Economics, Planning, and Policy, **9**(2), 120–129, 2014. doi: 10.1080/15567241003773192

[17] L. Charfeddine, Z. Mrabet, "The impact of economic development and social-political factors on ecological footprint: A panel data analysis for 15 MENA countries," Renewable and Sustainable Energy Reviews, **76**, 138–154, 2017. doi: 10.1016/j.rser.2017.03.031

[18] Y. Chiu, "Deforestation and the environmental Kuznets curve in developing countries: A panel smooth transition regression approach," Canadian Journal of Agricultural Economics/Revue Canadienne d'agroeconomie, **60**(2), 177–194, 2012. doi: 10.1111/j.1744-7976.2012.01251.x

[19] K. Ahmed, M. Shahbaz, A. Qasim, W. Long, "The linkages between deforestation, energy and growth for environmental degradation in Pakistan," Ecological Indicators, **49**, 95–103, 2015.

[20] M. Mazzanti, A. Montini, R. Zoboli, "Municipal waste generation and the EKC hypothesis new evidence exploiting province-based panel data," Applied Economics Letters, **16**(7), 719–725, 2009. doi: 10.1080/13504850701221824

[21] P.K. Narayan, S. Narayan, "Carbon dioxide emissions and economic growth: Panel data evidence from developing countries," Energy Policy, **38**(1), 661–666, 2010. doi: 10.1016/j.enpol.2009.09.005

[22] Z. Zoundi, "CO2 emissions, renewable energy and the Environmental

Kuznets Curve, a panel cointegration approach," Renewable and Sustainable Energy Reviews, **72**, 1067–1075, 2017. doi: 10.1016/j.rser.2016.10.018

[23] S. Sinha Babu, S.K. Datta, "The relevance of environmental Kuznets curve (EKC) in a framework of broad-based environmental degradation and modified measure of growth–a pooled data analysis," International Journal of Sustainable Development & World Ecology, **20**(4), 309–316, 2013. doi: 10.1080/13504509.2013.795505

[24] F. Abbasi, K. Riaz, "CO2 emissions and financial development in an emerging economy: an augmented VAR approach," Energy Policy, **90**, 102–114, 2016. doi: 10.1016/j.enpol.2015.12.017

[25] S. khoshnevis Yazdi, B. Shakouri, "The renewable energy, CO2 emissions, and economic growth: VAR model," Energy Sources, Part B: Economics, Planning, and Policy, **13**(1), 53–59, 2018. doi: 10.1080/15567249.2017.1403499

[26] U. Al-Mulali, S.A. Solarin, I. Ozturk, "Investigating the presence of the environmental Kuznets curve (EKC) hypothesis in Kenya: an autoregressive distributed lag (ARDL) approach," Natural Hazards, **80**(3), 1729–1747, 2016. doi: 10.1007/s11069-015-2050-x

[27] E. Dogan, B. Turkekul, "CO 2 emissions, real output, energy consumption, trade, urbanization and financial development: testing the EKC hypothesis for the USA," Environmental Science and Pollution Research, **23**(2), 1203–1213, 2016. doi: 10.1007/s11356-015-5323-8

[28] W. Ali, A. Abdullah, M. Azam, "Re-visiting the environmental Kuznets curve hypothesis for Malaysia: fresh evidence from ARDL bounds testing approach," Renewable and Sustainable Energy Reviews, **77**, 990–1000, 2017. doi: 10.1016/j.rser.2016.11.236

[29] U.K. Pata, "Renewable energy consumption, urbanization, financial development, income and CO2 emissions in Turkey: testing EKC hypothesis with structural breaks," Journal of Cleaner Production, **187**, 770–779, 2018. doi: 10.1016/j.jclepro.2018.03.236

[30] B. Raggad, "Carbon dioxide emissions, economic growth, energy use, and urbanization in Saudi Arabia: evidence from the ARDL approach and impulse saturation break tests," Environmental Science and Pollution Research, **25**(15), 14882–14898, 2018. doi: 10.1007/s11356-018-1698-7

[31] A.S. Alshehry, M. Belloumi, "Study of the environmental Kuznets curve for transport carbon dioxide emissions in Saudi Arabia," Renewable and Sustainable Energy Reviews, **75**, 1339–1347, 2017. doi: 10.1016/j.rser.2016.11.122

[32] H. Mahmood, T.T.Y. Alkhateeb, "Trade and environment nexus in Saudi Arabia: An environmental Kuznets curve hypothesis," International Journal of Energy Economics and Policy, **7**(5), 291–295, 2017. https://EconPapers.repec.org/RePEc:eco:journ2:2018-03-5

[33] M. Shahbaz, R. Sbia, H. Hamdi, I. Ozturk, "Economic growth, electricity consumption, urbanization and environmental degradation relationship in United Arab Emirates," Ecological Indicators, **45**, 622–631, 2014. doi: 10.1016/j.ecolind.2014.05.022

[34] L. Charfeddine, K. Ben Khediri, "Financial development and environmental quality in UAE: Cointegration with structural breaks," Renewable and Sustainable Energy Reviews, **55**, 1322–1335, 2016. doi: 10.1016/j.rser.2015.07.059

[35] Z. Mrabet, M. Alsamara, "Testing the Kuznets Curve hypothesis for Qatar: A comparison between carbon dioxide and ecological footprint," Renewable and Sustainable Energy Reviews, **70**, 1366–1375, 2017. doi: 10.1016/j.rser.2016.12.039

[36] N.D. Woods, "Interstate competition and environmental regulation: a test of the race-to-the-bottom thesis," Social Science Quarterly, **87**(1), 174–189, 2006. doi: 10.1111/j.0038-4941.2006.00375.x

[37] C. Dick, A.K. Jorgenson, "Sectoral foreign investment and nitrous oxide emissions: A quantitative investigation," Society & Natural Resources, **23**(1), 71–82, 2009. doi: 10.1080/08941920802392690

[38] P. Grimes, J. Kentor, "Exporting the greenhouse: foreign capital penetration and CO? Emissions 1980 1996," Journal of World-Systems Research, 261–275, 2003.

[39] U. Soytas, R. Sari, B.T. Ewing, "Energy consumption, income, and carbon emissions in the United States," Ecological Economics, **62**(3–4), 482–489, 2007. doi: 10.1016/j.ecolecon.2006.07.009

[40] M. Shahbaz, M.M. Rahman, "Foreign capital inflows-growth nexus and role of domestic financial sector: an ARDL co-integration approach for Pakistan," Journal of Economic Research, **15**(3), 207–231, 2010.

[41] T. Chang, W. Fang, L.-F. Wen, "Energy consumption, employment, output, and temporal causality: evidence from Taiwan based on cointegration and error-correction modelling techniques," Applied Economics, **33**(8), 1045–1056, 2001. 10.1080/00036840122484

[42] M.H. Pesaran, "The role of economic theory in modelling the long run," The Economic Journal, **107**(440), 178–191, 1997. https://www.jstor.org/stable/2235280

[43] M.H. Pesaran, R.P. Smith, "Structural analysis of cointegrating VARs," Journal of Economic Surveys, **12**(5), 471–505, 1998. doi: 10.1111/1467-6419.00065

[44] H.H. Pesaran, Y. Shin, "Generalized impulse response analysis in linear multivariate models," Economics Letters, **58**(1), 17–29, 1998. doi: 10.1016/S0165-1765(97)00214-0

[45] M.H. Pesaran, Y. Shin, R.J. Smith, "Bounds testing approaches to the analysis of level relationships," Journal of Applied Econometrics, **16**(3), 289–326, 2001. doi: 10.1002/jae.616

[46] S. Johansen, K. Juselius, "Maximum likelihood estimation and inference on cointegration—with appucations to the demand for money," Oxford Bulletin of Economics and Statistics, **52**(2), 169–210, 1990. doi: 10.1111/j.1468-0084.1990.mp52002003.x

[47] D.A. Dickey, W.A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," Journal of the American Statistical Association, **74**(366a), 427–431, 1979. doi: 10.1080/01621459.1979.10482531

[48] P.C.B. Phillips, P. Perron, "Testing for a unit root in time series regression," Biometrika, **75**(2), 335–346, 1988. doi: 10.1093/biomet/75.2.335

[49] D. Kwiatkowski, P.C.B. Phillips, P. Schmidt, Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?," Journal of Econometrics, **54**(1–3), 159–178, 1992. doi: 10.1016/0304-4076(92)90104-Y

[50] M. Bouznit, M. del P. Pablo-Romero, "CO2 emission and economic growth in Algeria," Energy Policy, **96**, 93–104, 2016. doi: 10.1016/j.enpol.2016.05.036

[51] Y. Sakamoto, M. Ishiguro, G. Kitagawa, "Akaike information criterion statistics," Dordrecht, The Netherlands: D. Reidel, **81**(10.5555), 26853, 1986. doi: 10.1080/01621459.1988.10478680

[52] P.K. Narayan, "The saving and investment nexus for China: evidence from cointegration tests," Applied Economics, **37**(17), 1979–1990, 2005. doi: 10.1080/00036840500278103

[53] R.L. Brown, J. Durbin, J.M. Evans, "Techniques for testing the constancy of regression relationships over time," Journal of the Royal Statistical Society: Series B (Methodological), **37**(2), 149–163, 1975. doi: 10.1111/j.2517-6161.1975.tb01532.x

[54] B.A. Demena, S.K. Afesorgbor, "The effect of FDI on environmental emissions: Evidence from a meta-analysis," Energy Policy, **138**, 111192, 2020. doi: 10.1016/j.enpol.2019.111192

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 6, No. 4, 196-211 (2021)
www.astesj.com
Special Issue on Innovation in Computing, Engineering Science & Technology

ASTES Journal
ISSN: 2415-6698

Boltzmann-Based Distributed Control Method: An Evolutionary Approach Using Neighboring Population Constraints

Gustavo Alonso Chica Pedraza[1*], Eduardo Alirio Mojica Nava[2], Ernesto Cadena Muñoz[3]

[1]*School of Telecommunications Engineering, Universidad Santo Tomás, Bogotá D.C., 10111, Colombia*

[2]*Department of Electric and Electronic Engineering, Universidad Nacional de Colombia, Bogotá D.C., 10111, Colombia*

[3]*Department of Systems and Industrial Engineering, Universidad Nacional de Colombia, Bogotá D.C., 10111, Colombia*

ARTICLE INFO

Article history:
Received: 30 April, 2021
Accepted: 06 July, 2021
Online: 27 July, 2021

Keywords:
Distributed control
Entropy
Learning
Population dynamics
Selection-Mutation

ABSTRACT

In control systems, several optimization problems have been overcome using Multi-Agent Systems (MAS). Interactions of agents and the complexity of the system can be understood by using MAS. As a result, functional models are generated, which are closer to reality. Nevertheless, the use of models with permanent availability of information between agents is assumed in these systems. In this sense, some strategies have been developed to deal with scenarios of information limitations. Game theory emerges as a convenient framework that employs concepts of strategy to understand interactions between agents and maximize their outcomes. This paper proposes a learning method of distributed control that uses concepts from game theory and reinforcement learning (RL) to regulate the behavior of agents in MAS. Specifically, Q-learning is used in the dynamics found to incorporate the exploration concept in the classic equation of Replicator Dynamics (RD). Afterward, through the use of the Boltzmann distribution and concepts of biological evolution from Evolutionary Game Theory (EGT), the Boltzmann-Based Distributed Replicator Dynamics are introduced as an instrument to control the behavior of agents. Numerous engineering applications can use this approach, especially those with limitations in communications between agents. The performance of the method developed is validated in cases of optimization problems, classic games, and with a smart grid application. Despite the information limitations in the system, results obtained evidence that tuning some parameters of the distributed method allows obtaining an analogous behavior to that of the conventional centralized schemes

1 Introduction

This original research paper is an extension of the work initially presented in the Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI) 2020 [1]. In this version, readers can find a full view of the proposed learning distributed method, which uses concepts from Game Theory (GT) to control complex systems. This paper also presents an evaluation of the method from an evolutionary perspective of the obtained equations. This work also simulates a modified version of the case study presented in the conference, which includes different communication constraints and attributes of the generators employed in the power grid. Moreover, some additional cases in the context of classic games and maximization problems are introduced to make clearer the incidence of some control parameters in the behavior of agents.

The idea to model and control complex systems has increased over time. In this context, Engineering applications have received special interest due to their affinity with the use of mathematical techniques to prove new models and concepts on applications closer to reality [2]. In recent decades, research has been focused on the study of distributed systems with large-scale control. Numerous models and techniques have been developed to overcome issues such as the expensive computational requirements, the structure of the communication, and the calculation of the data required to complete a task in large-scale systems. These issues can be managed by using Multi-Agent Systems (MAS) and concepts from game theory [3]. In this sense, the interactions of agents have been thoroughly studied, as some strategies can help agents maximize their outcomes. For example, [4] establishes relations among games, learning, and

*Corresponding Author: Gustavo Alonso Chica Pedraza, Carrera 9 No 51-11, Bogotá D.C., Colombia, +5715878797 Ext 1654, gustavochica@usantotomas.edu.co

optimization in networks. Other studies have focused on games and learning [5] or on algorithms for distributed computation in topologies of dynamic networks [6]. Authors in [7] studied the main applications of power control in the frameworks of distributed and centralized game theory. Regarding smart grid control applications, please refer to [8]. Other research has concentrated on cases with issues in coordination and negotiation that guide the study of the interactions of agents[9]. For further studies on applications of power control using game theory, please refer to [10]. Research on game theory considers three types of games. First, continuous games consider the way an agent can have a pure strategy looking for maximum profit. Second, in matrix games, agents are regarded as individuals and can take only one shot to play simultaneously. Finally, dynamic games suppose that players can learn in some way about the environment, that is, their actions and states. This assumption means agents can learn and correct their behavior based on the outcomes of their actions [11]. Dynamic games must deal with the following challenges: modeling the environment for agents interaction, modeling the agents goals, the prioritization of the agents actions, and the estimation of the amount of information owned by a player [12].

The study of dynamics of agents changing over time is a concept of dynamic games introduced by Evolutionary game theory (EGT) [13]. The concept of the evolutionary stable strategy popularized EGT thanks to the analogy with biology concepts and the comparison with natural behaviors [14]. Some real-life control applications have employed EGT, whose understanding serves as a basis for the replicator dynamics (RD) approach. The revision protocols describe the way agents choose and modify their strategies, while population games determine the agents' interactions. The combination of both revision protocols and population games produces the concept of evolutionary game dynamics [14]. This perspective of evolution is often used to model large-scale systems because its mathematical background helps to describe this process with differential equations [13]. Many areas of Engineering have applied EGT, for example, optimization problems, control of communication access, systems of microgrids, etc. [11]. The use of EGT to model engineering problems has revealed the following benefits: ease to relate a game to an engineering problem, where payoff functions can be defined with the objective function and the strategies, and the relationship between the optimization concept and Nash Equilibrium, which is enabled under particular conditions that met the conditions of the first-order optimization of the Karush–Kuhn–Tucker. Last but not least, EGT uses local information to achieve solutions. In this sense, distributed approaches emerge to tackle engineering problems, which is useful when considering the implementation cost of centralized schemes and their complexity [11]. Distributed schemes of population dynamics have outstanding features over techniques like the method of dual decomposition, which requires a centralized coordinator [15]. This characteristic reduces the associated cost with the structure of communication. Additionally, in comparison with distributed learning algorithms in normal-form-games, there are no failures in distributed population dynamics when all the variables involved in decision-making have limitations[16]. This makes Distributed Population Dynamics suitable for solving issues regarding allocation of resources like in a smart city design [17]. For these purposes, the distributed power generation needs to be integrated so that electric grids be more reliable, robust, efficient, and flexible. Nevertheless, modeling a grid using a distributed approach instead of the classic centralized, is an option to consider due to its realism and flexibility, according to microgrids constraints [18]. In this sense, control operations are considered individually in microgrids, as they make a distinction among the power generation, the secondary frequency, and the economic dispatch [19]. Static optimization concepts are employed to manage the economic dispatch [20] or even methods like the offline direct search [21]. The analysis may be more complicated if it includes loads, the generator, and power line losses in the distributed model. Other approaches cannot consider the dynamic conditions like the economic dispatch time dependence [22]. Some approaches have been developed to face these challenges. For instance, [23] presents a management system for a microgrid with centralized energy and stand-alone mode to study its static behavior. Other research employs a distributed control strategy considering power line signaling for energy storage systems [24]. The employment of the MAS framework in economic problems using a distributed approach was gathered in [25], taking into account the delays in the communication system. Microgrid architectures have also been proposed considering distributed systems like the microgrid hierarchical control [26].

This paper presents an approach to overcome some of the issues identified in the literature review. The aim is to show how to develop a control method of learning to study the influence of the exploration concept in MAS, that is, interaction between agents. RD was developed from simple learning models [27], so this research seeks to bring the exploration concept into the traditional exploration-less expression of RD, using the Q-learning dynamics. As a result, the combination of these frameworks opens up a path to tackle dynamics in a scenario where the feedback of each agent is determined by the agent itself and by other agents, and where interaction between them is limited. For the analysis, the Boltzmann distribution includes a distributed perspective of the Replicator Dynamics as a way to regulate the agents' behavior in a determined scenario. The developed method employs a temperature parameter and the presence of entropy terms, to modify the learning agents' behavior and link the selection-mutation process from EGT and the exploration-exploitation concept from RL. This attribute complies with the traditional positive condition of EGT techniques (modeling agents' interaction). Nevertheless, In the control area, the employment of these techniques has to be understood more on the normative side of things. To explain these features, this approach employs theory of RL, EGT, and decision-making to solve some cases in the context of classic games and maximization problems using a novel distributed model of learning. It also uses experimental data to tackle an economic dispatch problem, which is a common problem in smart grids. The results obtained by the proposed approach are contrasted with the classical centralized framework of RD.

The remainder of this paper is organized as follows. Section 2 presents, a short synopsis of game theory and reinforcement learning, as well as the relationship between EGT and Q-learning using the Boltzmann distribution. Section 3 explains a distributed neighboring concept used for the Boltzmann control method, considering the behavior of replicator dynamics. Section 4 introduces important concepts from the previous Section, related to evolutionary game theory and reinforcement learning. In Section 5, the employment of

the learning method on traditional cases of GT and maximization problems presents the background to analyze the application of the Boltzmann model behavior on a smart grid real-life case. Finally, the main conclusions of the study are summarized in Section 6.

2 Preliminaries

Game theory includes a group of equations and concepts to study the background in decentralized control issues. Most of the time, a game comprises a group of players (agents) with similar population behavior that choose the best way to execute actions. The strategy of a player can decrease rewards after performing a wrong action or increase rewards when the action was correct [28]. The theory of learning is used to understand this behavior. In this sense, the scheme of RL explains the relationship among the environment, signals, states, and actions. In the interaction, at each step, each player gets a notification with the current state of the environment and a reinforcement signal, then, the player chooses a strategy. Each player of the game aims to find the policy that produces the best rewards after recognizing the consequence of its actions, that is, reward or punishment. A structure of estimated value functions is characteristic of traditional RL methods [29]. The total reward that a player can obtain is usually a pair state-action or a state value. This means that the optimal value function is needed to find the policy that correctly fulfills payoffs. The Markov decision process and value iteration algorithm can be employed for this purpose [30] when the scenario is familiar. In other cases, Q-learning can be used as an adaptable method of value iteration where the model of the scenario does not require to be specific. Equation (1) depicts the Q-learning interaction process [31]:

$$Q_{t+1}(s,a) \leftarrow (1-\alpha)Q_t(s,a) + \alpha(\Gamma + \gamma max_{a'} Q_t(s',a')) \quad (1)$$

The whole process begins at time Q_{t+1} with an initial pair of action-state (s,a), then, after performing action a achieves the $Q_t(s',a')$, where (s',a') represents the newest values of s and a, respectively. $max_{a'}$ obtains the uppermost value of Q from s' by selecting the action that increases its value. α represents the general step size parameter, Γ is the instant reinforcement, and γ is a deduction parameter. When players have complete access to the game information and there are no communication limitations, the theory of learning and games are valuable instruments to deal with control applications that use a centralized approach. Nevertheless, these models aim to provide a close description of optimal circumstances, but they have some drawbacks when dealing with more realistic conditions, communication constraints, and the individuals rationality. In this vein, EGT tries to loosen the idea of rationality, by substituting it with biological notions like evolution, mutation, and natural selection [32, 33]. In EGT, there is a genetic encoding of the strategies of the players, which are called genotypes and represent the conduct of every player employed to calculate its outcomes. The quantity of other types of agents in the scenario determines the payoff of the genotype of each player genotype. In EGT, the population strategies begin to evolve employing a dynamic process that allows finding the expected value of this process through the use of the Replicator dynamics equation. An evolutionary system often returns to two concepts: mutation and selection. On the one hand, mutation provides variety to the population. On the other hand, selection provides priority to some varieties where every genotype is a pure strategy $Q_j(n)$, where the RD offspring expresses this behavior. The general equation of RD [27] is presented in Equation (2).

$$\frac{dx_i}{dt} = [(Ax)_i - x \cdot Ax]x_i \quad (2)$$

where x_i is the portion of a population that plays the i-th strategy. The payoff matrix is written as A and it owns diverse payoff values that each replicator obtains from other agents. The vector of probability $x = (x_1, x_2, ..., x_J)$ often defines the population state (x), and evidence the diverse density values of each type of replicator. Consequently, $(Ax)_i$ is the payoff obtained by the i-th player with x state. Then, the average payoff would be written as $x \cdot Ax$. Similarly, $\frac{dx_i}{dt}$ symbolizes the growth rate of the population playing the i-th strategy, which is calculated using the obtained payoff value after playing the i-th strategy and its difference with the average population payoff. [34].

2.1 Relating EGT and Q-Learning

In [35], the frameworks of RD and Q-learning are related in the context of two-player games, where players have different strategies. This relationship is conceivable as players can also be considered Q-learners. For modelling this case, a differential equation is needed for player R (rows) and another one for player C (columns). When $A = B^t$, the standard RD Equation (2) is employed, where x_i is substituted by r_i or c_i. Thence, A or B, and the change in state (x) for r or c determine the payoff matrix for a specific player. Therefore, $(Ax)_i$ switches to $(Ac)_i$ or $(Br)_i$ and is the reward obtained by the i-th player with a r or c state. Likewise, for players R and C, the growth rate $\frac{dx_i}{dt}$ switches to $\frac{dr_i}{dt}$ or $\frac{dc_i}{dt}$, respectively. This behavior is explained using the following system of differential equations [27] below:

$$\frac{dr_i}{dt} = [(Ac)_i - r \cdot Ac]r_i \quad (3)$$

$$\frac{dc_i}{dt} = [(Br)_i - c \cdot Br]c_i \quad (4)$$

Equations (3) and (4) denote the group of replicator dynamics equations used to model the behavior of two populations. Each population has a growth rate determined by the other populations. For example, A and B denote two payoff matrices that are needed to estimate the rate of change for two different current players in the problem using this group of differential equations. To find the relationship between the Q-learning framework and the RD equations, Equation (5) is introduced:

$$x_i(\delta) = \frac{e^{\tau Q_{a_i}(\delta)}}{\sum_{j=1}^{n} e^{\tau Q_{a_j}(\delta)}} \quad (5)$$

where the notation $x_i(\delta)$ means the prospect of using strategy i at time δ, and τ symbolizes the temperature. Equation (5) is well-known as the Boltzmann distribution and is used in [35] to obtain the continuous time model of Q-Learning in the context of a game played by two players, as shown in Equation (6), where $\frac{dx_i}{dt}$ is written as $\dot{x}_i$.

$$\frac{\dot{x}_i}{x_i} = \tau\Big[\frac{dQ_{a_i}}{dt} - \sum_{j=1}^{n} \frac{dQ_{a_i}}{dt} x_j\Big] \quad (6)$$

The expression $\frac{dQ_{a_i(t)}}{dt}$ in Equation (6) can be solved by using Equation (1) to represent the Q-learner update rule. Equation (7) presents the equation of difference for the function Q.

$$\Delta Q_{a_i}(\delta) = \alpha\Big[\Gamma_{a_i}(\delta+1) + \gamma \max Q - Q_{a_i}(\delta)\Big] \tag{7}$$

The term σ expresses the time spent between two repetitions of the Q-values updates, where $0 < \sigma \leq 1$, while $Q_{a_i}(\delta\sigma)$ denotes the Q-values at time $k\sigma$. Then, by assuming an infinitesimal scheme of this expression, Equation (7) converts to Equation (8) after taking the limit $\sigma \to 0$.

$$\frac{\dot{x}_i}{x_i} = \tau\alpha\Big[\Gamma_{a_i} - \sum_{j=1}^{n} x_j\Gamma_{a_j} + \sum_{j=1}^{n} x_j(Q_{a_j} - Q_{a_i})\Big] \tag{8}$$

As $\frac{x_j}{x_i}$ comes to $\frac{e^{\tau}\Delta Q_{a_j}}{e^{\tau}\Delta Q_{a_i}}$, the part after the sum in Equation (8) can be written in logarithm terms:

$$\alpha\Big[\tau\sum_{j} x_j(Q_{a_j} - Q_{a_i})\Big] = \alpha\Big[\sum_{j=1}^{n} x_j \ln\left(\frac{x_j}{x_i}\right)\Big] \tag{9}$$

The last expression in Equation (8) is reorganized and replaced, so it converts to Equation (10).

$$\frac{\dot{x}_i}{x_i} = \alpha\tau\Big[\Gamma_{a_i} - \sum_{j=1}^{n} x_j\Gamma_{a_j}\Big] + \alpha\Big[\sum_{j=1}^{n} x_j \ln\left(\frac{x_j}{x_i}\right)\Big] \tag{10}$$

For using payoff matrices in games with two players, Γ_{a_i} as $\sum_j a_{ij}y_j$ can be written, then, the expressions for players 1 and 2 are expressed as shown in Equations (11) and (12), respectively:

$$\dot{x}_i = x_i\alpha\tau\Big[(A\mathbf{y})_i - \mathbf{x}\cdot A\mathbf{y}\Big] + x_i\alpha\Big[\sum_{j=1}^{n} x_j \ln\left(\frac{x_j}{x_i}\right)\Big] \tag{11}$$

$$\dot{y}_i = y_i\alpha\tau\Big[(B\mathbf{x})_i - \mathbf{y}\cdot B\mathbf{x}\Big] + y_i\alpha\Big[\sum_{j=1}^{n} y_j \ln\left(\frac{y_j}{y_i}\right)\Big] \tag{12}$$

These expressions denote the derivation of the continuous-time model for Q-learning. For the full process of the derivation, see Annex A. The Equations (11) and (12) can be considered as a centralized perspective, analogous to the Equations (3) and (4) that represent the standard RD form to model actions of players R and C, in a game of 2 players. However, the Boltzmann model produces the main differences with the introduction of α and τ parameters, and the emergence of an additional term. This approach has been applied in some scenarios such as multiple state games, multiple player games, and in the context of 2×2 games [27]. Nevertheless, research is still needed to use this approach in real-life problems.

The following Section presents our approach, which is a learning method that uses a distributed population perspective to control agents' behaviors. This proposal uses some of the principles stated in [35] to introduce the Boltzmann-based distributed replicator dynamics approach. This paper also uses the concept of population dynamics but employing constraints in the agents communications and assuming players should use neighboring strategies, thus, having a scenario where players have no full information of the system.

3 The Boltzmann-based distributed replicator dynamics method

In the following paragraphs, we describe the Boltzmann-based distributed replicator dynamics method. The starting point needed to perform the development of this method is the Equation (11). This formalism is useful since it employs the Boltzmann concept and its first term has the classic form of the RD when modeling games that use payoff matrices. Considering the idea to have an analogous and more general form to express the RD expression, Equation (11) can be written as Equation (13):

$$\dot{x}_i = \alpha x_i\tau\Big[f_i(x) - \overline{f}(x)\Big] + \alpha x_i\Big[\sum_{j=1}^{n} x_j \ln\left(\frac{x_j}{x_i}\right)\Big] \tag{13}$$

In this equation, a fraction of a determined population can augment or diminish depending on the higher/lower fitness values of its individuals with respect to the population average. The population is represented by the state vector $x = (x_1, x_2, ..., x_n)^n$ with $0 \leq x_i \leq 1, \forall_i$ and $\sum_{i=1}^{n} x_i = 1$, which denotes the portions that belong to each of the n-types. In $f_i(x)$, i denotes the fitness type. Consequently, the fitness average of the population is expressed by $f(x) = \sum_j x_j f_j(x)$. Using these assumptions, this expression becomes:

$$\dot{x}_i = \alpha x_i\tau\Big[f_i(x) - \sum_{j=1}^{n} x_j f_j(x)\Big] + \alpha x_i\Big[\sum_{j=1}^{n} x_j \ln\left(\frac{x_j}{x_i}\right)\Big] \tag{14}$$

The first term of Equation (14) is written as the centralized equation for the RD. We propose to adapt it to a decentralized form, to compute the local information of the players to tackle limitations in communication. The decentralized expression of this step is written in Equation (15):

$$\dot{x}_i = \underbrace{\alpha x_i\tau\Big[f_i(x) - \sum_{j=1}^{n} x_j f_j(x)\Big]}_{Centralized} = \underbrace{\alpha x_i\tau\Big[f_i(x)\sum_{j=1}^{n} x_j - \sum_{j=1}^{n} x_j f_j(x)\Big]}_{Decentralized} \tag{15}$$

where $\sum_{j=1}^{n} x_j$ is equivalent to the unit, since the term x_j of the operation denotes the probabilities of selecting the *jth* strategy. Likewise, when using logarithms rules, the second term in Equation (14) becomes:

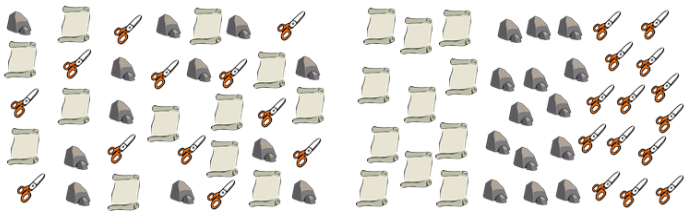
$$\underbrace{\alpha x_i\Big[\sum_{j=1}^{n} x_j \ln\left(\frac{x_j}{x_i}\right)\Big]}_{centralized} = \underbrace{-\alpha x_i\Big[\ln x_i - \sum_{j=1}^{n} x_j \ln x_j\Big]}_{Decentralized} \tag{16}$$

Finally, substituting Equations (15) and (16) in Equation (14), becomes Equation (17), that expresses the Decentralized form of the Replicator Dynamics equation in connection with Boltzmann probabilities.

$$\dot{x}_i = \alpha x_i\tau\Big[f_i(x)\sum_{j=1}^{n} x_j - \sum_{j=1}^{n} x_j f_j(x)\Big] - \alpha x_i\Big[\ln x_i - \sum_{j=1}^{n} x_j \ln x_j\Big] \tag{17}$$

This equation is studied in detail in Section 4 considering EGT with the selection-mutation concept and the exploration-exploitation

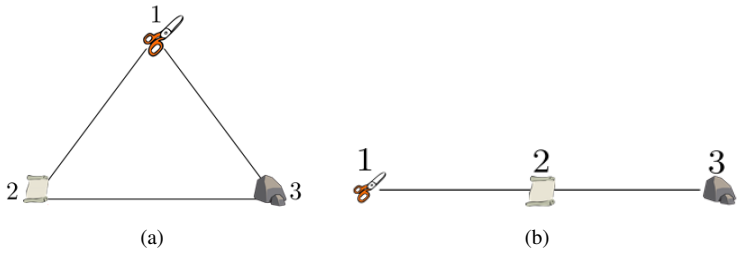
approaches with their influence on MAS. in Equation (17), the first parenthesis corresponds to alterations in the proportion of players that are using the i-th strategy and require complete information about the state of the whole population and the payoff functions. Consequently, complete information of the system is required so that population dynamics evolve. However, since this work aims to control scenarios where agents cannot access the complete information of the system, there should not be dependence on complete information, for example, in scenarios with limitations in communication infrastructure, big systems, or privacy matters that obstruct the process of sharing information. Since the population structure determines the features that explain players behaviors, the population structure in the classic approach owns a complete and well-mixed structure, which means that players can choose any strategy with the same probability as the others. Figure 1a illustrates this concept with some players in a game. We use element shapes such as scissors, paper, or stone to represent the chosen strategies of each agent.



(a) (b)

Figure 1: (a) Population Structure Without Constrains, (b) Constrained Population Structure

Considering EGT, each player can equally obtain a revision opportunity. When receiving this opportunity, players choose arbitrarily one of their neighbors and switch their chosen strategy to one of their neighbors based on the selected revision protocol. As players are supposed to have a full and well-mixed structure, any opponent has the same possibility of selecting and playing any strategy of the structure (Figure 1a). On the contrary, Figure 1b shows a case where constraints in the structure limit the capacity of an agent to select some strategies, which is also an approach closer to reality. In this case, all agents are equally likely to be given a revision opportunity, but a neighbor does not have the same probability to choose and play a particular strategy. For instance, when a player obtains a revision opportunity with a paper strategy, there is no opportunity of choosing an opponent with scissors. The reason is that no papers are close to any scissors. Nevertheless, in this player case, the prospect of choosing an adversary with a paper or stone plan is higher than in the scissors situation. The graph $G = (T, L, M)$ establishes a mathematical way to represent the behavior of agents and their dynamics. The set T symbolizes the strategies an agent can choose. Set L is the meeting probability between strategies. For contextualizing, the notation $M = [a_{ij}], a_{ij} = 1$ suggests that strategy j and i can find each other, but $a_{ij} = 0$ indicates that these strategies cannot meet. Thence, it is possible to define Ni as the set of neighbors of agent i. Full and well-mixed and constrained mixed populations can be represented by two types of graphs. Figure 2a depicts a complete graph for the full and well-mixed structures, while Figure 2b illustrates the case with constraints in the structure. The form of the graph is determined by the particular structure of the population. In this research, undirected graphs are employed, which means that the probability that strategies j and i find each other are the same as in strategies i and j.



(a) (b)

Figure 2: Graphs topology for (**a**) full and well-mixed structure and (**b**) constrained structure.

Now, for regulating agents interactions, the limitation of incomplete information dependency in the population structure of Equation (17) must be overcome. For this purpose, the work proposed by [2] is considered. Therefore, to incorporate the neighboring concept, we use the pairwise proportional imitation protocol, as expressed in Equation (18):

$$p_{ij} = p_j[f_j(p_{Ni}) - f_i(p_{Ni})]_+ \tag{18}$$

where the calculation of p_i only requires knowing the portions of the population that are playing neighboring strategies. Then, the following expression is assumed:

Assumption 1 *Operations that update behaviors of agents by employing the pairwise proportional imitation protocol use the neighboring concept, which means that the iterations in the sums and the payoff function are determined by those neighbors communicating effectively with the i-th player.*

In this vein, Equation (19) denotes the obtained distributed replicator dynamics that fulfill the limitations of the population structure and enable agents to regulate the calculation of incomplete information:

$$\dot{x}_i = \alpha x_i \tau\Big[f_i(x_{Ni}) \sum_{j\in Ni}^{n} x_j - \sum_{j\in Ni}^{n} x_j f_j(x_{Nj})\Big] \tag{19}$$

where $f_{i/j}(x_{Ni/j})$ is the payoff function for the ith or jth player, estimated by the proportion of population that effectively communicates with neighbors, and $\sum_{j\in Ni}^{n} x_j$ is a sum that just considers those neighbors who communicate effectively. As our statement about the neighboring concept was implemented just in the first part of Equation (17), the second part of the equation (second parenthesis) including this concept is written as follows:

$$-\alpha x_i\Big[\ln x_i - \sum_{k\in Ni}^{n} x_k \ln x_k\Big] \tag{20}$$

In this equation, k represents i-th neighbor with an active communication link that employs strategy j. The end of the equation expresses the way the i-th player behaves regarding the proposed method using the Boltzmann concept. Equation (21) denotes in a complete manner the Boltzmann-Based Distributed Replicator

Dynamics (BBDRD) which includes both concepts: the distributed and the neighboring.

$$\dot{x}_i = \alpha x_i \tau \underbrace{\left[f_i(x_{Ni}) \sum_{j \in Ni}^{n} x_j - \sum_{j \in Ni}^{n} x_j f_j(x_{Nj}) \right]}_{Exploitation} - \alpha x_i \underbrace{\left[\ln x_i - \sum_{k \in Ni}^{n} x_k \ln x_k \right]}_{Exploration} \tag{21}$$

As stated above, the BBDRD equation evidences the implementation of the exploration and the exploitation notions of RL, and the selection-mutation approach of EGT, as explained in the next section. The implementation of this approach and examples of its application in the context of classic games, maximization problems, and for a smart grid control are developed in Section 5.

4 Evolutionary Approximation

This section presents the control method stated in Equation (21) from the perspective of RL and in an evolutionary approximation, which is helpful to comprehend the introduction of the notion of exploration in the classic RD expression.

4.1 Evolutionary Perspective

The traditional structure of RD is represented in the first part of the dynamics of Equation (21). This allows approximating to the Q-learner dynamics from EGT, because the mechanism for selection is contained in it. Then, the mechanism for mutation is found in the complementary part of the expression, which means:

$$x_i \alpha \left(\sum_{k \in Ni}^{n} x_k \ln(x_k) - \ln(x_i) \right) \tag{22}$$

In Equation (22), there are two recognizable entropy values: the distribution of probability x and the value of the strategy x_i. The expressions for entropy can be written as:

$$E_i = -x_i \ln(x_i) \tag{23}$$

and

$$E_n = -\sum_{k \in Ni}^{n} x_k \ln(x_k) \tag{24}$$

where E_i represents the available information regarding strategy i, while E_n is the information of the complete distribution. Consequently, the mutation equation can be expressed now as:

$$-(\alpha x_i E_n - \alpha E_i) \tag{25}$$

The following expression is the mutation equation derived, considering the difference between old and new states of x_i.

$$\sum_{k \in Ni}^{n} \epsilon_{ik} x_k - x_i \tag{26}$$

In Equation (26), ϵ_{ik} expresses the rate of mutation of agents that employ the i-th strategy and select another strategy from the pool of the k neighbors, for example, strategy j. When k is higher or equal to 1, ϵ_{ik} becomes bigger than or equal to zero. Considering EGT, in the framework of Q-Learning dynamics, mutation is directly connected with entropy that expresses the strategy state. However, this connection already existed, since it has been evidenced that entropy augments with mutation [36]. This connection is described in [37] from the perspective of thermodynamics, taking into account the trend of mutation to augment to increase entropy. Additionally, the Q-learning dynamics evidence that RD is the basis for the development of the selection concept. In RD, the resulting payoff can favor or be independent of a strategy, and the behavior of its opponent is strongly related to the resulting payoff. The concept of mutation can be found too. This fact is estimated by comparing the value of the entropy strategy with the value of entropy of the entire population.

4.2 Reinforcement Learning Perspective

Reinforcement learning aims to compensate the exploration and exploitation mechanisms. For gaining the maximum profit, a player must execute an action. Commonly, the player chooses actions that paid a high compensation before. Nevertheless, if the player wants to identify these actions, it must choose actions that were not chosen before. The notion from RL of exploitation-exploration is understood from a biological perspective by establishing connections between exploitation/exploration and mutation/selection. For clarity purposes, the first term of Equation (21) always chooses the best courses of actions, which matches the exploitation concept. Likewise, the exploration term is introduced into the RD expression due to its direct connection with the terms of entropy in Equation (22). Note that high values of entropy produce a high level of uncertainty in choosing one course of action. Therefore, the term of exploration augments entropy and gives diversity all at once. Consequently, the exploration and mutation concepts are strongly related, as both of them give variety, and a feature of heterogeneity to the environment. Being in control of particular scenarios like heterogeneity and communication limitations in a system is a demanding task when addressing real-life cases. In this sense, the compensation of the exploration-exploitation mechanisms can be quite difficult since a fine adjustment is often required for the parameters involved in the learning process. This adjustment must be performed to regulate the behavior of players in the process of decision-making. This problem can be solved by using the BBDRD control method as demonstrated in Equation (21) and explained in the following section.

5 Illustrative cases

5.1 Rock-Paper-Scissors as a classic game

In this part of the document, the concept introduced in Equation (21) is implemented in one of the classic games for excellence, the rock-paper-scissors game. For this purpose, a single population with three strategies has been considered, where $x = [x_1, x_2, x_3]^\top$ represent each of them respectively. In the same sense, the expression $F(x) = Ax$ denotes the fitness function, where A represents the payoff matrix showed in Equation (27). It is worth noting that the classic payoff matrix has been modified to guarantee positive values

of the payoffs in all cases.

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 3 & 2 & 1 \\ 1 & 3 & 2 \end{pmatrix} \tag{27}$$

To start running the simulation, a time of 30 units was considered with 300 agents and 5000 iterations. Additionally, the following initial conditions were stated $x_0 = [0.2, 0.7, 0.1]^\top$. The classic behavior of the rock-paper-scissors game proposes that every single strategy has the same probability to be selected, which means the absence of a dominant strategy. This behavior can be evidenced using Δ representation, which is defined as follows:

Definition 1 *Let Δ be the representation of a triangle of n-dimensions known as a Simplex.*

Using a simplex helps in the understanding of the implicit dynamics. Since the simplex is composed of three vertices, each of them represents a strategy e.g. rock, paper, scissors, then, the classic expected simulation of this situation is depicted as shown in Figure (3)a. Similarly, Figure (3)b, shows how the evolution of the population strategies is completely symmetrical, which means they keep constant along the time.

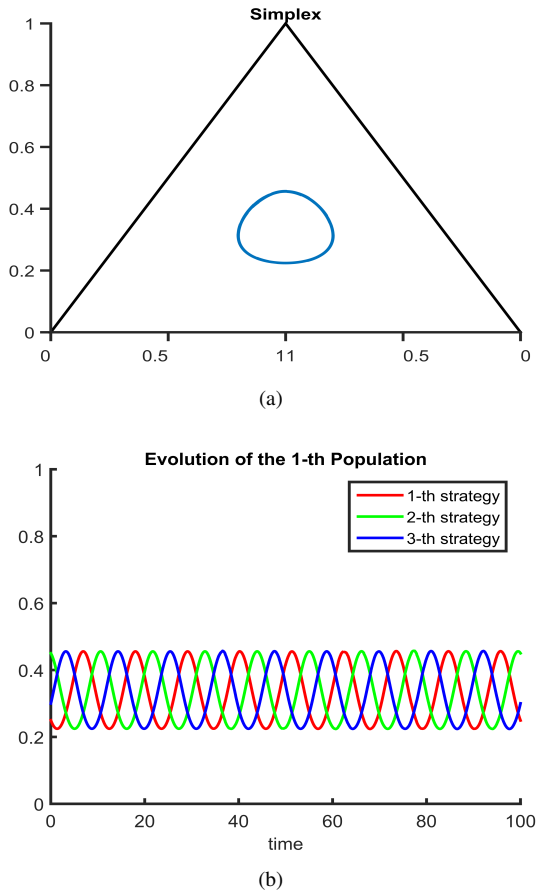


Figure 3: (a) Classic Rock-Paper-Scissors Behavior in a Simplex. (b) Evolution of the Population Behavior.

We also consider simulating a general distributed case to further compare it with the results of the BBDRD method. In both cases, the same simulation parameters were considered, but communication between agents was limited in the following way: agents playing strategy 1 were not allowed to communicate with agents playing strategy 3 and vice versa. Figure 4a shows the behavior of the distributed case, where the graphic seems to be an oval. This means that the interaction between strategies using only the first part of Equation (21) (general distributed case without entropy) tends to have a similar behavior to the one found in Figure 3a. Additionally, results depicted in Figure 4b show the evolution of the population strategies under the distributed case, where the symmetry is altered by the constraints in the communication of agents.

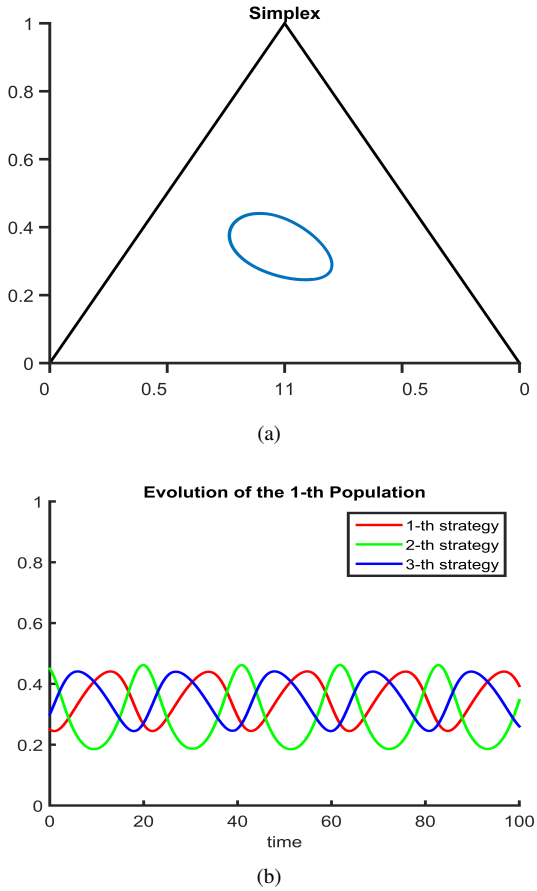
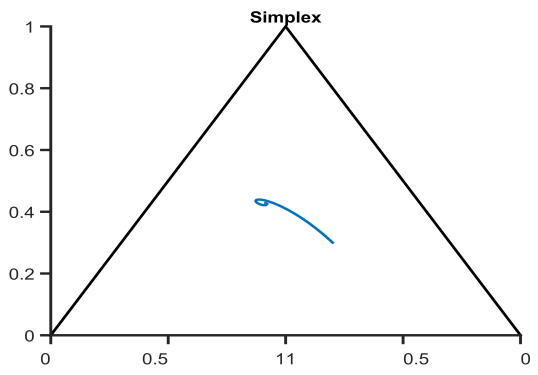
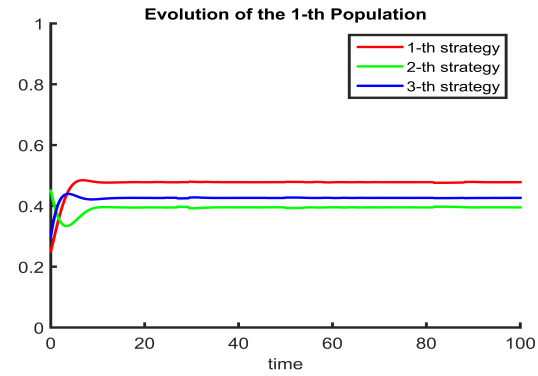


Figure 4: (a) Distributed Rock-Paper-Scissors Behavior in a Simplex. (b) Evolution of the Distributed Behavior.

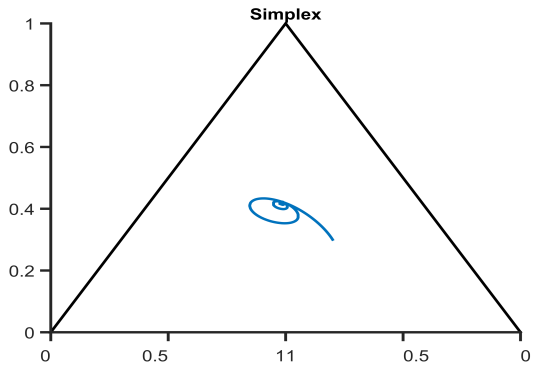
As mentioned previously, to compare these results with those obtained using the BBDRD method of Equation (21) (Distributed + Entropy case), Figure 5a shows that using τ= 1 the blue line depicts just one part of the oval (in contrast to 4a). Additionally, Figures 5c and 5d show the behavior of the model using τ values of 10 and 100 respectively. As evidenced, the bigger the term τ is, the more similar the behavior is to that obtained in the distributed case i.e. the contour of the oval seems to be equal to that obtained in Figure 4b), which at the same time is similar to the classic case. Finally, In Figure 5b, the evolution of the strategies population seems to be stable in all cases. This can be understood due to the introduction of the entropy term.



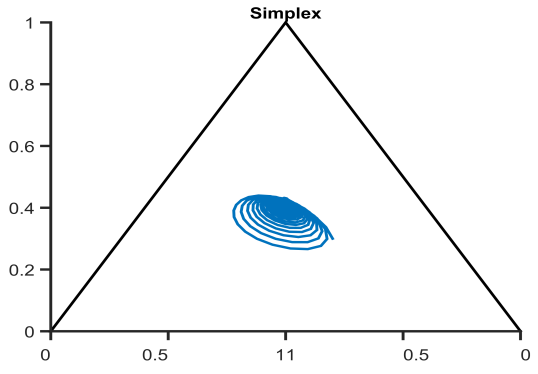
(a)



(b)



(c)



(d)

Figure 5: (a) Distributed + Entropy Rock-Paper-Scissors Behavior in a Simplex. (b) Evolution of the Population Behavior. (c) Simulation with τ =10. (d) Simulation with τ =100.

5.2 Solving maximization Problems

In this part of the document, we propose an application of the proposed method by understanding how it works under single and multi-population cases to solve maximization problems.

5.2.1 Single Population Case

This case considers a population where each agent can choose one of the $n+1$ strategies. In this case, the first n strategy corresponds to one variable of the objective function and the $n+1$ th strategy can be seen as a slack variable. Thus, x_k is the proportion of agents that use the *kth* strategy, and it corresponds to the *kth* variable, i.e., $x_k = z_k$. The fitness function of the *kth* strategy F_k is defined as the derivative of the objective function with respect to the *kth* variable, thus,

$$F_k(x) \equiv \frac{\partial}{\partial x_k} f(x)$$

Note that if $f(x)$ is a concave function, then its gradient is a decreasing function. As mentioned previously, users attempt to increase their fitness by adopting the most profitable strategy in the population, e.g. the *kth* strategy. This lead to an increase of x_k, which in turns decrease the fitness $F_k(x)$. Furthermore, the equilibrium is reached when all agents that belong to the same population have the same fitness. Thus, at equilibrium $F_i(x) = F_j(x)$, where $i, j \in \{1, \ldots, n\}$. If we define $F_{n+1}(x) = 0$, then, at equilibrium $F_i(x) = 0$ for every strategy $i \in \{1, \ldots, n\}$. Since the fitness function decreases with the action of users, it can be concluded that the strategy of the population evolves to make the gradient of the objective function equal to zero (or as close as possible). This resembles a gradient method to solve optimization problems. Recall that the evolution of the strategies lies in the simplex, that is, $\sum_{i \in S^p} z_i = m$, thence this implementation solves the following optimization problem:

$$\begin{aligned} \underset{z}{\text{maximize}} \quad & f(z) \\ \text{subject to} \quad & \sum_{i=1}^{n} z_i \leq m, \end{aligned} \tag{28}$$

where m is the total mass of the population.

Figure 6 shows an example of the setting described above for the function

$$f(z) = -(z_1 - 5)^2 - (z_2 - 5)^2. \tag{29}$$

Figure 6a shows the classic behavior to solve the maximization problem using a centralized approach. The simulation is executed during 0.6 time units. The black line founds the maximum with a very short deviation. Figure 6b depicts the case using a decentralized maximization approach. Once again, the maximum is reached but the deviation is bigger than the centralized approach. Finally, Figures 6c, 6d, 6e and 6f show the behavior of the Boltzmann-Based Distributed Replicator Dynamics, i.e. communication between agents is limited (Equation 21). In these cases, values of τ of 0.1, 0.5, 1, and 10 were used, respectively. Using the obtained model, it can be observed that as τ grows, the behavior of the simulation tends to be very similar to that of the centralized approach. Conversely, the shorter the τ value, the farther it is from the maximization point.

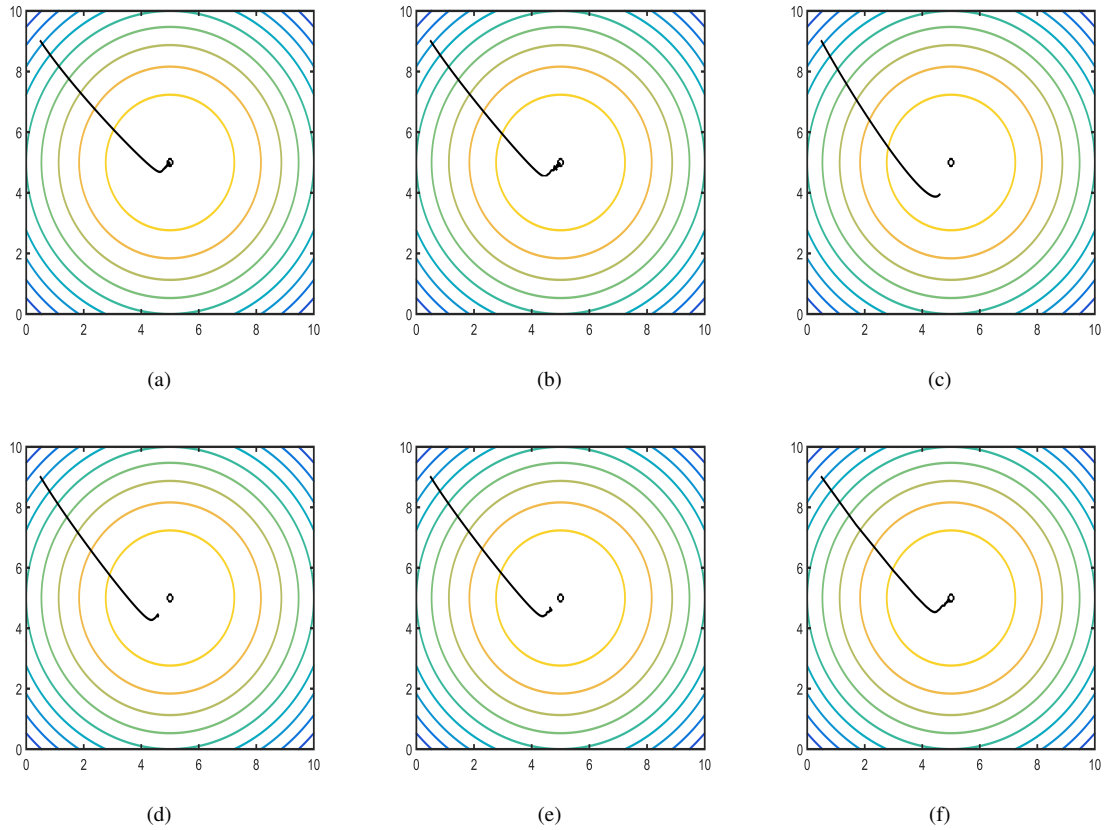


Figure 6: (a) Centralized Maximization Approach. (b) Decentralized Maximization Approach. (c) Distributed Maximization + Entropy Approach for τ =0.1. (d) Distributed Maximization + Entropy Approach for τ =0.5. (e) Distributed Maximization + Entropy Approach for τ =1. (f) Distributed Maximization + Entropy Approach for τ =10

5.2.2 *Multi Population Case*

Consider n populations where each agent can choose one out of two strategies. One population is defined per each variable of the maximization problem and also n additional strategies that resemble slack variables. Thus, x_i^p is the proportion of agents that use the *ith* strategy in the *pth* population. In this case x_1^k corresponds to the *kth* variable, that is, $x_1^k = z_k$, while x_2^k is a slack variable. The fitness function F_1^k of the *kth* population is defined as the derivative of the objective function with respect to the *kth* variable, that is, $F_1^k(x) \equiv \frac{\partial}{\partial x_1^k} f(x)$. Additionally, $F_2^k(x) = 0$. This implementation solves the following optimization problem:

$$\begin{aligned} &\underset{z}{\text{maximize}} && f(z) \\ &\text{subject to} && z_i \leq m^i, i = \{1, \ldots, n\}. \end{aligned} \tag{30}$$

Figure 7a shows the way the system gets to the maximum point using the centralized approach. Using a multi-population, the plotted line is made almost without deviations. Similarly, Figure 7b depicts the result for the classical distributed approach, where the multiple populations reach the maximum, but the following form has some deviations before reaching it. Figures 7c, 7d, 7e and 7f show the behavior of the Extended Distributed Replicator Dynamics (see full model of Equation (21)). In these cases, values of τ of 0.1, 0.5, 1, and 10 were used respectively. Results show once again, that using the Boltzmann-Based Distributed Replicator Dynamics method evidences that as τ grows, the behavior of the simulation tends to be very similar to that of the centralized approach (where full information is assumed within agents). Conversely, the shorter the φ value, the farther and the more deviant it is from the maximization point.

5.3 *Smart Grids Application*

This part of the paper presents how the use of the Boltzmann-based distributed replicator dynamics can be developed in a power grid. Some of the main issues to solve in these kinds of applications are cases of the economic dispatch problem (EDP). In these problems, first, it is necessary to reduce the global value of the power generation and, second, to maximize the overall effectiveness of the power generators, thus fulfilling the limitations of generation capacity and power balance simultaneously [38]. In this sense, traditional approaches to EDP have employed offline direct-search methods [21, 15], or static optimization algorithms [20]. One of the first works that introduced a different approach to deal with EDP is [39], where the authors proposed changing the resource allocation as a solution to this issue. Our work takes into account this approach and complements it with the introduction of the Boltzmann-based distributed replicator dynamics as a way to find the place to execute the dispatch algorithm at a microgrid, by using distributed population dynamics. Our work also assumes that loads, generators, and other devices in the grid share information in the system and have

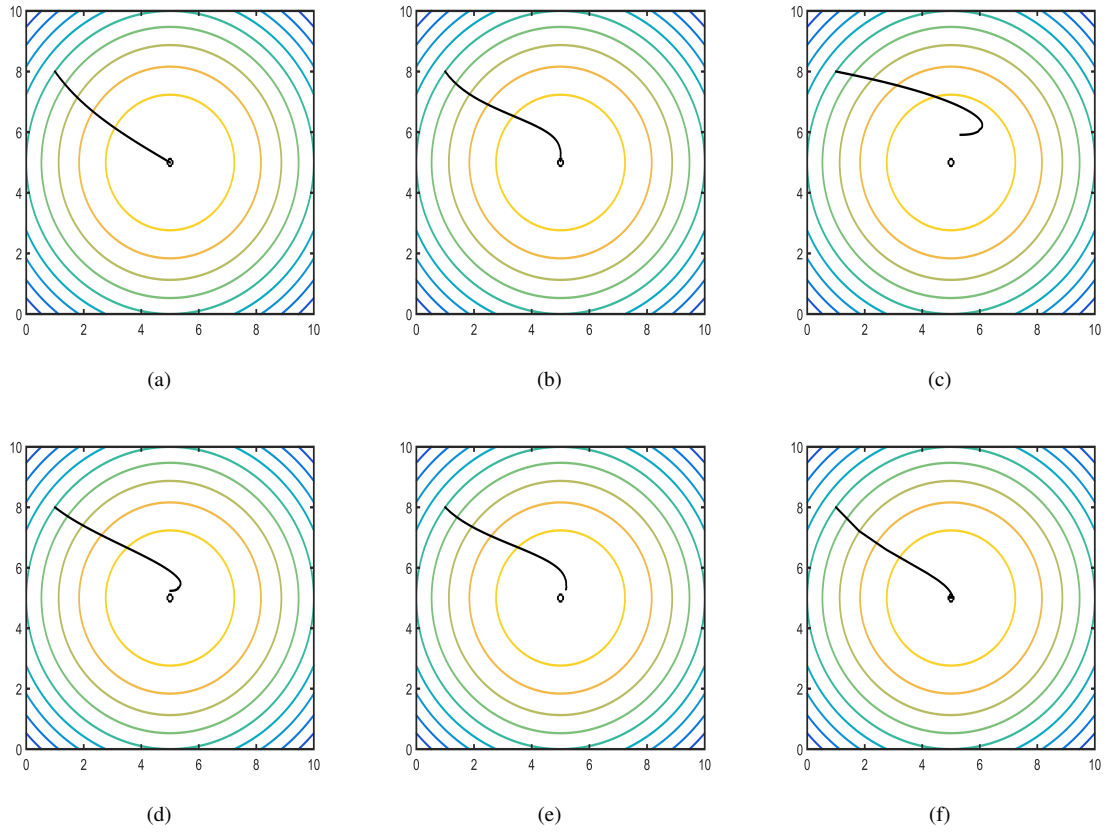


Figure 7: (a) Centralized Maximization Approach. (b) Decentralized Maximization Approach. (c) Distributed Maximization + Entropy Approach for τ =0.1. (d) Distributed Maximization + Entropy Approach for τ =0.5. (e) Distributed Maximization + Entropy Approach for τ =1. (f) Distributed Maximization + Entropy Approach for τ =10

a cooperative role with other controllable devices in the grid. The general case of the microgrid is explained in [40], where authors formulate a grid with two different control levels. At the lowest level, an inverter attaches loads to a source of voltage comprised of seven distributed generators (DGs). The output voltage and the operation frequency are controlled by a drop-gain regulator. Figure 8 depicts the distribution of the microgrid.

The uppermost level employs a strategy that can dynamically dispatch setpoints of power. The economic limitations, like load demands and power production costs, come from the inferior level of control and are directed to the central controller of the microgrid. Therefore, a classic RD is implemented. The controller obtains dynamic values of load demands and costs, which means that it is possible to include renewable energy resources. As a result, the dispatch is carried out , that is, the uppermost control level. The expression of the EDP is written as follows:

$$\begin{aligned} &\text{maximize} \quad J(\varphi) = \sum_{i=i}^{n} J_i(\tau_i), \\ &\text{subject to} \sum_{i=1}^{n} \varphi_i = \sum_{i=1}^{n} \psi_i = \varphi_D \end{aligned} \tag{31}$$

In Equation (31), $0 \leq \varphi_i \leq \varphi_{\max i}, \forall i \in \mathbb{Z}$, n represents the quantity of distributed generators, φ_i denotes the the i-th DG set-point of power, ψ_i simbolizes the loads, φ_D represents the total load that the grid requires, $\varphi_{\max}$ establishes the i-th DG maximum capacity of generation, and $J_i(\varphi_i)$ represents the utility function of every DG. The criterion of the economic dispatch determines the utility function [38], which in turn settles the performance of all the generation units with the same marginal utilities stated in Equation (32)

$$\frac{dJ_1}{d\varphi_1} = \frac{dJ_2}{d\varphi_2} = ... = \frac{dJ_n}{d\varphi_n} = \delta, \tag{32}$$

Consider $\delta > 0$, so that $\sum_{i=1}^{n} \varphi_i = \varphi_D$. According to the EDP criterion expressed in Equation (32), it is possible that the EDP of Equation (31) obtain a solution by employing utility functions with quadratic form for every DG [39].

5.3.1 *The Economic Dispatch Problem Using a Population Games Perspective*

From the Population Games Perspective, the EDP can be managed using the Replicator Dynamics approach. For the simulation purposes, we limit the communications constraints among agents at random, which allows us to have another point of view to compare results with those obtained in [1]. Using the population games approach, n represents the quantity of DGs in the grid. Consider the selection of a DG as the i-th strategy, then, φ_i would be the amount of power allocated to each DG, which is associated with the number of players that choose the i-th strategy in S. The term φ_D represents the sum of every power set-point, which means $\sum_{i=1}^{n} \varphi_i = \varphi_D$ to obtain an appropriate steady-state performance. Likewise, to accomplish

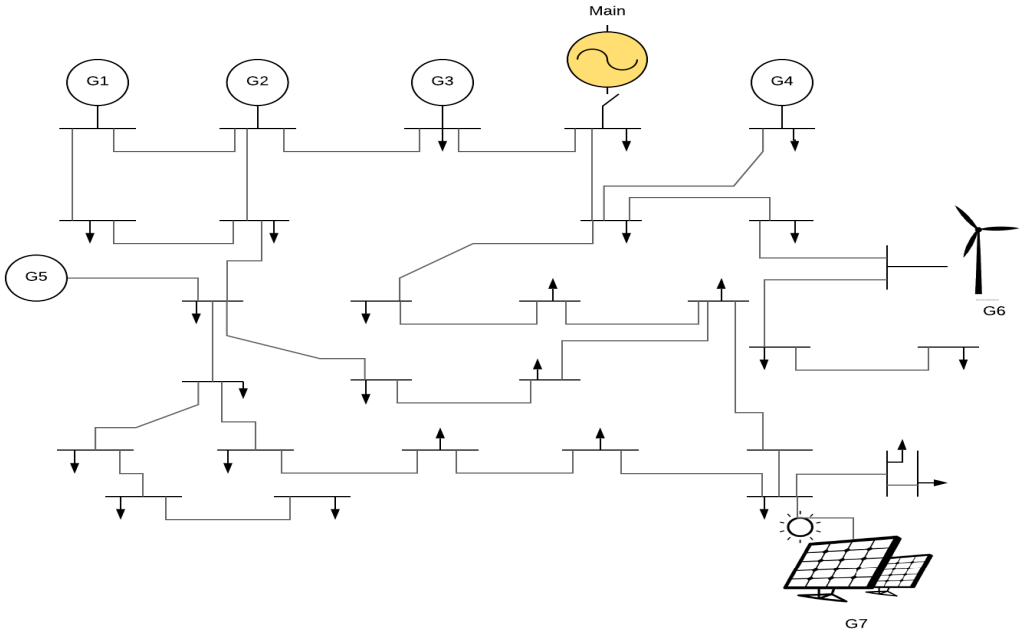


Figure 8: Distribution of a Power Grid. Adapted from [11].

the power balance, $\bar{f} = (1/\varphi_D) \sum_{i=1}^{n} \varphi_i f_i$ must be implemented to enable the invariance of set $\Delta = [\varphi \epsilon^t \geq 0 : \sum_{i \epsilon S} \varphi i = m]$ [41]. This equation ensures that in case $\varphi(0) \epsilon \Delta$, then $\varphi(t) \epsilon \Delta, \forall t \geq 0$, that is, the strategy of control have to determine set-points to guarantee the correct equilibrium between the demanded and generated power by generators. This behavior makes it possible to perform a proper control of frequency. To include in the control strategy economic and technical criteria, the capacity of power generation and the associated cost are relevant factors for determining the final power dispatched to every DG. RD seems to be suitable, since its stationary state is achieved once the average outcome equals all the outcome functions. This characteristic relates RD to EDP, as it is the same as the economic dispatch criterion of Equation (32) when the outcome function is chosen as follows:

$$f_i(\varphi_i) = \frac{dJ_i}{d\varphi_i}, \forall i = 1, 2, ..., n, \tag{33}$$

It is worth noting that the EDP approach in Equation (32) ensures an optimal solution of the system if constraints are satisfied. These kinds of optimization issues may be tackled using marginal utilities for the outcome functions. This is possible because the outcome functions are equal to $\bar{f}$. The outcome chosen can be modeled as an expression whose growth/reduction depends on the distance of the desired set-point from/to the power. In this vein, RD allocate resources to generators according to the average result. The following function [42] can illustrate this phenomenon:

$$f(\varpi) = r\varpi\left(1 - \frac{\varpi}{k}\right) \tag{34}$$

where k represents the carrying capacity so that the independent variable $\varpi \epsilon (0, K)$. Here, parameters such as the carrying capacity and a cost factor of generation, among other parameters, are used by the outcome functions. As a result, the outcome function of each DG can be expressed as:

$$f_i(\varphi_i) = \frac{dJ_i}{d\varphi_i} = \frac{2}{ci}\left(1 - \frac{\varphi i}{\varphi_{max}}\right), \forall i = 1, 2, ..., n, \tag{35}$$

The population game can transform into a potential game by the addition of marginal utilities to the outcome functions [43]. The outcome functions in Equation (35) become functions of quadratic utility for every DG in the optimal EDP [39]. This outcome function has been implemented in other research, e.g. [39, 40, 44].

$$J_i(\varphi_i) = \frac{1}{ci}\left(2\varphi_i - \frac{\varphi i^2}{\varphi_{maxi}}\right), \quad \forall i = 1, 2, ..., n, \tag{36}$$

5.4 Simulation Results

The BBDRD control model presented in this paper is validated in a study case that considered a low voltage smart grid comprising seven DGs. The system used $\varphi_D = 9$ kW as the overall power demand in the network; DG 4 had the lowest cost and DG 7 the highest. DGs 1, 3, 5, and 6 had no significant differences in cost, and DG 2 had their lowest cost. The system employs 60 Hz and a nominal capacity of 3.6 kW for all generators, except for DGs 2 and 6 that employ 1.5 KW and 4 KW, respectively. Note that these initial conditions differ from those used in [1], where DG 3 had the lowest cost and the nominal capacity of DGs 2 and DG 6 was 3.6 KW and 2 KW, respectively.

For comparison purposes, first, the classic centralized case was simulated, taking into account the availability of full information. Figure 9a shows the results of this step. There is an unexpected rise in the load of 3 KW and various values for each generator. The frequency was stable, except for t = 0.8, where there is a variation of approximately 0.2 Hz produced by an increase in the load,

however, it returns to stability right after it. Figure 9a also depicts the quantity of power delivered to each DG. First, generator DG 7 transmits a minimum power when there is low demand, owing to its costly behavior. On the contrary, DG 3 approximates to its maximum capacity and remains near this value without affectations by changes in the load. In case that the demand augments, DG 7 augments its capacity too, intending to counter-weigh the demand. DG 6 approximated to its maximum capacity just after changes in the load. DGs 1, 4, 5, and 6 evidence a comparable behavior, since they present analogous conditions. Finally, DG 2 approximates to its maximum performance thanks to its low-cost performance. With the results of the classic centralized case, the BBDRD control method was employed to contrast its behavior. Figure 9b–d present the outcomes of this step for different values of τ. For simulation purposes, we use constraints in the communication of agents at random. When the τ value augments, the system imitates the centralized approach. Low values of τ (Figure 9b) produced the biggest differences, as DGs need more time to achieve their working level. On the contrary, high values of τ (Figure 9d) allow DGs to achieve their working levels faster. Concerning the results obtained in [1], we observe a similar behavior of the microgrid. Despite using limitations in the communications of the generators at random, after employing the BBDRD method with high values of τ, once more the behavior tends to be equal to that of the centralized approach of the classic RD. The main difference evidenced is the behavior of DG4 in comparison with the centralized approach, that is, when the demand augments, it delivers more power as a result of the communication limitations topology and the effect of the exploration concept (second term in Equation (21). This effect was also evidenced in [1] with the behavior of DG5.

6 Conclusions

The Boltzmann-based distributed replicator dynamics shown in Equation (21) might be defined as a learning method of distributed control that includes the exploration scheme from RL in the classic equation of RD. In this sense, exploration can be related to the mutation concept of EGT, and involves a method for measuring variety in the system with the entropy approach. The Boltzmann-based distributed replicator dynamics also employs the scheme of the Boltzmann distribution to include the τ parameter for controlling purposes. An appropriate temperature function can be chosen using methodological search and reliably set to fulfill an anticipated convergence distribution. Regarding stability, Section 3 presents a derivation process that has low or no significant variations in the presence of multiple agents. This behavior is explained with the inclusion of the population approach in the BBDRD method. The neighboring approach provides the missing piece to prevent centralized schemes from happening and compels players to consider just the available information of other players before performing an action. The method was validated in the context of classic games, maximization problems, and in a smart grid that allowed initializing parameters beforehand, and providing evidence that behavior using the BBDRD approach tends to be similar to cases using centralized schemes.

Engineering problems represent real scenarios whose complexity can be simulated using MAS, through the analysis of the communication between the agents. EGT presents some helpful tools to tackle communication between players and control them. This paper evidences the advantages of applying a distributed control approach of EGT to a real-life smart grid. The BBDRD performance is presented using experiments that include limitations in communication, therefore, it emerges as a helpful tool for developing more realistic control strategies in Engineering problems with distributed schemes. This advantage becomes particularly relevant because it offers the opportunity to deal with complex systems using local information of the agents, taking into account communication limitations without the need of a centralized coordinator and evading expensive implementation costs, as in classic approaches, like the dual decomposition method. The distributed control concept proposed to tackle cases of classic games, maximization issues, and the Economic Dispatch Problem can be further applied to other real-life situations, including some other problems in the smart grid context, like as the physical limits of power-flow, the presence of power losses, and the inconsistency of renewable generation, among others.

Despite using incomplete information, results demonstrated that the system can imitate the performance of a centralized approach when the τ value increases. conversely, when τ takes values lower than the unit, the behavior was distant from outcomes obtained under the optimal communication scenario of a centralized approach. The possibility of adjusting the behavior and parameters of the method using communications limitations between players proved to be successful. This can also logically be extended to any number of players or populations. Results also evidenced that the Boltzmann-based distributed method has adequate performance for solving some cases of maximization problems, including the economic dispatch problem in a smart grid. This is possible since the features of the DGs were coherent with their power capacity and operation cost.

7 Future Work

For future work, the optimization of wireless sensor networks can be an option for the building automation field. Various critical issues may be tackled by implementing the distributed replicator dynamics approach to solve the EDP in a smart grid scenario, for example, considering power losses, the limitations of physical power-flow, or the uncertainty of renewable generation. Open issues should be considered with respect to the control strategies developed that must include decentralization, scalability, and robustness. In consequence, novel methods should incorporate economic incentives and the information necessary to ensure that more elements can be included in the system without a reconfiguration of the whole system. The Boltzmann-based distributed concept that tackles EDP will be expanded to other problems in the context of a smart grid framework, for example, the inconsistency in renewable generation, physical limitations of power-flow, and incorporation of power loss. In summary, distributed techniques used to manage open problems represent a suitable option for modeling the complexity of these scenarios. Innovative approaches are still required to include scalable solutions and features closer to reality.

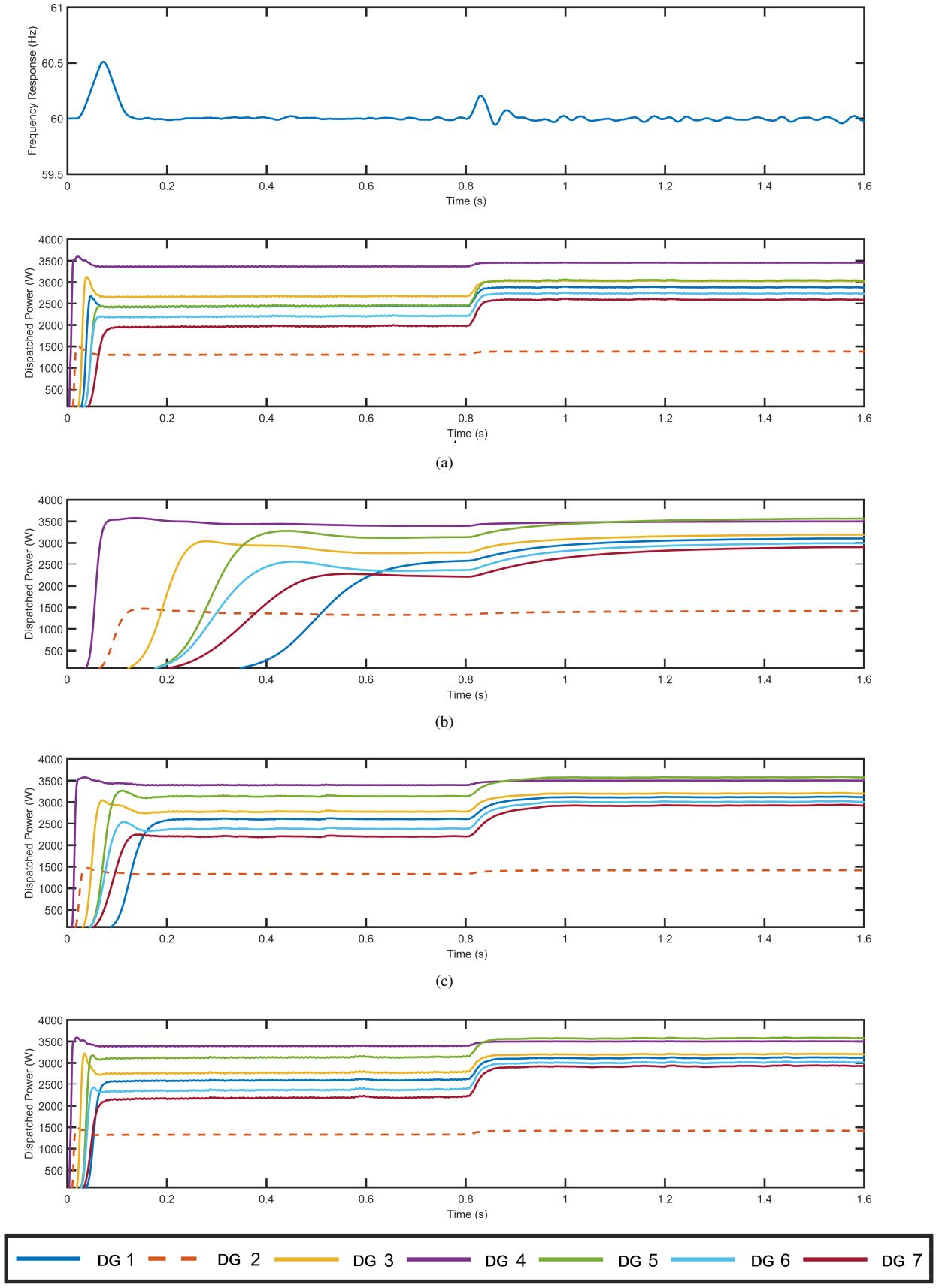


Figure 9: Results for a microgrid system. (**a**) Frequency response and active power response of DGs for the classic RD. The analysis of the performance of the Boltzmann-based distributed replicator dynamics for different values of τ: (**b**) $\tau = 0.4$ (**c**) $\tau = 2.5$, (**d**) $\tau = 7$.

Conflict of Interest The authors declare no conflict of interest.

Funding This research was funded by Colciencias, grant Doctorado Nacional number 727.

Acknowledgment We express our gratitude to Colciencias for the founding of this Project. We thank Universidad Nacional de Colombia and Universidad Santo Tomás for allowing us to use laboratories, hardware and software for the development of this work.

Annex A

In this part of the document, we reconstruct the full process of derivation necessary to have a continuous-time limit for the model of Q-learning, where the Q-values are considered as Boltzmann probabilities for action-selection mechanisms. For clarity purposes in the construction of the learning model, this analysis starts considering an extended version of the equations obtained in [35], where dynamics for the Q-learners in two-players games were defined.

To find the relationship between the Q-learning framework and the RD equations, the use of the Equation (1) that describes the Boltzmann probabilities is done.

$$x_i(\delta) = \frac{e^{\tau Q_{a_i}(\delta)}}{\sum_{j=1}^{n} e^{\tau Q_{a_j}(\delta)}} \tag{1}$$

Here, $x_i(\delta)$ represents the prospect of selecting the i strategy at δ step time, and τ symbolizes the temperature. From the Boltzmann distribution, it is easy to find the expression for $x_i(\delta+1)$ as follows:

$$x_i(\delta+1) = \frac{e^{\tau Q_{a_i}(\delta+1)}}{\sum_{j=1}^{n} e^{\tau Q_{a_j}(\delta+1)}}$$

now dividing $x_i(\delta+1)$ into $x_i(\delta)$:

$$\frac{x_i(\delta+1)}{x_i(\delta)} = \frac{e^{\tau Q_{a_i}(\delta+1)} \sum_{j=1}^{n} e^{\tau Q_{a_j}(\delta)}}{e^{\tau Q_{a_i}(\delta)} \sum_{j=1}^{n} e^{\tau Q_{a_j}(\delta+1)}}$$

after organizing terms it gets to:

$$\frac{x_i(\delta+1)}{x_i(\delta)} = \frac{e^{\tau Q_{a_j}(\delta+1)} e^{-\tau Q_{a_i}(\delta)}}{\sum_{j=1}^{n} e^{\tau Q_{a_j}(\delta+1)} \sum_{j=1}^{n} e^{-\tau Q_{a_j}(\delta)}}$$

Then, using Δ to denote a small difference between operations it takes the following form:

$$\frac{x_i(\delta+1)}{x_i(\delta)} = \frac{e^{\tau \Delta Q_{a_i}(\delta)}}{\sum_{j=1}^{n} e^{\tau \Delta Q_{a_j}(\delta)}}$$

This result can be rewritten in the following way:

$$x_i(\delta+1) = x_i(\delta) \frac{e^{\tau \Delta Q_{a_i}(\delta)}}{\sum_{j=1}^{n} x_j e^{\tau \Delta Q_{a_j}(\delta)}}$$

Now, considering the difference equation for x_i:

$$\begin{aligned} x_i(\delta+1) - x_i(\delta) &= \frac{x_i(\delta) e^{\tau \Delta Q_{a_i}(\delta)}}{\sum_{j=1}^{n} x_j(\delta) e^{\tau \Delta Q_{a_j}(\delta)}} - x_i(\delta) \\ &= x_i(\delta) \Big[\frac{e^{\tau \Delta Q_{a_i}(\delta)} - \sum_{j=1}^{n} x_j(\delta) e^{\tau \Delta Q_{a_j}(\delta)}}{\sum_{j=1}^{n} x_j(\delta) e^{\tau \Delta Q_{a_j}(\delta)}} \Big] \end{aligned}$$

At this point, to describe the continuous time version, it is assumed that σ, with $0 < \sigma \leq 1$, describes the time amount spent between game repetitions. In the case of $x_i(\delta\sigma)$, it represents the x-values at time $k\sigma = t$. Under these premises, the expression takes the following form:

$$\begin{aligned} \frac{x_i(\delta\sigma+\sigma) - x_i(\delta\sigma)}{\sigma} =& \Big[\frac{x_i(\delta\sigma)}{\sigma \sum_{j=1}^{n} x_j(\delta\sigma) e^{\tau \Delta Q_{a_j}(\delta\sigma)}} \Big] * \\ & \Big[e^{\tau \Delta Q_{a_j}(\delta\sigma)} - \sum_{j=1}^{n} x_j(\delta\sigma) e^{\tau \Delta Q_{a_j}(\delta\sigma)} \Big] \end{aligned}$$

Nevertheless, the main interest is finding the limit of $x_i(\delta\sigma)$, given $\sigma \to 0$, $\delta\sigma \to t$ and $t \geq 0$, then:

$$\begin{aligned} \lim_{\sigma\to\infty} \frac{\Delta x_i(\delta\sigma)}{\sigma} = \lim_{\sigma\to 0} \Big[& \Big(\frac{x_i(\delta\sigma)}{\sigma \sum_{j=1}^{n} x_j(\delta\sigma) e^{\tau \Delta Q_{a_j}(\delta\sigma)}} \Big) * \\ & \Big(e^{\tau \Delta Q_{a_j}(\delta\sigma)} - \sum_{j=1}^{n} x_j(\delta\sigma) e^{\tau \Delta Q_{a_j}(\delta\sigma)} \Big) \Big] \end{aligned}$$

This expression can be rewritten as follows:

$$\begin{aligned} \lim_{\sigma\to\infty} \frac{\Delta x_i(\delta\sigma)}{\sigma} = & \lim_{\sigma\to 0} \Big[\frac{x_i(\delta\sigma)}{\sum_{j=1}^{n} x_j(\delta\sigma) e^{\tau \Delta Q_{a_j}(\delta\sigma)}} \Big] * \\ & \lim_{\sigma\to 0} \Big[\frac{e^{\tau \Delta Q_{a_j}(\delta\sigma)}}{\sigma} - \frac{\sum_{j=1}^{n} x_j(\delta\sigma) e^{\tau \Delta Q_{a_j}(\delta\sigma)}}{\sigma} \Big] \end{aligned}$$

In the first limit, the expression $e^{\tau \Delta Q_{a_j}(\delta\sigma)}$ is equal to 0, and the summation becomes 1 because it is referred to the sum of all the probabilities. This means that the first limit becomes x_i.

$$\lim_{\sigma\to\infty} \frac{\Delta x_i(\delta\sigma)}{\sigma} = x_i * \underbrace{\lim_{\sigma\to 0} \Big[\frac{e^{\tau \Delta Q_{a_j}(\delta\sigma)}}{\sigma} - \frac{\sum_{j=1}^{n} x_j(\delta\sigma) e^{\tau \Delta Q_{a_j}(\delta\sigma)}}{\sigma} \Big]}_{L2}$$

In the second limit, an undefined situation is presented; the numerator and denominator become zero, therefore, after using l'hopital rule, this limit equals (for short *L2*):

$$L2 = \lim_{\sigma\to\infty} \Big[\frac{\tau \Delta Q_{a_j}(\delta\sigma) e^{\tau \Delta Q_{a_j}(\delta\sigma)}}{\sigma} \Big] - \sum_{j=1}^{n} x_j(\delta\sigma) * \lim_{\sigma\to\infty} \Big[\tau \Delta Q_{a_j}(\delta\sigma) \frac{e^{\tau \Delta Q_{a_j}(\delta\sigma)}}{\sigma} \Big]$$

Which allow finding the following expression:

$$L2 = \tau \frac{dQ_{a_i}(t)}{dt} - \sum_{j=1}^{n} x_j(t) \frac{dQ_{a_j}(t)}{dt}$$

Now, it is possible to find the total limit, that is, the Q-Learning continuous time model derived as shown in Equation (2):

$$\frac{\frac{dx_i}{d_t}}{x_i} = \tau \Big[\frac{dQ_{a_i}}{dt} - \sum_{j=1}^{n} \frac{dQ_{a_j}}{dt} x_j \Big] \tag{2}$$

To solve the expression $\frac{dQ_{a_i(t)}}{dt}$, the first player takes the following update rule:

$$Q_{a_i}(\delta+1) = Q_{a_i}(\delta) + \alpha\Big[\Gamma_{a_i}(\delta+1) + \gamma \max_{ai} Q - Q_{a_i}(\delta)\Big]$$

Therefore, the last expression represents the equation of difference for the Q-function and can be rewritten as follows:

$$\Delta Q_{a_i}(\delta) = \alpha\Big[\Gamma_{a_i}(\delta+1) + \gamma \max_{ai} Q - Q_{a_i}(\delta)\Big] \tag{3}$$

if Equation (3) takes an infinitesimal scheme, it is supposed that the amount of time spent performing two update iterations of the Q-values is given by σ with $0 < \sigma \leq 1$. Additionally, $Q_{a_i}(\delta\sigma)$ symbolizes the Q-values at time $\delta\sigma$. Applying these assumptions, Equation (3) gets to:

$$\Delta Q_{a_i}(\delta\sigma) = \Big[\alpha(\Gamma_{a_i}((\delta+1)\sigma) + \gamma \max_{ai} Q - Q_{a_i}(\delta\sigma)\Big] * \Big[(\delta+1)\sigma - \delta\sigma\Big]$$

which is equal to:

$$\Delta Q_{a_i}(\delta\sigma) = \alpha\sigma\Big[\Gamma_{a_i}((\delta+1)\sigma) + \gamma \max_{ai} Q - Q_{a_i}(\delta\sigma)\Big]$$

Once again, the limit $\sigma \to 0$ is the state sought. Taking the limit of $Q_{a_i}(\delta\sigma)$, it gets to Equation (4):

$$\frac{dQ_{a_i}}{dt} = \alpha\Big[\Gamma_{a_i} + \gamma \max_{ai} Q - Q_{a_i}\Big] \tag{4}$$

Now, substituting Equation (4) on Equation (2):

$$\frac{\frac{dx_i}{dt}}{x_i} = \tau\Big[\alpha\Gamma_{a_i} + \alpha\gamma \max_{ai} Q - \alpha Q_{a_i} - \sum_{j}^{n} x_j \alpha(\Gamma_{a_j} + \gamma \max_{ai} Q_{a_i} - Q_{a_j})\Big]$$

$$= \tau\alpha\Big[\Gamma_{a_i} - \sum_{j=1}^{n} x_j\Gamma_{a_j} - Q_{a_i} + \sum_{j=1}^{n} Q_{a_j}x_j\Big]$$

Taking into account that $\sum_j^n x_j = 1$ and using $\dot{x}_i$ to denote $\frac{dx_i}{dt}$, it is obtained:

$$\frac{\dot{x}_i}{x_i} = \tau\alpha\Big[\Gamma_{a_i} - \sum_{j=1}^{n} x_j\Gamma_{a_j} - Q_{a_i}\sum_{j=1}^{n} x_j + \sum_{j=1}^{n} Q_{a_j}x_j\Big]$$

$$\frac{\dot{x}_i}{x_i} = \tau\alpha\Big[\Gamma_{a_i} - \sum_{j=1}^{n} x_j\Gamma_{a_j} + \sum_{j=1}^{n} x_j(Q_{a_j} - Q_{a_i})\Big]$$

since $\frac{x_j}{x_i}$ equals $\frac{e^{\tau}\Delta Q_{a_j}}{e^{\tau}\Delta Q_{a_i}}$, the second part of the last expression can be expressed in logarithm terms:

$$\alpha\Big[\sum_{j=1}^{n} x_j ln(\frac{x_j}{x_i})\Big] = \alpha\Big[\tau\sum_{j} x_j(Q_{a_j} - Q_{a_i})\Big]$$

After reorganizing and substituting, the result is:

$$\frac{\dot{x}_i}{x_i} = \alpha\tau\Big[\Gamma_{a_i} - \sum_{j=1}^{n} x_j\Gamma_{a_j}\Big] + \alpha\Big[\sum_{j=1}^{n} x_j \ln\left(\frac{x_j}{x_i}\right)\Big]$$

To bring the concept of the payoff matrices into a 2 x 2 game, it can be expressed r_{a_i} as $\sum_j a_{ij}y_j$, thus obtaining the Equation (5) which represents the behavior for the first player as follows:

$$\dot{x}_i = x_i\alpha\tau\Big[(A\mathbf{y})_i - \mathbf{x}\cdot A\mathbf{y}\Big] + x_i\alpha\Big[\sum_{j=1}^{n} x_j \ln\left(\frac{x_j}{x_i}\right)\Big] \tag{5}$$

Similarly, for the second player, the expression is:

$$\dot{y}_i = y_i\alpha\tau\Big[(B\mathbf{x})_i - \mathbf{y}\cdot B\mathbf{x}\Big] + y_i\alpha\Big[\sum_{j=1}^{n} y_j \ln\left(\frac{y_j}{y_i}\right)\Big] \tag{6}$$

Since the approach of the classic RD can be used at this point, Equation (5) may be stated as shown in the following expression [27]:

$$\dot{x}_i = \alpha x_i\tau\Big[f_i(x) - \overline{f}(x)\Big] + \alpha x_i\Big[\sum_{j=1}^{n} x_j \ln\left(\frac{x_j}{x_i}\right)\Big]$$

It should be noted that depending on the value obtained from the fitness of a specific type of population, this value may increase or decrease depending on the average value obtained by the entire population.

References

[1] G. Chica, E. Mojica, E. Cadena, "Boltzmann-Based Distributed Replicator Dynamics: A Smart Grid Application," in 2020 Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI), 1–6, IEEE, 2020, doi: 10.1109/CONIITI51147.2020.9240335.

[2] J. Barreiro-Gomez, G. Obando, N. Quijano, "Distributed population dynamics: Optimization and control applications," IEEE Transactions on Systems, Man, and Cybernetics: Systems, **47**(2), 304–314, 2016, doi:10.1109/TSMC.2016.2523934.

[3] G. Chica-Pedraza, E. Mojica-Nava, E. Cadena-Muñoz, "Boltzmann Distributed Replicator Dynamics: Population Games in a Microgrid Context," Games, **12**(1), 1–1, 2021, doi:10.3390/g12010008.

[4] G. Bacci, S. Lasaulce, W. Saad, L. Sanguinetti, "Game theory for networks: A tutorial on game-theoretic tools for emerging signal processing applications," IEEE Signal Processing Magazine, **33**(1), 94–119, 2015, doi: 10.1109/MSP.2015.2451994.

[5] C. Mu, K. Wang, "Approximate-optimal control algorithm for constrained zero-sum differential games through event-triggering mechanism," Nonlinear Dynamics, **95**(4), 2639–2657, 2019, doi:10.1007/s11071-018-4713-0.

[6] M. Zhu, E. Frazzoli, "Distributed robust adaptive equilibrium computation for generalized convex games," Automatica, **63**, 82–91, 2016, doi: 10.1016/j.automatica.2015.10.012.

[7] S. Najeh, A. Bouallegue, "Distributed vs centralized game theory-based mode selection and power control for D2D communications," Physical Communication, **38**, 100962, 2020, doi:https://doi.org/10.1016/j.phycom.2019.100962.

[8] R. Tang, S. Wang, H. Li, "Game theory based interactive demand side management responding to dynamic pricing in price-based demand response of smart grids," Applied Energy, **250**, 118–130, 2019, doi:10.1016/j.apenergy.2019.04.177.

[9] K. Främling, "Decision theory meets explainable ai," in International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems, 57–74, Springer, 2020, doi:https://link.springer.com/chapter/10.1007/978-3-030-51924-7_4.

[10] A. Navon, G. Ben Yosef, R. Machlev, S. Shapira, N. Roy Chowdhury, J. Belikov, A. Orda, Y. Levron, "Applications of Game Theory to Design and Operation of Modern Power Systems: A Comprehensive Review," Energies, **13**(15), 3982, 2020, doi:10.3390/en13153982.

[11] N. Quijano, C. Ocampo-Martinez, J. Barreiro-Gomez, G. Obando, A. Pantoja, E. Mojica-Nava, "The role of population games and evolutionary dynamics in distributed control systems: The advantages of evolutionary game theory," IEEE Control Systems Magazine, **37**(1), 70–97, 2017, doi: 10.1109/MCS.2016.2621479.

[12] M. Lanctot, E. Lockhart, J.-B. Lespiau, V. Zambaldi, S. Upadhyay, J. Pérolat, S. Srinivasan, F. Timbers, K. Tuyls, S. Omidshafiei, et al., "OpenSpiel: A framework for reinforcement learning in games," arXiv preprint arXiv:1908.09453, 2019.

[13] W. H. Sandholm, Population games and evolutionary dynamics, MIT press, 2010.

[14] L. Hindersin, B. Wu, A. Traulsen, J. García, "Computation and simulation of evolutionary Game Dynamics in Finite populations," Scientific reports, **9**(1), 1–21, 2019, doi:https://doi.org/10.1038/s41598-019-43102-z.

[15] D. P. Palomar, M. Chiang, "A tutorial on decomposition methods for network utility maximization," IEEE Journal on Selected Areas in Communications, **24**(8), 1439–1451, 2006, doi:10.1109/JSAC.2006.879350.

[16] J. R. Marden, "State based potential games," Automatica, **48**(12), 3075–3088, 2012, doi:10.1016/j.automatica.2012.08.037.

[17] L. Zhao, J. Wang, J. Liu, N. Kato, "Optimal edge resource allocation in IoT-based smart cities," IEEE Network, **33**(2), 30–35, 2019, doi:10.1109/MNET.2019.1800221.

[18] A. Cagnano, E. De Tuglie, P. Mancarella, "Microgrids: Overview and guidelines for practical implementations and operation," Applied Energy, **258**, 114039, 2020, doi:10.1016/j.apenergy.2019.114039.

[19] J. P. Lopes, C. Moreira, A. Madureira, "Defining control strategies for microgrids islanded operation," IEEE Transactions on power systems, **21**(2), 916–924, 2006, doi:10.1109/TPWRS.2006.873018.

[20] T. Ibaraki, N. Katoh, Resource allocation problems: algorithmic approaches, MIT press, 1988.

[21] S.-J. Ahn, S.-I. Moon, "Economic scheduling of distributed generators in a microgrid considering various constraints," in 2009 IEEE Power & Energy Society General Meeting, 1–6, IEEE, 2009, doi:10.1109/PES.2009.5275938.

[22] G. Strbac, "Demand side management: Benefits and challenges," Energy policy, **36**(12), 4419–4426, 2008, doi:10.1016/j.enpol.2008.09.030.

[23] D. E. Olivares, C. A. Cañizares, M. Kazerani, "A centralized optimal energy management system for microgrids," in 2011 IEEE Power and Energy Society General Meeting, 1–6, IEEE, 2011, doi:10.1109/PES.2011.6039527.

[24] P. Quintana-Barcia, T. Dragicevic, J. Garcia, J. Ribas, J. M. Guerrero, "A distributed control strategy for islanded single-phase microgrids with hybrid energy storage systems based on power line signaling," Energies, **12**(1), 85, 2019, doi:10.3390/en12010085.

[25] B. Huang, L. Liu, H. Zhang, Y. Li, Q. Sun, "Distributed optimal economic dispatch for microgrids considering communication delays," IEEE Transactions on Systems, Man, and Cybernetics: Systems, **49**(8), 1634–1642, 2019, doi:10.1109/TSMC.2019.2900722.

[26] J. C. Vasquez, J. M. Guerrero, J. Miret, M. Castilla, L. G. De Vicuna, "Hierarchical control of intelligent microgrids," IEEE Industrial Electronics Magazine, **4**(4), 23–29, 2010, doi:10.1109/MIE.2010.938720.

[27] D. Bloembergen, K. Tuyls, D. Hennes, M. Kaisers, "Evolutionary dynamics of multi-agent learning: A survey," Journal of Artificial Intelligence Research, **53**, 659–697, 2015, doi:10.1613/jair.4818.

[28] H. Peters, Game theory: A Multi-leveled approach, Springer, 2015.

[29] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction, MIT press, 2018.

[30] W. Ertel, "Reinforcement Learning," in Introduction to Artificial Intelligence, 289–311, Springer, 2017.

[31] F. L. Da Silva, A. H. R. Costa, "A survey on transfer learning for multiagent reinforcement learning systems," Journal of Artificial Intelligence Research, **64**, 645–703, 2019, doi:10.1613/jair.1.11396.

[32] T. Başar, G. Zaccour, Handbook of Dynamic Game Theory, Springer, 2018.

[33] J. Newton, "Evolutionary game theory: A renaissance," Games, **9**(2), 31, 2018, doi:10.3390/g9020031.

[34] J. W. Weibull, Evolutionary game theory, MIT press, 1997.

[35] K. Tuyls, K. Verbeeck, T. Lenaerts, "A selection-mutation model for q-learning in multi-agent systems," in Proceedings of the second international joint conference on Autonomous agents and multiagent systems, 693–700, 2003, doi: 10.1145/860575.860687.

[36] A. E. Eiben, J. E. Smith, Introduction to Evolutionary Computing, Springer, 2015.

[37] D. Stauffer, "Life, love and death: Models of biological reproduction and aging," Institute for Theoretical physics, Köln, Euroland, 1999.

[38] W. Aj, B. Wollenberg, "Power generation, operation and control," New York: John Wiley & Sons, 592, 1996.

[39] A. Pantoja, N. Quijano, "A population dynamics approach for the dispatch of distributed generators," IEEE Transactions on Industrial Electronics, **58**(10), 4559–4567, 2011, doi:10.1109/TIE.2011.2107714.

[40] E. Mojica-Nava, C. A. Macana, N. Quijano, "Dynamic population games for optimal dispatch on hierarchical microgrid control," IEEE Transactions on Systems, Man, and Cybernetics: Systems, **44**(3), 306–317, 2013, doi: 10.1109/TSMCC.2013.2266117.

[41] J. Hofbauer, K. Sigmund, et al., Evolutionary games and population dynamics, Cambridge university press, 1998.

[42] N. F. Britton, Essential mathematical biology, Springer Science & Business Media, 2012.

[43] H. P. Young, S. Zamir, "Handbook of Game Theory with Economic Applications," Technical report, Elsevier, 2015.

[44] E. Mojica-Nava, C. Barreto, N. Quijano, "Population games methods for distributed control of microgrids," IEEE Transactions on Smart Grid, **6**(6), 2586–2595, 2015, doi:10.1109/TSG.2015.2444399.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 212-222 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Initial Experiments using Game-based Learning Applied in a Classical Knowledge Robotics in In-Person and Distance Learning Classroom

Márcio Mendonça[*,1], Rodrigo Henrique Cunha Palácios[1,2], Ivan Rossato Chrun[3], Diene Eire de Mello[4], Henrique Cavalieri Agonilha[5], Elpiniki Papageorgiou[6], Konstantinos Papageorgiou[6]

[1]*Programa de Pós-Graduação em Engenharia Mecânica (PPGEM-CP), Universidade Tecnológica Federal do Paraná, Av. Alberto Carazzai, 1640, Cornélio Procópio, Brazil*

[2]*Programa de Pós-Graduação em Informática (PPGI-CP), Universidade Tecnológica Federal do Paraná, Av. Alberto Carazzai 1640, Cornélio Procópio, Brazil*

[3]*Departamento de Engenharia Química, Universidade Estadual de Maringá, Av. Colombo 5790, Maringá, Brazil*

[4]*Programa de Pós-Graduação em Educação (PPEDU- UEL), Universidade Estadual de Londrina, Av. Colombo 5790, Maringá, Brazil*

[5]*Departamento de Acadêmico de Computação (DACOM), Universidade Tecnológica Federal do Paraná, Av. Alberto Carazzai 1640, Cornélio Procópio, Brazil*

[7]*Department of Energy Systems, University of Thessaly, Geopolis Campus, 41500 Larissa, Greece*

ARTICLE INFO

Article history:
Received: 16 March, 2021
Accepted: 13 July, 2021
Online: 27 July, 2021

Keywords:
Digital games
Game-based learning
Mobile robotics
Robotic manipulators
Autonomous robotics

ABSTRACT

This paper addresses experiments with Scratch-developed games in the robotics introduction course at the Federal University of Technology – Paraná. It aims at assisting learning of classical and initial robotics concepts. This proposal, similar to the classic 80s war tanks game on Atari 2600, was developed using an autonomous vehicle. In the first experiment, applied to the class 2019/2, the students (players) had to battle against another autonomous tank developed (in two different ways, in Person and Distance Learning), using keyboard inputs to control their tank. In this game, the students were asked to create states machine models while were being introduced to fundamental concepts such as pose, other basic notions concerning controlled and autonomous robots, and the hierarchy of actions. At the end of the games, a questionnaire answered by the students extracted valuable findings of the examined concepts. In 2021/1 class, the second and third experiments were applied. The former was an extension of the first experiment, using autonomous parking cars. The latter was inspired by the classic Pong game, with the addition of more degrees of freedom (DOF). In this case, the player attempts to reach and catch a ball through the operation of a robotic arm with two rotating joints, using keyboard inputs. Each block or scenario will become more complex, and the student has time to perform a task. In the case of the third experiment, the concepts including 2-D workspace, multiple solutions, inverse, and direct kinematics were explored. Delivery rates for the first and second experiments were 90% and 80%, respectively. Even though three individual experiments were investigated, the single objective was achieved: the implementation of modern didactic tools to deliver critical pedagogical concepts to students in the robotics class.

1. Introduction

The first 20 years of 2000 have been highlighted by the development and creation of computer technologies. This study is devoted to the integration of certain aspects of these technologies to either the working or leisure areas of everyday human life. In the field of leisure, games have been transformed into digital, triggering increased attention to users of all ages around the world,

[*]Corresponding Author Márcio Mendonça, Email: mmendoncautfpr@gmail.com

https://dx.doi.org/10.25046/aj060425

while the extended popularization of the web helped in this direction. Electronic games are characterized by an enormous variety of levels in terms of their type and difficulty, targeting a wider age group, no longer being a niche market [1], as it used to be in the past.

This paper is an extended version of the previous work published in ILSA 2020 [2] approaching game-based learning. The proposal of active methodologies such as game-based learning aims to involve students by establishing a process that implies action-reflection-action and not just internalization of what has been exposed (as is the case in in-person classes) [3].

With Game-based learning, students learn while playing. Thus, making the learning process more enjoyable, causing a positive effect on cognitive development. Games are combined with traditional classes because the conventional learning process can be monotonous, and game-based learning can improve students' motivation to learn. It is not just about using games to review and reinforce concepts [4].

The games include many problem-solving features, adding elements of competition and opportunity. That is, the student player needs to deal with an unknown result, choose between several paths to an objective, construct a context of the problem, and collaborate with several players [4].

Among the benefits of games implementation in learning, according to [5], the following attributes should be mentioned:

- Games can easily attract the attention of individuals across various demographic boundaries (for example, age, sex, ethnicity and educational status).
- Games can assist young people in setting their goals as they can provide feedback and reinforcement or record changes in human behaviour.
- Games offer fun and excitement to the players. Hence, it is not difficult to attract and maintain a person's attention.
- Games also offer the chance to participants to explore their curiosity and new challenges, thus stimulating their motivation for learning.

On that basis, gamification is the practice of using elements of game design, mechanics and thinking in non-game activities to motivate participants. Gamification in the field of education exploits, among others, sets of games rules, players' experiences, and cultural roles in shaping students' behaviour [4]. Thus, it uses points, badges, ratings, and incentives to engage students in the learning process.

The benefits of gamification in education are [4]:

- Improved learning experience.
- Enhanced learning environment.
- Instant Feedback.
- Promotion of behavioral changes.
- Feasibility to integrate into different learning needs.

Nowadays, the utilization of novel learning technologies has overcome the conventional problems of distance, time, and cost in learning. However, the lack of student motivation is a problem that e-learning still faces. The application of gamification in e-learning is being used aiming at giving participation and increasing student motivation. Within the same learning content, the characteristics of different users and static gamification elements do not increase the expected motivation. To overcome this problem, gamification must be adapted to the characteristics and needs of the learners [6].

The development of computers, tablets, and smartphones with enhanced capabilities derived from improved processing power and low cost has boosted the popularity and wide use of digital games among young people. On this basis, researchers have recently focused on exploring the intrusion of games into the teaching and learning process. In this context, the authors, through the examined scenario, investigate how technological means could improve the teaching-learning process using digital games.

This approach is based on traditional teaching and learning methods, whereas it exploits diverse aspects of the learning process [7]. Additionally, it utilizes question games, where students are rewarded with marks when delivering the correct answers, thus creating active cooperation and healthy competition, which stimulates the learning process; this process is also known as gamification. This gamification scenario has been appropriately applied to exhibit the contribution of games in the learning processes in students' interests [8].

The successful integration of games in education to complement traditional learning is highly attributed to the valuable features of games, including the playful aspects, interactivity, feedback, problem-solving, experimentation, competition, and students' engagement in learning [7, 9]. To instantiate the methodology, the work [10, 11] presents, in addition to learning the proposed concepts, that the games favour students' cognitive and social development through the solution of problems and cooperation between them. An example that can be cited in the literature is the work with children playing with robots. Some elements are relevant in game development and were used in this article, shown as follows.

As a motivation for this research at university level, we can mention the work that uses game-based learning in conjunction with other methods as a teaching technique for children. In [12], a novel teaching structure assisted by a computer was presented for teaching maths in the 5th grade. Based on an award-winning curriculum program, this approach uses music and body gestures to develop certain connections between mathematical concepts and culturally inspired metaphors. Utilizing Virtual Augmented Reality (VAR) and sensors, the presented approach deals with the successful transformation of the class from traditional to digital.

Another factor that supports this research is the fact that the Brazilian Computer Society (SBC) considers the basic concepts of computing as important as those of mathematics, philosophy, physics, and other sciences for contemporary life. Thus, computer science, robotics and digital games have found new adherents with meaningful pedagogical experiences.

It is not the scope of this research to carry out any statistical analysis, only to present roughly the percentage of students who

suggested having abstracted the classical knowledge of robotics in each of the proposed experiments.

Digital games, since their introduction in 1974, have taken tremendous and rapid steps towards improving players' overall gaming experience and involvement. Actually, game developers intended to keep people engaged in playing their game uninterruptedly, targeting new goals every time, being determined to experience new challenges. Even though today's children have a different view towards the use of video and computer games than games were meant to offer, students still show an optimistic attitude characterized as interested, competitive, cooperative, results oriented. At the same time, they actively seek information, fundamentals (which is the purpose of this research, toward robotics) and even solutions [1].

As for the application of games based on Atari, a relevant consideration suggests a possible motivation of students in the application of games, even if it is worth mentioning that: the use of DeepMind Technologies Limited acquired by Google in 2014, has been breaking all records of Atari games and is capable of challenging any human to a match [13]. This motivated students about the importance of digital games in intelligent computational systems and for learning robotics. That said, the students answered questionnaires and developed state machines.

This article is divided into five sections. Section 1 presents the introduction and a brief review of the literature. Section 2 presents the theoretical aspects of games and learning. Section 3 presents the development of the research developed. The results are presented in Section 4, and, finally, in Section 5, the overall conclusions and discussions are presented.

2. Games and learning: theoretical aspects

This approach aims at improving learning, using the motivating effects of the elements and techniques of digital games. Student engagement is the criterion for integrating gamification into the learning process and therefore serves as an essential measure for its effectiveness. However, gamification should not be restricted in considering only games' main concepts such as points and leaderboards, which significantly reduce its educational value and the desired impact on students, who constitute the main target of this attempt [14].

Among the advantages of game-based learning that motivates the current study is the ability to provide learners with the understanding of concepts in a practical way, taking the student to a higher level of involvement with his learning in a more dynamic way. In this context, the work of [15] presents a model under study which illustrates the fact that an intense engagement in learning has a strong effect on human body as dopamine and serotonin do. Specifically, this neural model was based on the assumptions that dopaminergic activity increases as the expected reward increases and serotonergic activity increases as the expected cost of an action increases.

Following a brief history review regarding learning and technology, the view of Robert McClintock, Frank Moretti and Luyen Chou is devoted to the evolution of technology which goes side by side with the change and transformations in teaching and learning. Originally, education and training were a process of imitation and training - "picking up a stone and playing with the animal." If you are unable to do this the first time, practice several times until you succeed. "No, do it this way." The practice has become a way of playing to make this repetitive skill-based learning bearable and memorable. This type of "demonstration and practice" learning requires good coaches, usually in an individual relationship. This is how people learn to play sports, play musical instruments, and master other physical skills. In the most basic, not even language is necessary, e.g., athletes and musicians are often skillfully trained by people who do not or barely speak the same language [1].

According to [16], since the 1980s, some researchers carried out certain studies concerning the use of games in education and its benefits and approaches. In [17], the authors reported the popularization of gamification started only in 2010. The term gamification refers to the use of game elements, as aesthetics, game thinking and mechanics, in non-game-related contexts to involve people, motivate action, improve learning, and solve problems.

It also involves several concepts, such as rewarding and punishing the players. It has several connotations, such as the case of this research in which games are developed to assist in the learning process, in commercial games, or even in the commercial area. For example, in [6], the authors use a Systematic Literature Review (SLR) to explore adaptive gamification in terms of frameworks and methods proposed, as well as other research components. The first step is to define the research question (RQ) and then to search the literature published in popular scientific journal databases. Twenty-five selected articles were finally reviewed, in which the authors identified three elements that comprise the proposed framework. These are adaptive gamification engine, adaptive component, and gamification display. Additionally, eleven types of methods were implemented in adaptive gamification. Among them, Felder-Silverman Learning Style Model (FSLSM) is the most popular method. As for the components of adaptive gamification, four were mined, namely: player/learner profiles, learning style, behavior, and skill/knowledge.

Specifically, the following examples are related to this research. The authors in [18] conclude that in recent years, interest in the theme has been increased to such extend so that a theoretical game model was developed for educational purposes. In this case, not any particular game element could be used for gamification, except a subtly combination of elements that are used to contextualize learning. The initiative behind the game development was the integration of various types of activities into this game, as well as assisting teachers in working with modernized learning content.

Another study considers games as powerful experiences that exploit motivation and engagement [17]. In particular, the deployment of simplified elements reduces project complexity concerning badges, levels, points, and leaderboards, that fail to involve students and damage any existing interest for the learning process. A thorough consideration regarding game design must be given in gamification, apart from just implementing game components successfully. However, gamification is a broader

concept used in different areas of knowledge, for example, the management area [6].

In [19], it was observed that educators could increase the feedback mechanisms by exploiting specific game design elements by applying continuous feedback, visual cues, frequent question and answer activities as well a progress bar. Students are also prompted to interact with the content, experience real-life case studies, make decisions for specific tasks, and have realistic consequences for making wrong or unsatisfactory choices. In this way, they are getting involved within the game flow, which attracts students' attention. In conclusion, satisfactory learning outcomes come when facts are embedded adequately inside a game-based story rather than in a passive text format.

In the example of Sheldon's work [20], the professor of a higher education institution gamified his electronic games development class. Students' marks were starting from zero and increasing with their results. Moreover, there was an increase in the number of game-based activities against the traditional assessments. These activities had the form of missions to defeat enemies and were assigned to groups of students. The outcome of each mission was assessed and formed the student's final grade. Throughout this learning process, it was observed that the average grades of the students were notably improved at the end.

Gee also noted that students perform better when a level system is used in the game-based learning process. The players are more inspired when they apply what they have learned and get the corresponding feedback after completing each level. Each subsequent level requires, after all, skills acquired at previous levels [21]. The authors in [22] conclude that the game's dynamics are improved when levels or progress bars are involved. In addition, storytelling is another game design aspect that highly contributes to the successful integration of games into the learning process. Some examples of these types of games include SimCity, where players follow the story of building a city from scratch, as well as Monopoly, where the story behind the game is to become rich buying and selling property in every round, trying to avoid the risk of losing everything [16].

After an in-depth review of the available literature, certain characteristics were found mainly in [16,17,19-21] regarding the successful development of games integrated into learning environments. These include Freedom to Fail, Quick Feedback, Progression and Counting stories.

3. Development

There are several approaches that deal with the integration of digital games in learning environments. The simplest paradigm refers to the utilization of simple games used exclusively in the classroom whereas, more complex examples focus on the design of digital game concepts that are put into practice but are not necessarily used as learning simulators in virtual or real educational environments [7].

It is not the scope of this work, but it is possible to use board games, for example, as a learning method. Several research works can be found in the literature which study board games-based learning. More specifically, the authors in [7] propose cooperative learning as a means to design an educational course based on the board game. According to the curriculum, pre-service teachers cooperate in developing and delivering a board game prototype.

Because of the game's simplicity, the tutorial with the game instructions was presented during the beginning of the class in Brazilian and Portuguese (Brazil is the native language) and is presented in the metadata of this work.

In the first stage, a group discussion in an online platform took place where the participants were working on the fulfillment of specific activities according to the curriculum. Next, a questionnaire evaluated the performance of the pre-service teachers in the particular task. Based on the results produced, all participants' groups were able to successfully design games in line with the academic disciplines that are played sufficiently and without difficulties. Additionally, teachers' self-efficacy was improved through the process of cooperative learning.

Another study in this area that contributes to learning [23] presents a novel structure devoted to the general board game (GBG). GBG configures boardgame's standard interfaces, states and AI agents. This structure offers cooperation with various agents in different games and defines the parts of board game coding. GBG is particularly suitable for arbitrary 1, 2, 3 or multiplayer board games.

This work proposes three different games developed individually. The first one is devoted to designing a tanking battle game based on a classic Atari 2600. This game includes the definition of some concepts regarding robotics discipline fundamentals. These are the pose concept that complements the Cartesian coordinates (x, y) and the angle, which denotes the object's orientation. Two more games are also presented: one consisting of a manipulator arm and a ball, while the last one deals with car parking.

Goals: each experiment has its own distinct goals. In tanks, the objective is to destroy the opponent; in the robotic arm, the goal is to reach the ball and, finally, in the valet, a parking space must be found.

Rules: In the case of tank war, the opponent must be destroyed before the player is destroyed; in other games, the objectives must be accomplished in the shortest time possible.

The proposed digital game was built on the Scratch platform, which was developed by the Massachusetts Institute of Technology (MIT) team in 2007. Scratch is among the most accessible programming languages as it uses a graphical interface and does not require programming skills on any other programming language [24]. The scope of this game is to insert blocks, as shown in the example in Figure 1.

The methodology of the learning process was initially the presentation of various games, such as Pong, which was the first game of the era. Then an explanation followed on how the games work, and next, the commands and objectives were exhibited. In the next phase, the students were introduced to relevant areas of robotics and/or virtual agents related to the tank war game and continued with activities concerning the following two experiments presented.

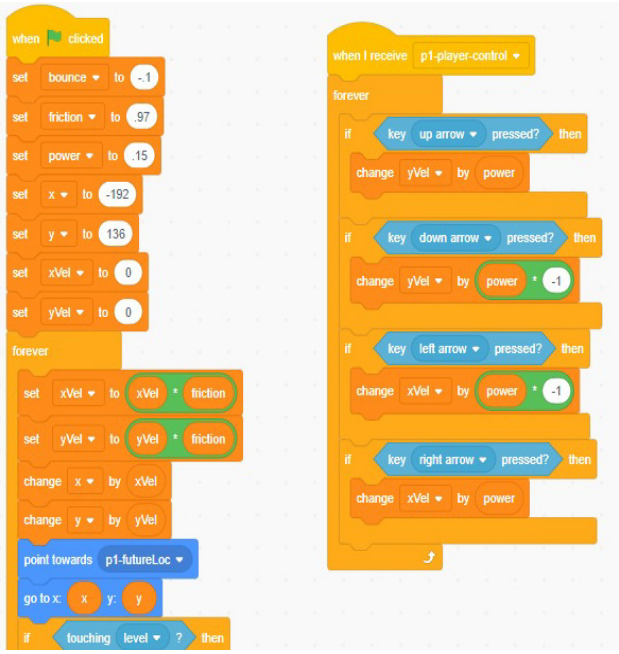


Figure 1: Screen example using Scratch.

4. Results

Three games built on Scratch were presented in different classes. The tanks war (experiment 1) and the autonomous valet (experiment 2) games were presented to the 2019/2 class, and various responses were collected. In sequence, the game with two robotic arms (experiment 3) was illustrated, in which students tried to simply catch a ball (target) by rotating the two arms joints (2 degrees of freedom, GDL) with the help of a keyboard. Concerning the first experiment, 27 students were involved. However, just a small number of results for all three experiments is presented for the purposes of this study due to the limited workspace.

4.1. Experiment 1

The purpose of this exercise was to assist students in understanding certain robotics concepts that concern the need for autonomy and hierarchy when actions are taken. Priority is given to some control actions in the scenario, such as obstacles avoidance and targeting the opposing tank. Additionally, new routines can be applied in the development of future works based on intelligent computer systems. For example, the opponent can become an autonomous entity by implementing intelligent computational methods based on Fuzzy logic. Fuzzy Cognitive Maps can help in this direction in the future, as it provides low computational complexity [25, 26].

In other words, this experiment added the abstraction of a robotic architecture inspired by fundamental concepts of Brooks' subsumption architecture [27]. The notion of priority when a robot, or in this game's case a virtual robot (bot), must prioritize its actions, such as staying alive, avoiding enemy attacks and obstacles before shooting the enemy.

In the initial phase of this game, as shown in Figure 2, both tanks are guided by players. The main objective of the first tank is to destroy the second by shooting at it while trying to avoid other fixed and mobile obstacles (opponent's shots). The next phase of the game design includes determining the game rules and then establishing the possible states of the tanks.

The game rules are defined according to possible damage suffered by the tanks. The game has a winning condition for the first player when he manages three hits to the opponent's tank. One point of damage is attributed either by a shot or physical contact with an obstacle. The distinct states of a tank are the following:

- Moving freely.
- Shooting.
- Avoiding a fixed obstacle.
- Avoiding a moving obstacle (opponent's shot).

According to the exercise proposal in the experiment developed in the seventh period of Control Engineering and Automation, the following tasks need to be accomplished by students:

- Develop tank's finite state machine.
- Identify the tank's position (coordinates) and pose (angle with the x-axis).

In this experiment, we had an abstraction of the mobile robotics concepts of approximately 80% for one class and 90% for another, despite having initial results with applications in only two classes. Objectivized concepts like hierarchy and pose in mobile robotics were abstracted, suggesting that this game is promising for introducing the basic concepts of programmed and mainly autonomous mobile robotics.

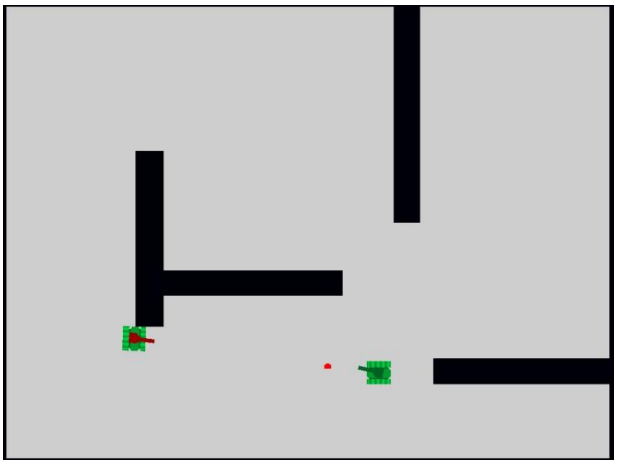


Figure 2: Representation of the first game

Based on the proposal, each student was asked to give a figure of the state machine and a description of the events in a doc or pdf file. In addition to the graphical representation, students were also asked to answer a group of questions to identify whether the nature of the activity made it possible for them to perceive aspects and concepts proposed. Such an activity can be considered as a Subjective Modeling when capturing the impression of the student in the short-term memory as listed below:

a) Is it possible to distinguish between the programmed and the autonomous control? Explain.

b) Is it possible to perceive the robot's hierarchy during the game, and in what way?

c) Is an exemplified angle during the battle necessary, in addition to the x, y coordinates?

d) Is It possible to perceive the different states, attacks and defences?

e) Develop a state machine that models the player's actions to defeat the opponent.

One exemplary and one complete answer provided by the students are listed below for each question:

Question a:

- Student 1: Yes, because the user-controlled tank responds to the given commands, whereas the autonomous only responds to the events following its logical construction.
- Student 2: It is possible. The opponent's tank moves without any human intervention. On the other hand, the player's tank is controlled at each cycle by the keyboard.

Question b:

- Student 1: The hierarchy is noticed as the robot tank is looking for its own defense, and then is trying to destroy the opponent with the intention to win. Our tank has no hierarchy since it is controlled by the user.
- Student 2: Hierarchy is perceived when activating opponents, avoiding obstacles and the opponent's attack for keeping its integrity (game's objective).

Question c:

- Student 1: The angle is used to increase the control precision. For example, position x = 0, y = 0 is where the tank is located in the environment. When X = 1, the tank will not necessarily move forward, but it may be rotated.
- Student 2: The question was not answered.

Question d: only two students answered the questions. However, the state machine abstraction was the main objective of this experiment.

Student 1: Student 1 provided a satisfactory interpretation of the robot's movement finite-state machine, shown in Figure 3. The possible states are: 1 - shoot avoidance; 2 - stopped; 3 - chasing the enemy; 4 - aiming at the enemy tank; 5 - shooting the enemy and 6 - obstacle avoidance.

Only two students were inserted due to a large amount of space for a more accurate analysis. However, students were chosen at random, and their responses can provide feedback on their learning from the most complex experiment, the tank battle. Other students have already managed to develop state machines for game strategy, which are shown below.

The response by the student is presented below.

a) Do not press anything.

b) Press arrow keys.

c) Mouse left-click.

d) Spin the mouse.

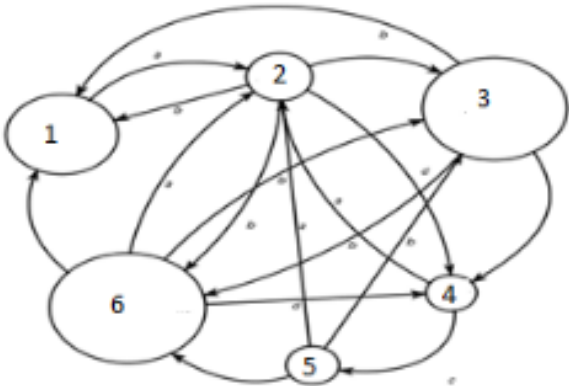


Figure 3: Finite-state machine by Student 1.

On the other hand, it is worth noticing that student 2 (Figure 4) provided a more complete interpretation than those submitted by the rest of the class. It seems to be more accurate and closer to the actual operation of the tanks in the game.

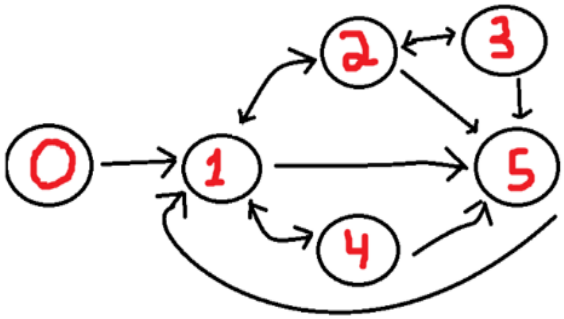


Figure 4: Finite-state machine by Student 2

- State 0: The game is stopped and has not yet started.
- State 1: The game started; tank not moving.
- State 2: The tank is moving.
- State 3: The tank moves, aiming and firing.
- State 4: The tank is stopped, aiming and firing.
- State 5: The tank is dead, awaiting re-entering to the game.
- State 0 → 1: Press the button to start the game.
- State 1 → 2: Press the directional buttons to move the tank.
- State 2 → 1: Release the movement buttons. Tank in motion.
- State 2 → 3: Press the shoot button to target. Tank in motion.
- State 3 → 2: Release the fire button. Tank in motion.
- State 1 → 4: Press the fire button while aiming at the target. Tank is stopped.
- States 1, 2, 3, 4 → 5: The tank got damage.
- State 5 → 1: The tank respawns after an x amount of time has passed.

Regarding student 3, he developed the state machine but left the vocabulary unfinished. Through this process, it emerges that the students have delivered various interpretations. Additionally, the students seemed to comprehend specific concepts in the field of robotics and their autonomous mode as well.

The graphical representation that student 4 provided (see Figure 5) illustrates the finite-state machine of the game combined with the respective vocabulary, offering a satisfactory interpretation of the activity concept. Although different interpretations emerged from students, these interpretations were similar concerning the game strategy. Overall, the number of different ways for learning assessment in each of the experiments needs to be highlighted.

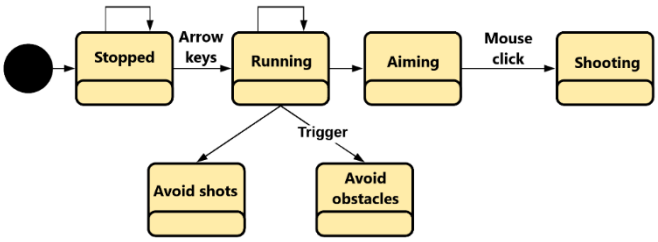


Figure 5: Finite-state machine by Student 4

In this tank warfare experiment, using the state machine was necessary due to different behaviours in combat, such as an attack, defence, avoiding obstacles, enemy shots, and chasing the enemy, among others. However, in the last experiment, four questions were enough to assess students' learning. In general, each experiment has a complexity level, whereas different evaluation methods were used.

State machines have been promising in terms of students' abstraction. Most of the students, about 40%, did not even know the state machines yet (because they are students of an Automation and Control course, whose robotics discipline is offered from the seventh period on). So, the results of this experiment showed promising abstraction of the game's strategy, especially the concepts of robotics involved, mentioned in the text.

New experiments took place in 2021 with distance learning due to the pandemic (University policy to mitigate Covid contagion). The difficulty of experimenting compared to previous ones before the pandemic was much greater [24]. It was necessary to do a live meeting and provide videos of the game's action to help students abstract the strategy and do their required tasks. Figure 6 shows details of the video provided to students.

UTFPR-CP, a federal university (in Portuguese Universidade Tecnológica Federal do Paraná Campus Cornélio Procópio), gave a subsidy to low-income students to purchase computers. In addition, it also promoted aid in cash to low-income students, which was also maintained during the pandemic. Only three examples (see Figures 7, 8, 9) will be presented, which took place in a distance learning class during the pandemic.

The ICT Panel COVID-19 (2020), which took place with young people aged 16 or over, showed that the cell phone was mentioned by 22% of users in social classes A and B, 43% of users in class C and 54% of users in classes D and E. The inequalities of student access to connected devices are striking. The same document also points out: three-quarters of Internet users aged 16 or over in classes D and E (74%) used the network exclusively by cell phone, a percentage that was 11% among users of classes A and B (no reference because it was written in Portuguese).

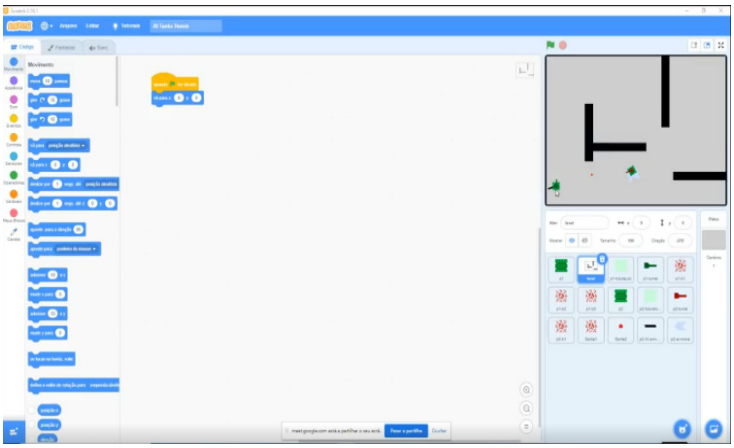


Figure 6: shows details of the video provided to students.

Student 1 (Distance Learning):

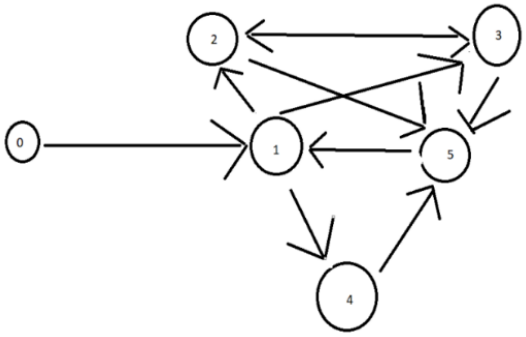


Figure 7: Finite-state machine by Student 1 (distance learning)

Vocabulary state machine.

- State 0 → 1: Start the game.
- State 1 → 2: The player moves the tank without shooting.
- State 1 → 3: The player moves the tank and shoots.
- State 1 → 4: The player stops the tank and shoots.
- State 4 → 5: The tank was stopped and died.
- State 2 → 5: the tank was moving and died.
- State 3 → 5: the tank was moving and shooting and died.
- State 5 → 1: The tank was dead and reappeared.
- State 3 → 2: Tank moving stopped shooting.
- State 2 → 3: Tank moving started shooting.

Student 2 (Distance Learning):

Vocabulary state machine.

- State 0 → 1: Starts the game, with the tank stopped.

- State 1 → 2: Tank is moving.
- State 2 → 3: Tank is moving and firing.
- State 3 → 5: The tank, while firing, dies and awaits reappearance.
- State 5 → 1: The tank was dead and reappeared.
- State 1 → 4: Stopped Tank aims and shoots.
- State 4 → 5: Tank was stopped and died.
- State 5 → 1: The tank was dead and reappeared.

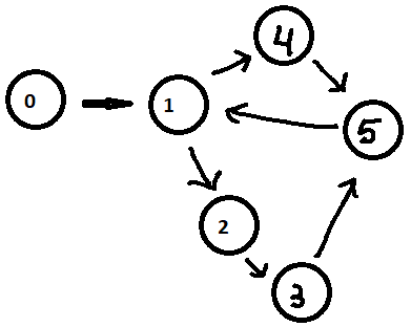


Figure 8: Finite-state machine by Student 2 (distance learning)

Student 3 (Distance Learning):

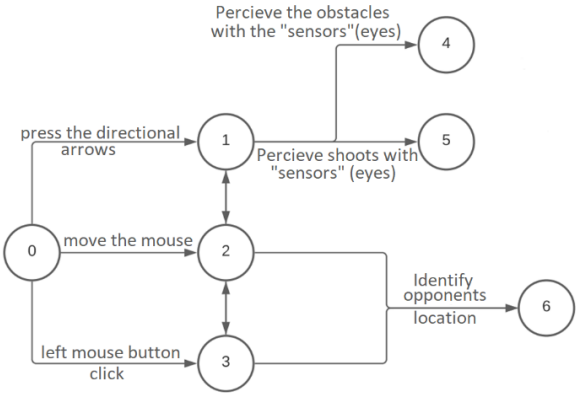


Figure 9: Finite-state machine by Student 3 (distance learning)

Vocabulary state machine.

- State 0: Game start.
- State 1: Player walking.
- State 2: Player aiming.
- State 3: Player shooting.
- State 4: Player dodging obstacles.
- State 5: Player dodging shots.
- State 6: Player hitting opponent.

The first two results were compatible with the experiments carried out in the classroom. The third result presented was one of the best in abstraction, if not the best of all the experiments presented in distance learning and on-site teaching. In addition to the states, the figure shows the events used to control the player's tank.

4.2. Experiment 2

Another simpler game was produced on Scratch, contributing to this work. In this game, the student must drive a car to find a parking space (autonomous valet), as shown in Figures 10 and 11. It was held in the same class (2019/2) following the tank war game.

It was also introduced in the robotics discipline, aiming to present students a brief notion of the maneuver difficulty, even if performed on the keyboard, by an autonomous valet using Fuzzy logic as an example.

The coexistence of several agents or virtual robots in the same scenario was emphasized. It is noteworthy that in the discipline of robotics, students have a posterior example of an autonomous valet. This game reinforces the concepts of tank warfare: autonomy and hierarchy of control actions. That is, it is necessary to avoid the cars reaching a parking space.

One student's answer: despite being a relatively simple activity for human beings, the game allowed us to observe the difficulties of developing an autonomous system, which would make these maneuvers without any human intervention. It helped us to understand a little more about autonomous robotics.

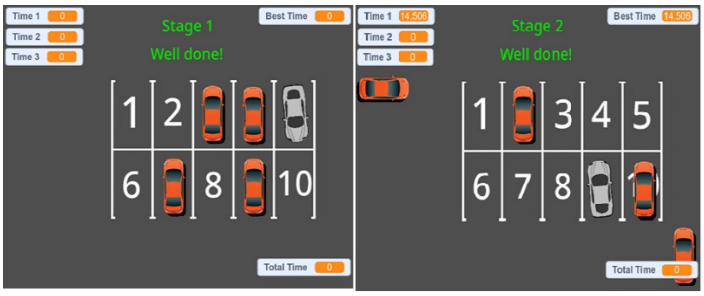


Figure 10: Student 1 answers; initial stages

Compared to Experiment 2, different results were obtained and aided in the autonomous vehicle abstraction. However, two of several examples will be cited in which can be seen that the second student was faster than the first one, as shown in Figures 10 to 13. Nevertheless, practically all students were successful in all three stages. The first objective is to help develop an autonomous valet through Fuzzy logic, for example, fuzzy cognitive maps [28].

The course has already clarified the difference between programmed and autonomous robots, a relevant concept in robotics [29].

Figure 11: Student 1 answers; final stage

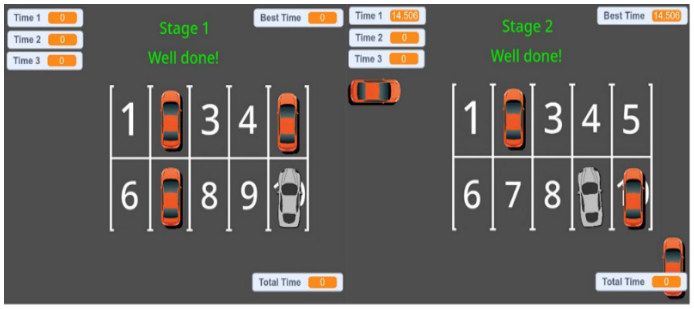


Figure 12: Student 2 answers; initial stages

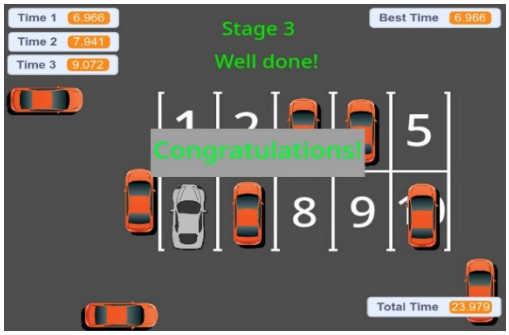


Figure 13: Student 2 answers; final stage

4.3. Experiment 3

The robotic arms game (experiment 3) was applied in the class of 2020/1, which has approximately 23 students. The game works as follows: a ball (target) appears on the scene at random as shown in Figures 14 and 15. The student's goal is to try and catch it using the cursors.

The game objective is to emphasize on direct and inverse cinematics concepts and the action radius (the robot inserts in a 2D environment). Future work could also implement a 3D game addressing a sphere or a search surface. In addition to the objectives, a classic problem regarding robotics manipulation became clear to students: the multiple solutions that can occur.

According to the figures presented, the game strategy is oriented to reaching three targets in sequence within the space area. However, the ball's movement cannot be achieved in certain situations, especially when the robotic arm is out of reach. It should be mentioned that the robotic arm has two degrees of freedom and also have rotational joints.

Due to workspace restrictions in this paper, the presentation of the answers to the provided questions will be limited, and only a few examples will be presented. It also needs to be highlighted that the results illustrated below were randomly selected to maintain the authenticity and truthfulness of the current research. Experiment 3 involves the following questions:

a) Is it possible to use more movements to catch the ball?

b) Would the movement become more natural if the joints move at the same time?

c) Does the game have direct or inverse kinematics?

d) Are there inaccessible positions of the arm?

In the meantime, the results in Figure 15 should be considered as they are the second successful attempt for student 2. The answers of five students are shown below for each of the four questions.

Question a:

- Student 1: Since the arm has two DOF, it is possible to have more than one solution.
- Student 2: It is possible for some ball positions, according to the image of the first line, in the same position, that there is more than one way to capture it. However, in the second line, there is only one possibility.
- Student 3: Yes, the movement is more harmonious when the joints move simultaneously, as this situation results in faster movements to reach the ball, more agile and more natural compared to a real human arm.
- Student 4: Yes, with the combination of the angles and the ball's position.
- Student 5: Yes, more than one solution is possible.

Question b:

- Student 1: Yes, this process would require greater computational power, but it would have a more harmonious movement closer to the human arm movement.
- Student 2: It would be, but it is easier to use keys to move both together simultaneously in the game.
- Student 3: Yes, it is possible to have different solutions to catch the ball, as each player has the freedom to choose the arm's movements by joining different angles for each joint but reaching the same point in the Cartesian plane.
- Student 4: Yes, it mimics the human body more.
- Student 5: Yes.

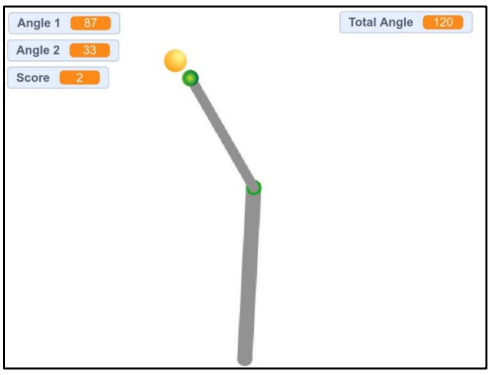


Figure 14: Student 1 result: first attempt

Question c:

- Student 1: Inverse, as the movement orientation of the joints occurs through the ball's position.
- Student 2: Inverse kinematics because the arm has a desired position and orientation (ball).

- Student 3: The game has inverse kinematics, as the player must define the joints' positions after a given ball orientation.
- Student 4: Inverse, the arm needs to follow the position of the given ball.
- Student 5: Direct, because the angles are already given, so the position of the arm can be found. In the game, we are going after the target "ball" visually.

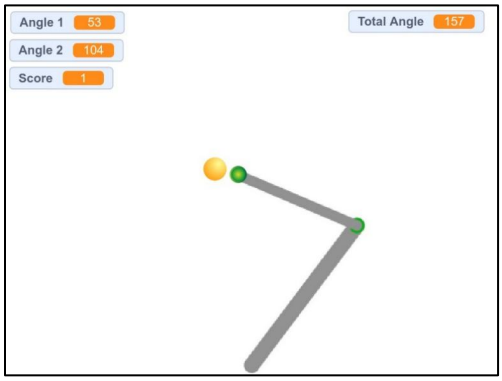


Figure 15: Student 1 result: second attempt

Question d:

- Student 1: Yes.
- Student 2: Yes, according to the figures, the ball travels practically the entire space of the screen and some points that are not unattainable.
- Student 3: The question was not answered.
- Student 4: Yes, so that a key is needed in the game to reset the ball.
- Student 5: Yes, then there is the ball moving.

The given answers emerge that they were similar since students could identify the problem just by playing the game. The correct answer to the fourth question is that the arm cannot reach the ball (target) because of its structure. Moreover, student 4 gave a more consistent answer to the second question, explaining his response assertively. Overall, the game seemed to have its objective fulfilled. In specific, considering a sample corresponding to approximately 20% of the class, it emerges that some answers were more competent than others. However, it is concluded that there was an abstraction of the contents and objective foundations.

Figures 14 and 15 illustrate in a more precise way the structure of the game's components and how the robotic arm changes its position to reach its target (ball), according to the game's random logic. The experiment was carried out in pairs, and the students had to repeat it three times and search for the target. The game attempted to leverage active learning [26] through objective concepts visualization. Additionally, the game demanded multiple solutions from students, as they had to find the target near or similar positions with different angles.

To provide a further understanding of the game, it is necessary to examine the arm's angles. Based on the representation inspired by the geometric model of MathWorks (see Figure 16), the angle θ1 is formed by the intersection of Arm 1 (L1) and the X-axis. The angle θ2 is formed by the continuation of Arm 1 (L1) and Arm 2 (L2), considering that (X, Y) is the desired position.

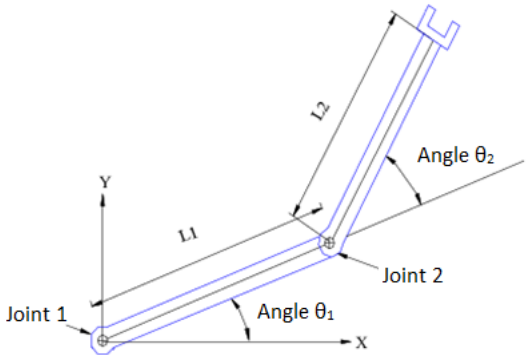


Figure 16: Representation of 2-DOF robotic manipulator.

5. Conclusion

The results, although initial, suggest the feasibility of the Scratch's proposals based on classic games of the 80s. The most important part of the experiments was the fact that the students experienced and became familiar with significant concepts in the domain of robotics through playing games. Among the produced outcomes was students' perception who considered the class "lighter" and liked the experiments. In short, the results were seemed convincing in terms of game-based learning since over 80% of the students had shown interest in these experiments, even if they were relatively simple. Therefore, the application of this methodology is considered quite promising.

In addition, it was possible to carry out some examples of experiments 1 and 3 in the distance learning modality, with a slight reduction in the percentage of students' interest, approximately 70%. Furthermore, experiment 3 had one of the most complex state machines and one of the best (if not the best) of the tank battle game, as mentioned above, which reinforces the promising results of the method.

The participation and motivation of the students in general in both classes was over 80%. According to the few results of the experiments, it can be observed that the objectives of introducing the basic concepts of robotics initially proposed were abstracted by the students with an even higher percentage than the motivational. That is approximately 85%.

Future works can emphasize the control of the opponent's tank and the appearance of obstacles like trees in the tank battle game. For the robotic arm, a 3D arm can be introduced to increase the difficulty of reaching the target significantly. Thus, the solution of the inverse kinematics would be more complex and the number of solutions to the problem. And finally, future work can be oriented to the development of realistic prototypes that would increase the difficulty of the game and introduce and utilize soft computing based on Fuzzy Cognitive Maps (FCMs). Considering that this intelligent technique has low computational complexity, it could be exploited to construct prototypes of low computational and financial cost, such as Arduino. Finally, as mentioned, the next generation of the games may also deploy other programming platforms, for instance, UNIT. Overall, considering the promising

results of distance learning, it would be interesting to conduct new experiments.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors gratefully acknowledge the Federal University of Technology – Paraná, campus Cornélio Procópio.

References

[1] M, Prensky, Digital game-based learning. McGraw-Hill, 2001.
[2] M. Mendonça et al., "Digital Game-Based Learning in a Robotics Course," in 2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA), 1–7, 2020, doi: 10.1109/IISA50023.2020.9284366.
[3] R. Al-azawi, F. Al-faliti, M. Al-blushi, "Educational gamification vs. game based learning: Comparative study," in International Journal of Innovation, Management and Technology, **7**(4), 132–136, 2016, doi: 10.18178/ijimt.2016.7.4.659.
[4] M. Ebner, A. Holzinger, "Successful implementation of user-centered game based learning in higher education: an example from civil engineering," Computers & Education, **49**(3), 873–890, 2007, doi:10.1016/j.compedu.2005.11.026.
[5] F. Rozi, Y. Rosmansyah, B. Dabarsyah, "A Systematic Literature Review on Adaptive Gamification: Components, Methods, and Frameworks," in 2019 International Conference on Electrical Engineering and Informatics (ICEEI), 187–190, 2019, doi:10.1109/ICEEI47359.2019.8988857.
[6] S. Sena, et al., "Aprendizagem Baseada em Jogos Digitais: A Contribuição dos Jogos Epistêmicos na Geração de Novos Conhecimentos," in Revista Novas Tecnologias na Educação (RENOTE), **14**(1), 1–11. 2016.
[7] S. Freitas, "Are games effective learning tools? A review of educational games," Educational Technology & Society. **21**(2), 74–84, 2018.
[8] P. F. C. Santana, D. X Fortes, R. A. Porto, "Jogos digitais: a utilização no processo ensino aprendizagem," Revista Científica da FASETE, 1, 218–229, 2016.
[9] T.-H. Tsai, H.-C. Lin, K.-C. Huang, "Digital Game-Based Learning on Digital Archives: A Case Study of Taiwanese Classical Poems," in 2012 IEEE Fourth International Conference on Digital Game and Intelligent Toy Enhanced Learning. 132–134, 2012, doi:10.1109/DIGITEL.2012.36.
[10] N. P. Zea, J. L. G. Sanchez, F. L. Gutierrez, "Collaborative Learning by Means of Video Games: An Entertainment System in the Learning Processes," in 2009 Ninth IEEE International Conference on Advanced Learning Technologies, 215–217, 2009, doi:10.1109/ICALT.2009.95.
[11] A. Barmpoutis, et al., "Exploration of Kinesthetic Gaming for Enhancing Elementary Math Education Using Culturally Responsive Teaching Methodologies," in 2016 IEEE Virtual Reality Workshop on K-12 Embodied Learning through Virtual & Augmented Reality (KELVAR). 1–4, 2016, doi:10.1109/KELVAR.2016.7563674.
[12] E. Gibney, "DeepMind algorithm beats people at classic video games," Nature, **518**, 465–466, 2015, doi:10.1038/518465a.
[13] F. Ymran, O. Akeem, S. Yi, "Gamification Design in a History E-Learning Context," in 2017 International Conference on Information, Communication and Engineering (ICICE). 270–273, 2017, doi:10.1109/ICICE.2017.8479194.
[14] D. E. Asher, A. Zaldivar, J. L. Krichmar, "Effect of Neuromodulation on Performance in Game Playing: A Modeling Study," in 2010 IEEE 9th International Conference on Development and Learning, 155–160, 2010, doi:10.1109/DEVLRN.2010.5578851.
[15] E. Klopfer, Augmented learning: research and design of mobile educational games, Cambridge, 2008.
[16] S. Deterding, "Gamification: designing for motivation," Interactions, **19**(4), 14–17, 2012, doi:10.1145/2212877.2212883.
[17] E. Sanchez, V. Emin-Martinez, "Towards a Model of Play: an Empirical Study," in C. Busch (Ed.), Proceedings of the 8th European Conference on Games Based Learning, **2**, 503–512, 2014.
[18] K. M. Kapp, The gamification of learning and instruction: Game-based methods and strategies for training and education, Pfeiffer & Company, 2016.
[19] L. Sheldon, The multiplayer classroom: designing coursework as a game, Cengage Learning, 2011.
[20] J. Gee, The ecology of games: connecting youth, games, and learning, MIT Press, 2008.
[21] K. Hogan, M. Pressley, Scaffolding student learning: instructional approaches and issues, Brookline Books, 1997.
[22] W. Konen, "General Board Game Playing for Education and Research in Generic AI Game Learning," in 2019 IEEE Conference on Games (CoG), 1–8, 2019, doi:10.1109/CIG.2019.8848070.
[23] MIT Media Lab Scratch (2019, August 1), Scratch, Available at: scratch.mit.edu
[24] M. Mendonça, H. S. Kondo, L. B. de Souza, R. H. C. Palácios, J. P. L. S. de Almeida, "Semi-Unknown Environments Exploration Inspired by Swarm Robotics using Fuzzy Cognitive Maps," in 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 1–8, 2019, doi:10.1109/FUZZ-IEEE.2019.8858847.
[25] E. I. Papageorgiou, Fuzzy cognitive maps for applied sciences and engineering, Springer, 2014.
[26] R Brooks, "A robust layered control system for a mobile robot," IEEE Journal of Robotics and Automation, **2**(1), 14–23, 1986, doi:10.1109/JRA.1986.1087032.
[27] A. M. El-Sherbini, M. A. Aboul-Dahab, M. M. Fouad and M. F. Abdelkader, "Distance learning during Covid-19: Lessons learned and Case studies from Egypt," in 2021 IEEE Global Engineering Education Conference (EDUCON), 1743–1748, 2021, doi:10.1109/EDUCON46332.2021.9454051.
[28] A. Wilby, E. Lo, "Low-Cost, Open-Source Hovering Autonomous Underwater Vehicle (HAUV) for Marine Robotics Research based on the BlueROV2," in 2020 IEEE/OES Autonomous Underwater Vehicles Symposium (AUV), 1–5, 2020, doi:10.1109/AUV50043.2020.9267913.
[29] C. Persello, L. Bruzzone, "Active and Semisupervised Learning for the Classification of Remote Sensing Images," IEEE Transactions on Geoscience and Remote Sensing, **52**(11), 6937–6956, 2014, doi:10.1109/TGRS.2014.2305805.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 223-232 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

An Efficient Combinatorial Input Output-Based Using Adaptive Firefly Algorithm with Elitism Relations Testing

Abdulkarim Saleh Masoud Ali[*,1], Rozmie Razif Othman[1], Yasmin Mohd Yacob[1], Haitham Saleh Ali Ben Abdelmula [2]

[1]*Faculty of Electronic Engineering Technology, Universiti Malaysia Perlis, Pauh Putra Campus, 02600, Malaysia*

[2]*College of Computer Technology Zawia, Ministry for Technical and Vocational Education,00218, Libya*

ARTICLE INFO

Article history:
Received: 13 June, 2021
Accepted: 12 July, 2021
Online: 27 July, 2021

Keywords:
Combinatorial optimization
Input output relations testing
Adaptive firefly algorithm
Elitist operator

ABSTRACT

Combinatorial software testing is regarded as a crucial part when it comes to the software development life cycle. However, it would be impractical to exhaustive test highly configurable software due to limited time as well as resources. Moreover, a combinatorial testing strategy would be to employ input-output-based relations (IORs) due to its benefits versus other forms of testing as it concentrates on program output as well as interactions amongst certain input value parameters. However, there are few studies focused on IOR strategies. Although the IOR strategy has been demonstrated to minimize test suite size because of its inherent properties, size could be decreased by appropriately choosing the "don't care value" pertaining to the test cases. To achieve a result, this paper demonstrates a unified strategy by considering the new meta-heuristic algorithm known as the adaptive firefly algorithm (AFA) in order to design an IOR strategy. In contrast to the existing work, the adaptive firefly algorithm represents a novel approach to integrate between test cases pertaining to t-way test suite generation by deploying elitism operator in classical firefly algorithm. The optimization algorithm method has been put forward to be adopted along with this strategy. Because of this, AFA is expected to deliver promising results when employing the IOR strategy. As per the experimental results backed by non-parametric statistical analysis, AFA demonstrated to offer competitive performance versus its counterparts. In particular, AFA has been found to achieve and match 68% with regards to the best sizes based on the published benchmark results including 32% new known best sizes. This finding could aid in the area of software testing by reducing the number of test cases pertaining to test execution.

1. Introduction

Software failure is deemed as a dreadful outcome because it could impact the costs associated with software growth. Software testers usually employ dynamic testing to identify if there are any errors in the system under test. Determination of a number of test cases is done to execute a testing activity. Within the path pertaining to the testing activity, a comparison of the real behavior pertaining to the system is done with the expected behavior [1]. All the values pertaining to each of the parameters have to be tested one time at least. It would be inefficient to test the values individually. It also results in an exhaustive parameter-value combination testing [2]. We need more time for exhaustive testing, also we need to run various test cases. As an alternative, combinatorial testing could be used[1-3]. Most of the current methods do not detect errors caused by a set of multi-input parameters like interfaces that involve greater than two parameters [4], [5]. Recently, t-way testing (t denotes an interaction strength), is being widely employed to overcome the issue T-way testing consists of three forms: variable strength, uniform strength, and relations according to input-output. T-way strategies could be employed by either pure computational, algebraic, or optimization algorithm. Much of the study focused on optimization algorithm; however, it covers only strength less than three, variable strength, input/output based interaction, constraints, seeding, and sequence [6]. A major concern of meta-tests by using optimization is including an interaction test suite with the fewest test cases [7]. Even though the majority of current T-strategies are considered essential and very useful, there is no specific strategy or

[*]Corresponding Author: Abdulkarim Saleh Masoud Ali, Email: abdulkarimali84@gmail.com

https://dx.doi.org/10.25046/aj060426

optimization algorithm method that suits the needs of input-output relationships [6], [8], [9]. Several optimization methods have employed nature-inspired strategies like Genetic Algorithm, Simulated Annealing, Particle Swarm Optimization, Firefly Algorithm, and Ant Colony Optimization. These strategies were designed to obtain optimal, yet efficient results. However, the NP-hard problem is deriving the best test cases. Therefore, not one strategy could be considered the best [6], [9]. These strategies have a proven track record when they can generate optimized test suites pertaining to variable and uniform strength interaction. This method has not yet been explored by the IOR. There is a need for additional investigation on this method in order to reinforce the input-output based relations [10]. For this research, the nature-inspired strategy was used to produce a test suite in IOR as it allowed achieving optimum results. It has been known that this strategy yields encouraging results in terms of variable and uniform strength [10], [11]. A meta-heuristic algorithm named Firefly Algorithm (FA) has been employed in numerous applications as well as industries as it can address combinatorial optimization problems [12], [13]. Certain researchers who are involved in the field of software testing select the algorithm without change regarded to be efficient in producing test data pertaining to structural testing. FA yields competitive results versus Simulated Annealing and Genetic Algorithm with regards to an effective level, average coverage, and average convergence Components [14].

The IOR strategy has been introduced in this paper by combining it with the AFA algorithm, in order to achieve the optimum test suite size. A smaller test suite size is produced by IOR itself versus combinatorial testing. Also, the test suite size could be minimized by choosing the best don't care values. AFA has been put forward to be combined with the IOR strategy because it could have chosen a test case with better weightage, from the start of the random firefly.

The research is organized in the following sequence: In the second part, the relationships based on inputs and outputs are elucidated comprehensively. The Firefly algorithm is explained in more detail in the third part. In part four, the recommended approach is outlined, and part five provides the summary.

Schroeder P.J. et al. put forward IOR [15]. To manage multiple outputs and input. It is not the same strategy when we compare it to a unified and variable strength. The main variation is we should consider a software tester as the relationship based on input and output of the values of the interaction parameters. It is also considered that the IOR application covers uniform and variable strength [16].

Generally, the tested system parameters are distinguished from one another. Indirectly, the associated input parameters possess various values. Tester assumption was used to apply variable or uniform strength interaction testing instead of actual interaction [17].

This could result in the exclusion of fundamental interactions as well as having irrelevant test cases. It is regarded that IOR can solve this issue. This kind of interaction takes into account just interaction pertaining to input parameter values that may influence the output. This happens since just some of the input interact amongst each other or there is a variation of strength pertaining to each interaction [18]. Even though a combination of every input is not needed in this strategy and possesses a smaller number of combinatorial test, there is an enhancement in the ability to identify any error while also reducing the redundancy of the test suite [18], [19] There are currently several IOR-supporting t-way approaches that use pure computational e.g. Greedy, Union, and Integrated t-Way test Data Generator (ITTDG), Test Vector Generator, Aura, ReqOrder and Para Order. Each strategy possesses its own weakness and strength. With regards to the two experiments carried out by [9], [10] it could be said that the ultimate optimum test suite size is produced by ITTDG and Density while the least favorable result is yielded by Union. A research study in [10], [15] discovered that IOR can decrease the size of the test suite by almost fifty percent. There is a decrease in size since the testing is executed according to program1 output. There is less interaction required to be run versus full combinatorial testing. The other key reason is due to only one interaction that could encompass many matching values pertaining to input parameters.

For a detailed explanation for the IOR strategy, Program 1 has been employed as a simple. The program1 has been demonstrated in Figure 1. It includes four inputs, A, B, C, and D. Three outputs are yielded by the program: V, W, and X [15]. The output of the function pertaining to V includes inputs A and C, while function output pertaining to W is usually an interaction amongst inputs B, C, and D. The function output pertaining to X covers a combination of input A and D. The general IOR program could be expressed as:

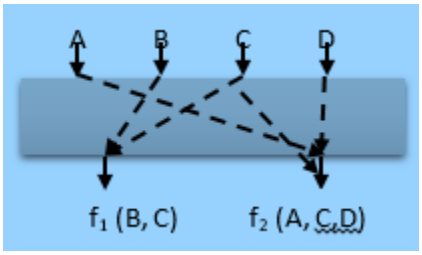


Figure 1: IOR for program 1

CA IOR test suite symbole is expressed in Equation (1), which is taken from [3].

$$TS = IOR\ (N, V_1^{r_1}\ \ V_1^{r_2} \ldots\ldots V_n^{r_n}, R) \quad (1)$$

In the equation,

- TS = test suite,
- MCA = mixed covering array,
- N = the final test suite size,
- t = the interaction strength,
- v = the value of parameters,
- r = number of parameters,
- n = the nth parameter or value.

In this case, if we know the input-output relationships (for example through experimentation), then a generation of input-output-based relations t-way test suite could be done accordingly.

To demonstrate, let us consider (f1 and f2) which are system outputs that conform to the following relationships.

- f (1) is a function of the output of parameters, B and C, that is, f1 =f (B, C).
- f (2) is a function of the output of parameters A, C, and D, that is, f2=f (A, C. D).

Table 1: Input Parameter Values for Program 1

Parameters of Input	A	B	C	D
Values	a1	b1	c1	d1
	a2	b2	c2	d2

Figure 2 (i.e. the shaded input parameter-value) shows the tuples (i.e. parameter-value interaction elements) generated by these two input-output relations. The un-shaded parameters could be regarded as don't care in which any valid value could be employed.

To form a comprehensive input-output-based relation test suite, all tuples need to be tested at least once. One such approach would be to convert all tuples into a test case by allocating any valid value to all who don't care. An input-output-based relation test suite can be defined as a collection of all form test cases (i.e. without duplication of any test case) as presented in Figure 2.

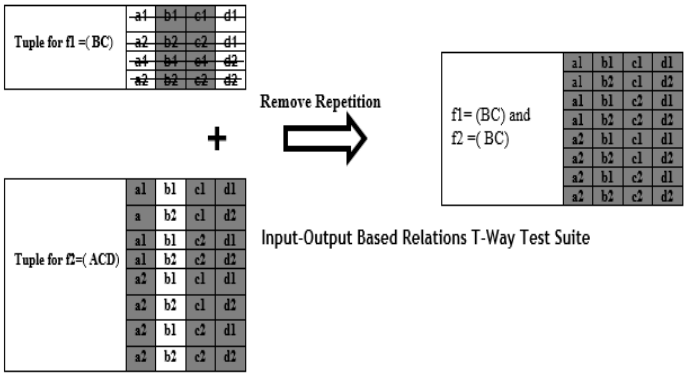


Figure 2: IOR interaction test cases for program 1

Figure 2 shows the IOR interaction derived from the proposal of the College Consultation System program1. The interactions provide only Function output f (1) until f (2). Only function outputs-based interactions are handled while different not linked input interactions or parameters are ignored. Just 8 test cases are needed to be performed out of 16 test cases.

To produce test cases according to on IOR strategy; generation of combinatorial test cases pertaining to the outputs F1 and F2 of the program is done independently. After this, all two sets pertaining to the test cases are compounded, while eliminating redundant test cases. At the end; only thirteen cases are produced for Program 1, as shown in Fig. 3. It denotes a decrease of forty-five percent.

Based on the last test cases generated, it was observed that certain parameters are don't care values. They can be defined as any value that has been randomly selected. If we select the don't care value appropriately; the number of test cases could be decreased. The best test suite size could be produced by carefully choosing the don't care value. In this paper work, they don't care values were wisely chosen by integrating the AFA algorithm into the approach.

2. Mathematical of Fireflies Algorithm (FA)

In [20], the author developed FA which is an optimization-based algorithm, and was based on the Firefly characteristics. There are more than 2000 Firefly species in the world, which show a unique flashing configuration. The Firefly flashing was based on a bioluminescence technique. The Firefly flashes light for attracting a mating partner, warning of the predators, or as a form of communication. As the fireflies are unisexual, any fly gets attracted to light. Furthermore, distance and light intensity are inversely proportional, which indicates that the light absorption decreases with the distance increase between the fireflies. Thus, the real function is optimized for determining the Firefly light intensity. These functions were combined for deriving a novel solution. FA is regulated using 3 parameters, i.e., randomization absorption coefficient, c; attractiveness, b; and parameter, a; as described in Figure. 1 [14–17]. FA has been recently developed as a swarm intelligence technique, which was developed in 2008 by Yang. This was a nature-inspired, stochastic, meta-heuristic algorithm, which could be used for solving some difficult optimization problems (or NP-hard problems). As it is a stochastic algorithm, it uses a type of randomization process which searches for a solution set. This algorithm was based on the flashing lights of the fireflies. A heuristic is defined as `to find' or `to derive solutions by trial and error'. Thus, there was no assurance that an optimal solution could be derived within a reasonable time period. Also, meta-heuristic indicates a 'higher level', wherein the search process applied in the algorithms was based on the trade-off between the local search and randomization. In a FA process, the `lower level' (heuristic) was based on the development of novel solutions in a specific search space and selects the optimal solution for survival. Furthermore, randomization helps the search process avoid all solutions that were trapped in the local optima. This local search also improves the candidate solutions till better improvements were noted. The FA was developed in 2007 by Yang and was a result of the light flashing pattern and attitude of the fireflies. Thus, FA was based on the 3 rules below:

As the fireflies are unisexual, any fly gets attracted to the light, irrespective of its sex.

There is a direct proportional between the attractiveness of a Firefly and the light intensity, which decreases with an increasing distance. Hence, for any flashing fireflies, the one with a lesser light intensity moves in the direction of the brighter Firefly. If there is no brighter Firefly compared to the Firefly, it moves randomly. Firefly flashing brightness is influenced by the nature of the actual function [21], [22].

For an exaggerating problem, the brightness can be directly proportional to the objective function value. As a firefly's attractiveness is directly proportional to the intensity seen light by neighboring fireflies; the variation of attractiveness β can be now be defined with the distance r by [23]

$$\beta_{ij} = \beta_0 e^{-\gamma r y^2} \quad (2)$$

Where; β_0 was the attractiveness at r = 0; whereas the attractiveness, which simulated the movement of a Firefly (i), in the direction of a brighter Firefly (j), is defined as:

$$x_i = x_i + \beta_0 e^{-\gamma r^2}(x_j - x_i) + \alpha(rand - 0.5) \quad (3)$$

The 2nd part is because of attraction, while the 3rd part was a randomization term, wherein αt was a randomization factor, and ϵit was a vector of some arbitrary numbers derived from the Gaussian distribution or the regular distribution at the time, t. If β_0= 0, it is considered a simple random move. Otherwise, if γ = 0, it decreases to a different of particle swarm optimization [24]. The secondary phase of the proposed technique includes to the Adaption with Elitism (AE) as can be seen in Figure 3, which is also employed as a development algorithm[8]. The preferred result obtained by using the FA to assessed the population is accepted to the AE algorithm to give a neighbor solution. If the new solution is better than the existing one, it is approved. If the new solution is not as good as the previous one, the possibility rule is proved, and the new solution is approved if it showed the possibility rule as in Equation 3. Then, the superior solution is gone the FA to give a new population according to the best candidate. During this procedure, the search for the best solution is continuous.

For minimizing the test cases and improving the results, the researchers implemented the FA with the test suite generator.

A random list of some test cases is generated, called the fireflies. Thereafter, all test cases are evaluated for deriving the weightage. Another condition is present wherein the generated test case is added according to all combination pairs coverage in the combination list. It was seen that if a test case covered one combination pair from every combination list, it was believed to include a maximum coverage (i.e., weightage). Thereafter, it is added to the final test suite. On the other hand, if it did not include the maximum coverage, it is added to the memory of the fireflies, wherein their memory (population) is filled with the candidate fireflies (i.e., test cases). These test cases undergo an improvisation process, for deriving a better value of the intensity, which indicates the test case weightage. It was seen that if these improvised selected test cases showed a better weightage value, the primary test case is replaced by the improvised test case. Furthermore, when this process reached a maximal generation, the test case showing the maximal weightage amongst the fireflies was added to the test suite.

3. The Proposal Approach

For archiving, the performance Input-Output Based on Relations using adaptive firefly algorithm this paper is proposed adaptive firefly to minimum test cast.

3.1. Implementation of Test Case Generation

For minimizing the test cases and improving the results, the researchers implemented the FA with the test suite generator.

A random list of some test cases is generated, called the fireflies. Thereafter, all test cases are evaluated for deriving the weightage. Another condition is present wherein the generated test case is added according to all combination pairs coverage in the combination list. It was seen that if a test case covered one combination pair from every combination list, it was believed to include a maximum coverage (i.e., weightage). Thereafter, it is added to the final test suite. On the other hand, if it did not include the maximum coverage, it is added to the memory of the fireflies, wherein their memory (population) is filled with the candidate fireflies (i.e., test cases). These test cases undergo an improvisation process, for deriving a better value of the intensity, which indicates the test case weightage. It was seen that if these improvised selected test cases showed a better weightage value, the primary test case is replaced by the improvised test case. Furthermore, when this process reached a maximal generation, the test case showing the maximal weightage amongst the fireflies was added to the test suite.

The test of the T-way interaction deals with optimal test suite size as well as takes a lesser time to produce the test suite. Investigators have made efforts to deal with the problem by concentrating on either of these [19]. This study concentrates on the IOR strategy to get an ideal test suite size. As stated previously, although IOR could yield lesser sized test cases, this study involves the most optimal test suite size. A lesser test suite size signifies minimal time and costs to apply the testing activity.

3.2. Adapt Firefly Algorithm Process

In this research study, we have used the AFA algorithm due to its ability to choose the best test case with better weightage from the start of the tour. However, there are certain changes required to outfit this subject. In such a case, the input/output based relations become the subject matter, while the algorithm is adjusted to variable strength interaction. This is due to the fact that IOR includes a range of strengths. The algorithm can be employed even if the strength is alike for each function output. The key priority here is that IOR needs to support the limitations. The interaction amongst the values pertaining to the input parameters also has an impact on the program output. Thus, the put forward strategy overlooks the unrelated values.

For the IOR strategy, Section II needs to be revisited. The strategy only stresses on the interactions amongst input that generate the function output. All of the interaction input, as well as its function output, is categorized under SUT's requirements and not merely as the assumption of the tester. Processing of the produced function outputs is done to acquire interaction test cases by accounting for the IOR strategy.

A flow chart that represents how the AFA technique is employed in the scheme, is displayed in Figure 3. The IOR interaction test cases list is required to be produced. This list includes the test cases which are uncovered. It functions as the input for the AFA module. The AFA module begins with the initialization of the pheromone trail.

A constant value for the pheromone and a heuristic value is set. After that, every ant begins formulating solutions by exploiting or exploring the edges of every factor. The edges represent every parameter's value. The ant revises its local pheromone corresponding to the selected edge. This procedure goes on until all the fireflies finish the shifting of the parameters. Here, a fitness function is employed to search for the most optimal test. Only the most optimal test case which has a better weightage

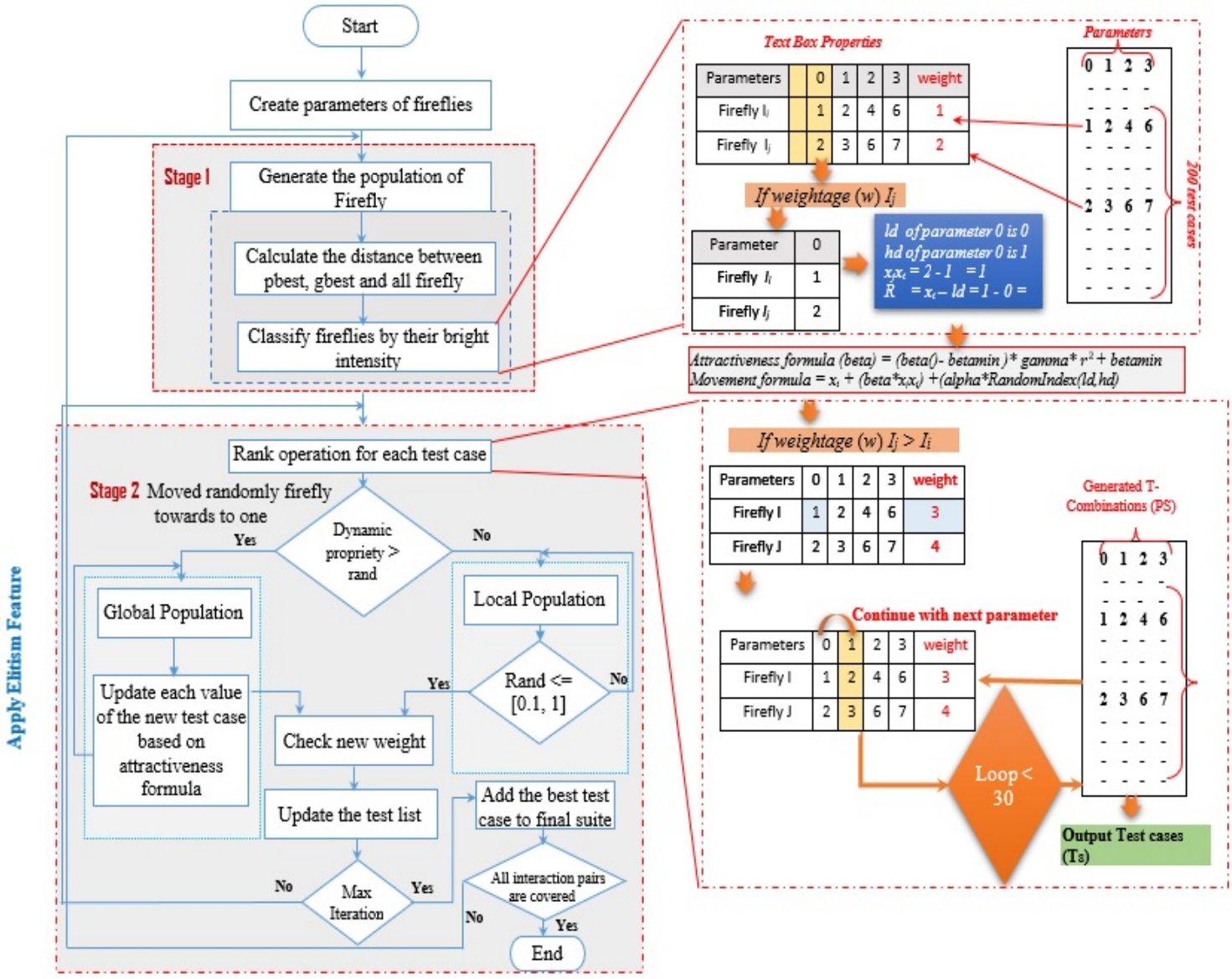


Figure 3: Input-Output Based on Relations using adaptive firefly algorithm

Some concerns to be regarded in the implementation of the IOR scheme combined with AFA are Combinatorial interaction – that the interaction of the input factor values of SUT is set. The one who is testing must take into account each of the function outputs as well as their corresponding input parameter value interaction.

Heuristic value – this heuristic value is needed to help in searching for a good quality solution in the metaheuristic technique and can become accustomed to any situation of the optimization.

In the AFA technique, a constant value of the pheromone and the heuristic value is computed during Pheromone Trail Initialization. In this recommended scheme, a heuristic is crucial to ensure that the ants start their exploitation or exploration with consistent information instead of random preliminary values which can possibly end up causing a bad edge selection. The value may perhaps help in decreasing the duration of time required in producing the test cases.

The heuristic value represents the number of interactions from one node (i) to the next node (j) over the interactions summation involved by node i. The formula for the heuristic value given in [11] can be applied with an adjustment to be appropriate for the IOR scheme. The formula is:

$$\eta_{i,j}(t) = \frac{IOR_{i,j}(t)+1}{max_{e_{i,j\epsilon}G\{IOR_{i,j}(t)+1\}}} \quad (4)$$

IORi.j (t) is a summation of IOR interactions between parameters i and j. G is the set of parameters and its edges.

Fitness function – it represents the amount of interactions which are covered by the current and not by the previous test.

In the case of the IOR scheme, the highest fitness function is the amount of the programmer outputs of the SUT. The formula that follows is employed to compute the amount of the fitness functions, f(t):

$$f(t_i) = \sum_{p=0}^{programoutput} W_p \quad (5)$$

Where (W_P) represents the IOR interactions that are remain uncovered by the previous test but gets covered by the test t_i.

AFA summary is represented by the shaded box in Figure 3: The technique adds the most optimum test case into the last test set, and after that, the covered interaction components are discarded from the list of the interactions. Further, the interaction components are examined. Once each of the interaction components is covered (that is, the list of the interactions is empty), the end of the iteration; otherwise, there is a repetition of the search procedure. The firefly procedure for the t-way examination is explained in more detail as follows:

1. Every test case is one firefly; every test case contains a list of the uncovered tuples (light intensity).
2. Compute the weight in every test case (list of uncovered tuples) using the value of the interaction components.
3. Apply the Firefly scheme randomly at the in-depth test set, beginning with an arbitrary position to compute the fitness function (value). If the existing firefly value is higher, then the preceding firefly does not move and keeps its current position (value), and moves with the other fireflies.
4. Generate the population of firefly: A random list of some test cases is generated, called the fireflies.
5. Move firefly to brighter one: These test cases undergo an improvisation process, for deriving a better value of the intensity, which indicates the test case weightage. It was seen that if these improvised selected test cases showed a better weightage value, the primary test case is replaced by the improvised test case.
6. Calculate the attractiveness and distance for each firefly: if a test case covered one combination pair from every combination list, it was believed to include a maximum coverage (i.e., weightage).
7. To improvisation-based dynamic elitism.
8. Adaptive elitism with dynamic property selection based on the local population and global population to obtain the best values for every test cases.
9. If the dynamic property is greater than the random it will get the global population and check for the length of FA by iteration for the new Wight (gbest). It will put the best test cases inside the memory else it will get the local population.
10. Carry out the operation of elitism for enhancing the local and universal population by using the steps given below:
11. If the random is less than or equal to the probability [0,1] to find several elite flies.
12. Then it will do iteration for maximum to check the new weight. As it gets the best weight for (test cases).
13. Added the memory for the final test suit until stop the iteration. In the proposed self-adaptive Firefly with elitism test list generation strategy, the poor solutions will be replaced by the new solutions based on local population or global population dynamically.

The integration of the FA technique with an Elitism technique is done to achieve a balance between exploitation and exploration to make sure that there is an effective convergence as well as an ideal solution.

It is used as a development technique as well because of the best outcome by using FA after the optimum solution is calculated. They adopted the Elitism to create a nearby solution, and in case the new solution is improved in comparison to the existing solution, then it is agreed.

4. Benchmarking and Discussion

Two experiments published in [10],[25] are conducted to evaluate the performance of AFA in IOR. The capability of the size of the T-way IOR generation optimal test set inspires the experts to investigate further in this field. Several schemes that are in support of IOR are proposed in the literature, as cited in the "related work" part of section 2. In this part, we standardize the results with nine other schemes, i.e. Union [7], Greedy [10], the Generator of Test Vector, ITTDG (Integrated T-Way Test Data Generator) [13], Para Order and, ReqOrder [18], AURA [26], Density [14], TVG (Test Vector Generator) [27], CTJ (Jaya algorithm) [25].

Nonetheless, these approaches use the computational search method, and the metaheuristic search method was not used. This study implements AFA on the sizes of the IOR test set. Moreover, two experiments published in [10] are conducted to evaluate the performance of AFA in IOR.

Furthermore, the input-output relationship (R) in the tests begins with the first 10 requested interactions, and then the subsequent 10 interactions are added every time until there are 100 interactions.

where R= [{1, 2, 7, 8}, {0, 1, 2, 9}, {4, 5, 7, 8}, {0, 1, 3, 9}, {0, 3, 8}, {6, 7, 8}, {4, 9}, {1, 3, 4}, {0, 2, 6, 7}, {4, 6}, {2, 3, 4, 8}, {2, 3, 5}, {5, 6}, {0, 6, 8}, {8, 9}, {0, 5}, {1, 3, 5, 9}, {1, 6, 7, 9}, {0, 4}, {0, 2, 3}, {1, 3, 6, 9}, {2, 4, 7, 8}, {0, 2, 6, 9}, {0, 1, 7, 8}, {0, 3, 7, 9}, {3, 4, 7, 8}, {1, 5, 7, 9}, {1, 3, 6, 8}, {1, 2, 5}, {3, 4, 5, 7}, {0, 2, 7, 9}, {1, 2, 3}, {1, 2, 6}, {2, 5, 9}, {3, 6, 7}, {1, 2, 4, 7}, {2, 5,8}, {0, 1, 6, 7}, {3, 5, 8}, {0, 1, 2, 8}, {2, 3, 9}, {1, 5, 8}, {1, 3, 5, 7}, {0, 1, 2, 7}, {2, 4, 5, 7}, {1, 4, 5}, {0, 1, 7, 9}, {0, 1, 3, 6}, {1, 4, 8}, {3, 5, 7, 9}, {0, 6, 7, 9}, {2, 6, 7, 9}, {2, 6, 8}, {2, 3, 6}, {1, 3, 7, 9}, {2, 3, 7}, {0, 2, 7, 8}, {0,1, 6, 9}, {1, 3, 7, 8}, {0, 1, 3, 7}, {1,4}, {0,9,3}, {3,7,9}, {0,6,8,4}, {3,5}, {1,2,8,9}, {0,6}, {0,3,7}, {2,4}, {7,8,9}, {3,7,6}, {3,8,9}, {2,5,6,9}, {4,7,9}, {5,8}, {4,6,7,9}), {6, 9}, {6,7}, {3,4,7}, {4,8}, {0,9}, {0,2,6}, {1,4,8,9}, {7,8}, {5,8,9}, {3,6,7,9}, {4,8,9}, {2,4,6,9}, {4,8,9}, {3,5,9}, {0,4,9}, {0,6,8,9}, {4,5,8}, {2,5}, {3,5,6,8}, {2,4,7}, {4,5,6,7}, {5,7,9}, {3,5,8,9}, {2,9})].

The results of test size are listed in tables 2 and 3. The proposed AFA Results are compared with the results of previous IOR based strategies. It is worth to mention that the previous studies are implemented with R values from 10 to 60 except the ITTG which is executed up to 100. In this work, the AFA is implemented from R = 10 to 100. The shaded cells refer to the minimum values of the test suite size of the different strategies and configurations.

Comparison between the test suite size results of the proposed

AFA and previous strategies at different IOR (N, 3^{10} , R).

As it is shown in Table 2, the AFA performed better than the other methods in obtaining the optimal solutions. The AFA obtained optimal solution in the nine strategies instance: Density, Tvg, Reqoreder Union Greedy, ITTDG, Aura, CTJ in addition to obtaining over 95% accuracy in the remaining instance.

Comparison between the test suite size results of the proposed AFA and previous strategies at different IOR (N, 2^3 3^3 4^3 5^1, R).

Table 2: Comparison of Size Generated by Different Strategies for IOR (N, 3^{10}, Rel)

Computational									Metaheuristic
R	Density	TVG	Req Order	Para Order	Union	Greedy	ITTDG	AURA	CTJ	AFA
10	86	86	153	105	503	104	81	89	88	83
20	95	105	148	103	858	110	94	99	100	**90**
30	116	125	151	117	1599	122	114	132	118	**106**
40	126	135	160	120	2057	134	122	139	128	**114**
50	135	139	169	148	2635	138	131	147	134	**122**
60	144	150	176	142	3257	143	141	158	145	**131**
70	NA	NA	NA	NA	NA	NA	139	NA	NA	**133**
80	NA	NA	NA	NA	NA	NA	140	NA	NA	**137**
90	NA	NA	NA	NA	NA	NA	110	NA	NA	**83**
100	NA	NA	NA	NA	NA	NA	101	NA	NA	**90**

Table 3: Comparison of Size Generated by Different Strategies for IOR *(N, $2^3 3^3 4^3 5^1$, Rel).*

Computational									Metaheuristic
R	**Density**	**TVG**	**Req Order**	**Para Order**	**Union**	**Greedy**	**ITTDG**	**AURA**	**CTJ**	**AFA**
10	144	144	154	144	505	**137**	144	144	144	144
20	160	161	187	161	929	**158**	**160**	182	165	**160**
30	165	179	207	179	1861	181	169	200	170	**160**
40	165	181	203	183	2244	183	173	207	173	**161**
50	182	194	251	200	2820	198	183	222	191	**180**
60	197	209	250	204	3587	207	199	230	209	**187**
70	NA	NA	NA	NA	NA	NA	190	NA	NA	**187**
80	NA	NA	NA	NA	NA	NA	249	NA	NA	**242**
90	NA	NA	NA	NA	NA	NA	268	NA	NA	**264**
100	NA	NA	NA	NA	NA	NA	260	NA	NA	**254**

The proposed AFA method gives the most optimal size of the test set among all schemes with the exception of R=10, for which the optimum size of the test set is given by ITTDG. In this case, nonetheless, AFA still provides the second-best optimal size of the test set.

Table-2 displays that the worst working is of the Union scheme for all the cases.

In view of the outcomes given in Table 3, the recommended AFA scheme gives the best size of the test set among all the schemes excepting of R=10, where, out of 9, 6 schemes produce an equivalent optimum size of the test set. Apart from AFA, the other schemes are TVG, AURA, ParaOrder, ITTDG and Density. Moreover, the optimum size of the test set is generated by ITTDG, and AFA at R=20. The Union scheme, among all the schemes, has the worst working in all the cases, same as that given in Table 2.

5. T-way Result Statically Analysis

The statistical analysis is performed using the Friedman [28], and Wilcoxon [29] signed-rank test with Bonferroni-holm correction ($\propto_{holm}$) at 95% confident level (i,e, $\propto = 0.005$). In this section, the statistical analysis is divided into two subsections. The First sub sections consider the result of the t- way strength benchmarking while the second subsections consider the results of the mixed-strength benchmarking. The strategies with N/A and N/S results are considered incomplete and ignore samples as there is no available result for the specified test configuration.

The statistics for Friedman test and Post-hoc Wilcoxon signed-rank test is used between AFA and each strategy and it is presented in Tables 4-5, through 2 and tables 6-7 through 3 with the confidence of 95% level (i.e. α =0.05). As the tables show the Post-hoc Wilcoxon Rank-Sum Tests give negative ranks (i.e. a number of cases that AFA unable to outperform another strategy), and positive ranks (i.e. number of cases that AFA is better than another strategy), along with ties. The column is labeled Asymp. Sig. (2-tailed) shows p-value probability: if p-value less than 0.005, as recommended in [30]. For the statistical significance, all the AFA (Size) results are based on 10 executions. The test is performed using an SPSS software tool.

Table 4: Friedman Test for Table 2

Friedman	Conclusion
Degree of freedom = 9, $\propto = 0.05$ Friedman statistic (p-vale) = 0.000 Chi-square vale (x^2) = 45.255	0.00 <0.05 (i.e p-value <$\propto$). Therefore, reject H_0 and proceed to the post-hoc test.

Table 5: Wilcoxon Ranks Tets of Table 2

Categories	Pair comparison	Ranks			Asymp. Sig.(2-tailed)	Conclusion
		Negative Ranks	Positive Ranks	Total		
Meta-heuristic-based strategies	**Density -AFA**	**0**	**6**	**6**	**0.027**	Reject the mull hypothesis H_0
	TVG - AFA	**0**	**6**	**6**	**0.027**	Reject the mull hypothesis H_0
	Req Order - AFA	**0**	**6**	**6**	**0.027**	Reject the mull hypothesis H_0
	Para Order - AFA	**0**	**6**	**6**	**0.027**	Reject the mull hypothesis H_0
	Union - AFA	**0**	**6**	**6**	**0.028**	Reject the mull hypothesis H_0
	Greedy - AFA	**0**	**6**	**6**	**0.027**	Reject the mull hypothesis H_0
	ITTDG - AFA	**1**	**5**	**6**	**0.046**	Reject the mull hypothesis H_0
	AURA- AFA	**0**	**6**	**6**	**0.027**	Reject the mull hypothesis H_0
	CTJ - AFA	**0**	**5**	**6**	**0.042**	Reject the mull hypothesis H_0

(courtesy: IBM SPSS version 26)

Table 6: Friedman Test for Table 3

Friedman	Conclusion
Degree of freedom = 9 , $\propto = 0.05$ Friedman statistic (p-vale) = 0.000 Chi-square vale (x^2) = 50.074	0.000 <0.05 (i.e p-value <$\propto$). Thus, reject H_0 and proceed to the post-hoc test.

Table 7: Wilcoxon Signed-Rank (post-hoc) Tets for Table3

Categories	Pair comparison	Ranks				Asymp. Sig.(2-tailed)	Conclusion
		Negative Ranks	Positive Ranks	Total	Mean Rank		
Meta-heuristic-based strategies	**Density -AFA**	**0**	**5**	**7**	**3.00**	**0.043**	Reject the mull hypothesis H_0
	TVG - AFA	**0**	**6**	**7**	**3.50**	**0.028**	Reject the mull hypothesis H_0
	ReqOrder - AFA	**0**	**7**	**7**	**4.00**	**0.018**	Reject the mull hypothesis H_0
	ParaOrder - AFA	**0**	**6**	**7**	**3.50**	**0.027**	Reject the mull hypothesis H_0
	Union - AFA	**0**	**7**	**7**	**4.00**	**0.018**	Reject the mull hypothesis H_0
	Greedy - AFA	**2**	**5**	**7**	**5.00**	**0.063**	Reject the mull hypothesis H_0
	ITTDG - AFA	**1**	**4**	**7**	**3.50**	**0.078**	Reject the mull hypothesis H_0
	AURA - AFA	**0**	**6**	**7**	**3.50**	**0.028**	Reject the mull hypothesis H_0
	CTJ - AFA	**0**	**6**	**7**	**3.50**	**0.027**	Reject the mull hypothesis H_0

6. Conclusion

The Input-Output Relation strategy is an excellent strategy for interaction testing. It is due to the fact that it is capable of dealing with the original input as well as programmer output causing the reduction in the size of the test set because of ignoring any interaction with the separate parameters of the input value. Thus, the recommended AFA is used to optimize IOR strategy to obtain an optimum size of the test set. Few elements are required to be examined so that they can match with the IOR and AFA schemes. The components are heuristic value, the number of a firefly, combinatorial interactions, and fitness functions. The strategy is assessed and compared with other strategies for optimization. Moreover, the statistical examination shows 49% statistical importance according to the compression of the pier of Wilcoxon signed-rank (as shown in Tables 5-7). Thus, this research summary that AFA is a helpful strategy for producing a t-way test set. The results demonstrated that the AFA scheme is an improvement over the traditional algorithm (FA) and other similar algorithms due to the enhancement of the diversity of its population by the elitism operatives. In the view of AFA's promising performance, it is proposed that other limitations of FA, like the weak examinations in high-dimensional.

Acknowledgment

The author would like to acknowledge the support from the Fundamental Research Grant Scheme (FRGS) under a grant number of FRGS/1/2018/ICT01/UNIMAP/02/1 from the Ministry of Education Malaysia.

References

[1] R. N. Kacker, D. R. Kuhn, Y. Lei, and J. F. Lawrence, "Combinatorial testing for software: An adaptation of design of experiments," Measurement, **46**(9) 3745–3752, 2013, doi.org/10.1016/j.measurement.2013.02.021.

[2] R. C. Bryce, Y. Lei, D. R. Kuhn, and R. Kacker, "Combinatorial testing," in Handbook of Research on Software Engineering and Productivity Technologies: Implications of Globalization, IGI Global, 2010, 196–208, doi: 10.4018/978-1-60566-731-7.ch014.

[3] R. R. Othman and K. Z. Zamli, "T-Way Strategies and Its Applications for Combinatorial Testing," Int. J. New Comput. Architecture their Applications, **1**(2) 459–473, 2011, doi: 10.1109/PRDC.2007.55.

[4] R. Kuhn, Y. Lei, and R. Kacker, "Practical combinatorial testing: Beyond pairwise," It Prof., **10**(3) 19–23, 2008, doi**:** 10.1109/MITP.2008.54.

[5] N. Ramli, R. R. Othman, Z. I. A. Khalib, and M. Jusoh, "A Review on Recent T-way Combinatorial Testing Strategy," in 2017 MATEC Web of Conferences, 2017, doi.org/10.1051/matecconf/201714001016

[6] A. A. Al-Sewari and K. Z. Zamli, "An orchestrated survey on t-way test case generation strategies based on optimization algorithms," in 2014 The 8th International Conference on Robotic, Vision, Signal Processing & Power Applications, 2014, 255–263, doi.org/10.1007/978-981-4585-42-2_30.

[7] C. A. Floudas et al., "Handbook of test problems in local and global optimization. 1999." Kluwer Academic Publishers, Dordrecht.

[8] A. S. M. Ali, R. R. Othman, Y. M. Yacob, "Application Of Adaptive Elitism Operator In Firefly Algorithm For Optimization Of Local," **99**(3), 569–582, 2021.

[9] O. H. Yeh And K. Z. Zamli, "Development of a non-deterministic input-output based relationship test data set minimization strategy," in 2011 IEEE Symposium on Computers & Informatics, 2011, 800–805, doi:10.1109/ISCI.2011.5959020.

[10] A. A. Alsewari, N. M. Tairan, and K. Z. Zamli, "Survey on Input Output Relation based Combination Test Data Generation Strategies," ARPN Journal Engineering Applied Science, **10**(18) 8427–8430, 2015.

[11] A. A. Alsewari, L. M. Xuan, and K. Z. Zamli, "Firefly Combinatorial Testing Strategy," in Science and Information Conference, 2018, 936–944.

[12] K. Z. Alsewari, AbdulRahman A. Xuan, Lin Mee Zamli, "Firefly Combinatorial Testing Strategy," in Intelligent Computing, 2019.

[13] A. S. M. Ali, R. R. Othman, Y. M. Yacob, and J. M. Alkanaani, "ParametersTuning of Adaptive Firefly Algorithm based Strategy for t-way Testing." International Journal of Innovative Technology and Exploring Engineering, **9**, 4185–4191, doi: 10.35940/ijitee.A6111.119119.

[14] A. Pandey and S. Banerjee, "Test Suite Optimization Using Chaotic Firefly Algorithm in Software Testing," International Journal Applied Metaheuristic Computing, **8**(4) 41–57, 2017, doi: 10.4018/978-1-7998-3016-0.ch032.

[15] P. J. Schroeder and B. Korel, Black-box test reduction using input-output

analysis, **25**(5), 173–177, 2000, doi: 10.1145/347636.349042.
[16] N. Ramli, R. R. Othman, and M. S. A. R. Ali, "Optimizing combinatorial input-output based relations testing using Ant Colony algorithm," in 2016 3rd International Conference on Electronic Design (ICED), 2016, 586–590, doi**:** 10.1109/ICED.2016.7804713.
[17] A. B. Nasser and K. Z. Zamli, "A new variable strength T-way strategy based on the cuckoo search algorithm," in Intelligent and Interactive Computing, Springer, 193–203, 2019, doi: 10.1109/ICED.2016.7804713.
[18] W. Ziyuan, N. Changhai, and X. Baowen, "Generating combinatorial test suite for interaction relationship," in 2007 Fourth international workshop on Software quality assurance: in conjunction with the 6th ESEC/FSE joint meeting, 2007, 55–61, doi: 10.1145/1295074.1295085.
[19] P. J. Schroeder, P. Faherty, and B. Korel, "Generating expected results for automated black-box testing," in 2002 Proceedings 17th IEEE International Conference on Automated Software Engineering, 2002, 139–148, 2002, doi:10.1109/ASE.2002.1115005.
[20] X.-S. Yang, "Othman & Zamli," Nature-inspired metaheuristic algorithms, **20**, 79–90, 2008.
[21] A. H. Gandomi, X. Yang, S. Talatahari, and A. H. Alavi, "Commun Nonlinear Sci Numer Simulat Firefly algorithm with chaos," Communication Nonlinear Science Numerical Simulation, **18**(1) 89–98, 2013, doi: 10.1016/j.cnsns.2012.06.009
[22] X. Yang, "Firefly Algorithm , L ´ evy Flights and Global Optimization arXiv : 1003 . 1464v1 [math . OC] 7 Mar 2010," 1–10, 2010.
[23] X.-S. Yang, Nature-Inspried Metaheuristic Algorithms, Second Edition.(July 2010.
[24] X. Yang, "Firefly Algorithm : Recent Advances and Applications," **1**(1), 36–50, 2008. doi: 10.1504/IJSI.2013.055801.
[25] M. I. Younis, A. R. A. Alsewari, N. Y. Khang, and K. Z. Zamli, "CTJ: Input-Output Based Relation Combinatorial Testing Strategy Using Jaya Algorithm," Baghdad Science Journal, **17**(3), 1002-1009, 2020, doi: 10.21123/bsj.2020.17.3(Suppl.).1002.
[26] H. Y. Ong and K. Z. Zamli, "Development of interaction test suite generation strategy with input-output mapping supports," Sci. Res. Essays, **6(**16) 3418–3430, 2011, doi: 10.5897/SRE11.427.
[27] S. Esfandyari and V. Rafe, "GALP: a hybrid artificial intelligence algorithm for generating covering array," Software Computing, **25**(11) 7673–7689, 2021, DOI: 10.1007/s00500-021-05788-0.
[28] L. Statistics, "Wilcoxon Signed Rank Test in SPSS Statistics-Procedure, Output and Interpretation of Output Using a Relevant Example." 2017.
[29] D. C. Montgomery and G. C. Runger, Applied statistics and probability for engineers. John Wiley and Sons, 2014.
[30] D. C. Montgomery and G. C. Runger, Applied statistics and probability for engineers. John Wiley and Sons, 2014.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 233-240 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

A New Topology Optimization Approach by Physics-Informed Deep Learning Process

Liang Chen*, Mo-How Herman Shen

Mechanical and Aerospace Engineering, The Ohio State University, Columbus, OH, 43210, USA

ARTICLE INFO

Article history:
Received: 03 May, 2021
Accepted: 07 July, 2021
Online: 27 July, 2021

Keywords:
Physics-Informed Neural Network
Topology Optimization
Automatic Differentiation

ABSTRACT

In this investigation, an integrated physics-informed deep learning and topology optimization approach for solving density-based topology designs is presented to accomplish efficiency and flexibility. In every iteration, the neural network generates feasible topology designs, and then the topology performance is evaluated using the finite element method. Unlike the data-driven methods where the loss functions are based on similarity, the physics-informed neural network weights are updated directly using gradient information from the physics model, i.e., finite element analysis. The key idea is that these gradients are calculated automatically through the finite element solver and then backpropagated to the deep learning neural network during the training or intelligence building process. This integrated optimization approach is implemented in Julia programming language and can be automatically differentiated in reverse mode for gradient calculations. Only forward calculations must be executed, and hand-coded gradient equations and parameter update rules are not required. The proposed physics-informed learning process for topology optimization has been demonstrated on several popular 2-D topology optimization test cases, which were found to be a good agreement with the ones from the state-of-the-art topology optimization approach.

1. Introduction

1.1. Literature Review

Classical gradient-based optimization algorithms [1], [2] require at least first-order gradient information to minimize/ maximize the objective function. If the objective function can be written as a mathematical function, then the gradient can be derived analytically. For a complex physics simulator, such as finite element method (FEM), computational fluid dynamics (CFD), and multibody dynamics (MBD), gradients are not readily available to perform design optimization. To overcome this problem, a data-driven approach, such as response surface methodology or surrogate modelling [3], is applied to model the simulator's behavior, and the gradients are obtained rapidly. The response surface methodology and surrogate modeling approximate the relationship between design variables (inputs) and the objective values (outputs). Once the approximated model is constructed, gradient information for optimization are much cheaper to evaluate. However, this approach may require many precomputed data points to construct an accurate model depending on the dimensionality of the actual system. Therefore, generating many data points arises as one of the bottlenecks of this method in practice due to the high computational cost of physics simulations.

In addition, the data-driven approach may not generalize the behavior of the physics simulation accurately. This would cause the approximated model to produce invalid solutions throughout the design space except at the precomputed data points. This will be detrimental in optimization, where the best solution must be found among all true solutions.

A possible reason data-driven approach fails is its reliance on the function approximator to learn the underlying physical relationships using a finite amount of data points. In theory, a function approximator can predict the relationship between any input and output according to the universal approximation theorem [4], [5]. However, this usually requires enough capacity and complexity of the function and sufficient data points to train the function. These data points are expensive to obtain and a significant amount of effort must be placed towards computing the objective values of each data point through a physics simulator. In the end, only the design variables and corresponding objective values are paired for further analysis (surrogate modeling and optimization). All the information during the physics simulation is discarded and wasted. Many ideas have been proposed to solve topology optimization using data-driven approach. In [6], [7] the

*Corresponding Author: Liang Chen, 201 W. 19th Ave, Columbus, OH, 43210, chen.4853@osu.edu

showed the GAN's capability of generating new structure topologies. In [8], the authors proposed a TopologyGAN to generate a new structure design given arbitrary boundary and loading conditions. In [9], the authors designed a GAN framework to generate a new airfoil shape. All these approaches are data-driven methods where large dataset must be prepared to train the neural network for surrogate modeling.

To use the computational effort during physical simulations, we turn to a newly emerged idea called differential programming [10], [11] which applies automatic differentiation to calculate the gradient of any mathematical/ non-mathematical operations of a computer program. Because of this, we can make the physics model be fully differentiable and integrate it with a deep neural network and use backpropagation for training. In [12], the authors discussed the application of automatic differentiation on density-based topology optimization using existing software packages. In [13], [14] the authors demonstrated the ability to use automatic differentiation to calculate gradients for fluid and heat transfer optimization. Furthermore, in [15] the authors modeled the solution of a partial differential equation (PDE) as an unknown neural network. Plugging the neural network into the PDE forces the neural network to follow the physics governed by the PDE. A small amount of data is required to train the neural network to learn the real solution of the PDE. In [16], [17] the authors demonstrated the application of a physics-informed neural network on high-speed flow and large eddy simulation. In [18], the authors designed an NSFnet based on a physics-informed neural network to solve incompressible fluid problems governed by Navier-Stokes equations. In this work, we will design optimal structural topology by integrating the neural network and the physics model(FEM) into one training process. The FE model is made fully differentiable through automatic differentiation to provide critical gradient information for training the neural network.

1.2. Review of Solid Isotropic Material with Penalization Method (SIMP)

The SIMP method proposed by Bendsoe [19] established a procedure (Figure 1) for density-based topology optimization.

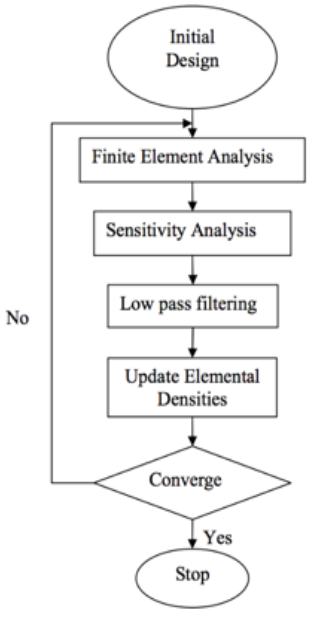


Figure 1: SIMP Method Optimization Process

The algorithm begins with an initial design by filling in a value x_e in each quadrilateral element of a predefined structured mesh in a fixed domain. The value represents the material density of the element where 0 means a void and 1 means filled. A value between 0 and 1 is partially filled, which does not exist in reality. It makes optimization easy but results in a blurry image of the design. Therefore, the author [19] proposed to add a penalization factor to push the element density towards either void or completely filled.

The objective of the SIMP method is to minimize the compliance, C, of the design domain under fixed loadings and boundary conditions. The compliance defined in (1), also described as total strain energy, is a measure of the overall displacement of a structure.

$$C(x) = U^T K U = \sum_{e=1}^{N} E_e(x_e) u_e^T K_0 u_e \tag{1}$$

$$\frac{dC}{dx_e} = \frac{dE_e}{dx_e} u_e^T K_0 u_e \tag{2}$$

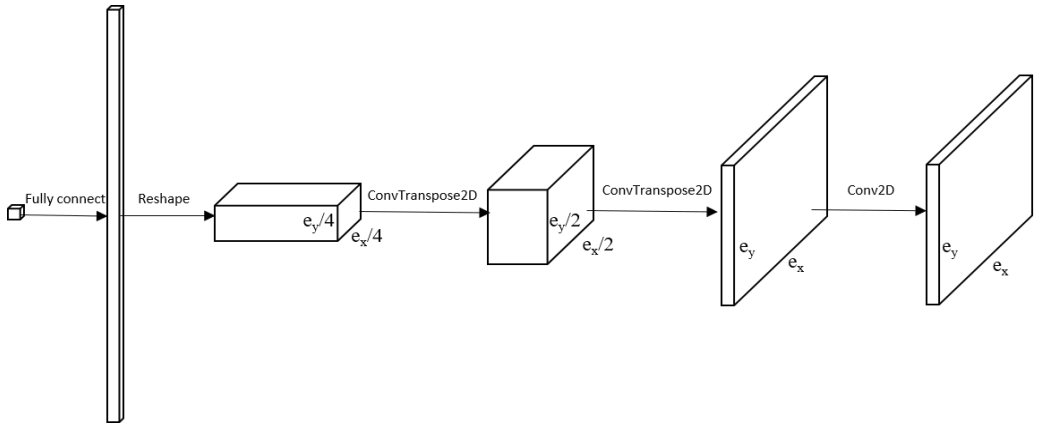
$$x_e^{new} = \begin{cases} \max(0, x_e - m) & if\ x_e B_e^{\eta} \le \max(0, x_e - m) \\ \min(1, x_e + m) & if\ x_e B_e^{\eta} \ge \min(1, x_e - m) \\ x_e B_e^{\eta} & otherwise \end{cases} \tag{3}$$

$$where\ B_e^{\eta} = \frac{-\frac{\partial C}{\partial x_e}}{\lambda \frac{\partial V}{\partial x_e}}$$

Inside the optimization loop, the density of each element x_e has to be updated to lower the compliance. For a gradient-based method, the gradient dC/dx is calculated as (2) by taking a derivative of (1). Then a heuristic update rule, (3), is crafted to ensure the optimality condition is satisfied for the design. In other popular density-based topology optimization techniques, such as level-set [20], [21] or bidirectional evolutionary optimization [22] an analytical formula of gradient, element density derivative, or shape derivative must be provided manually. In the proposed framework in Section 2, there is no need to provide such gradient information analytically as the gradients are calculated by reverse mode automatic differentiation.

1.3. Motivation

In this work, the goal is to integrate a physics-based model (e.g., finite element model) with a neural network for generating feasible topology designs so that the gradient information obtained from the finite element model can be used to minimize the loss function during the training process to achieve optimal topology designs. The critical gradient information and update rules are determined using automatic differentiation and ADAM optimizer, eliminating hand-coded equations, such as (5) and (6). Furthermore, the critical gradient information directly obtained in the finite element model would contain essential features which may not exist or are significantly diluted in the training data sets.



In addition, the proposed approach does not require a significant amount of time for preparing the training data set.

In Section 2, we have constructed a new topology optimization procedure by integrating a deep learning neural network with a wildly used classical SIMP topology optimization method to achieve efficiency and flexibility. This integrated physics-informed deep learning optimization process is presented in Section 2 and has been implemented in Julia programming language, which can be automatically differentiated in reverse mode for gradient calculations. The proposed integrated deep learning and topology optimization approach has been demonstrated on several famous 2-D topology optimization test cases, which were found to be a good agreement with the SIMP method. Further validation of the idea and approach via complex structural systems and boundary and/or loading conditions will be conducted in our future work.

2. Proposed Physic-Informed Deep Learning Process

2.1. Mathematical Formulation

In this work, we will perform the topology optimization with integrated physics-informed neural network. The neural network is physics-informed because its parameters are updated using gradient information directly from the finite element model.

A flow chart of the proposed physics-informed deep learning design optimization procedure is outlined in Figure 2. The equivalent mathematical formulation can be stated as:

$$\min_{\theta} L = U_{load}$$

subject to:

$$x = G(v;\theta)$$

$$M(x) = v$$

$$U = F \backslash K(x) \tag{4}$$

The U_{load} is the displacement at the applied load point and it is the loss function, L, to be minimized. The input to the generator is the target mass fraction,v. The generator, $G(.;\theta)$, is a deep neural network with parameter θ and it outputs a topology, x, based on the target mass fraction. The $M(x)$ returns the mass fraction of the topology produced by generated and compare it to the target mass fraction. The $K(x)$ and F are the stiffness matrix of the topology and load vector, respectively. They are processed in the FEM component in Figure 2, and a resultant displacement U is returned.

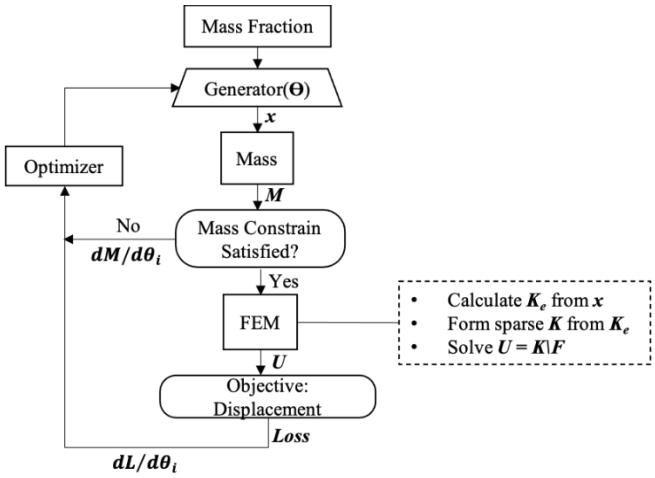
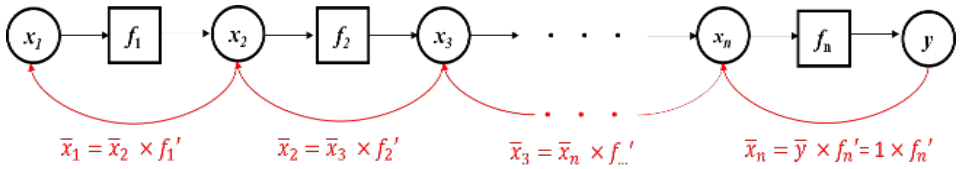


Figure 3: Proposed Physics-Informed Deep Learning Topology Optimization Approach

Instead of providing a hand-coded gradient equation, (1.2), and update rule, (1.3), the gradient of every operation throughout the calculation is determined by using reverse differentiation (ReverseDiff). The gradient is a vector of dL/dx_e, where L is a scalar value of the objective function. In the examples shown in Section 4, the compliance/displacement is the objective function, while a given total mass is equality constrain that must be satisfied. Instead of providing an update rule, the generator, *G(.)*, proposes a new design (x) for every iteration. The parameters of the network are adjusted simultaneously to generate better design while satisfying the mass constraint. Therefore, in the proposed algorithm, the gradients that guide the design approaching optimal are not dL/dx_e, but $dL/d\theta$ instead. For a 1-D design case where the design variables have no spatial dependencies, the neural network is a multilayer perceptron. While for 2D cases, the neural network architecture is composed of layers of the convolutional operator, which is good at learning patterns of 2-D images. Thus, the objective of this topology optimization becomes learning a set of parameters, θ, such that the generator can generate a 2D structure, x, such that it minimizes the displacement at the point of load while satisfying the mass constraint.

The optimizer component in Figure 2 is the algorithm to update the variables given the gradient information. In this work, ADAM (ADAptive Moments) [23] is used to handle the learning rate and update the parameter θ of the generator in each iteration.



2.2. Generator Architecture

The network architecture of the generator in is illustrated in Figure 3. Adapting the idea of the generator from Generative Adversarial Network (GAN) [24], which can generate high-quality images after training, the generator starts with a seed value (fixed or random) which followed by a fully connected layer. Then the fully connected layers are reshaped to a feature layer composed of a 3D array. The first two dimension relate to the feature layer's width and height, and the third dimension refers to channels. Then the following components are to expand the width and height of the feature layer but shrink the number of channels down to 1. In the end, the output will be a 2D (3rd dimension is 1) image with correct size $e_x \times e_y$ based on the learnable parameter θ. In Figure 3, the output size of the ConvTranspose2D layer doubles every time. The second last layer has an activation function, $tanh(.)$, making sure the output values are bounded. The last convolutional layer does not change the size of the input, and its parameters are predetermined and kept fixed. The purpose of the last layer is to average the density value around the neighborhood of each element to eliminate the checkerboard pattern. This is similar to the filtering technique used in the SIMP method [25], [26].

The generator architecture can vary by adding more layers or new components. Instead of starting with a size of $e_x/4 \times e_y/4$, one can make this even smaller and add more channels at the beginning. Hyper-parameters such as the kernel sizes and activation functions can be applied to any layers of the generator. However, in this paper, the architecture design is kept simple without experimenting too much with the hyper-parameters.

3. Automatic Differentiation

The idea of automatic differentiation uses derivative rules of general operations to calculate the derivatives of outputs to inputs of a composite function $F: R_m \rightarrow R_n$, where m and n are the dimensions of inputs and outputs, respectively. In general, the resulting derivatives form an $m \times n$ Jacobian matrix. When $m > n = 1$, the derivatives form a gradient vector with length m. The automatic differentiation can be done in two ways: forward mode [27] or reverse mode [11]. In the following, ForwardDiff and ReverseDiff will be used for short. Usually, the entire composite function F has no simple derivative rule associated with it. However, since it can be decomposed into other functions or mathematical operations where the derivatives are well defined, the chain rule can be applied to propagate the derivative information either forward (ForwardDiff) or backward (ReverseDiff) through the computation graph of a composite function$F$. The ForwardDiff has linear complexity with respect to the input dimension, and therefore, it is efficient for a function when $m \ll n$. Further detail of the forward differentiation can be found in reference [27]. Conversely, the ReverseDiff is ideal for a function when $m \gg n$, and applied in this paper to calculate derivatives. ReverseDiff can be taken advantage of in this situation as the dimension of the inputs (densities of each element) is very large, but the output dimension is just one scalar value, such as overall compliance or displacement.

3.1. Reverse Mode Differentiation (ReverseDiff)

The name, ReverseDiff, comes from a registered Julia package called ReverseDiff.jl and additionally has a library of well-defined rules for reverse mode automatic differentiation. The idea of ReverseDiff is related to the adjoint method [11], [20], [28], [29] and applied in many optimization problems where the information from the forward calculation of the objective function is reused to calculate the gradients efficiently. In structure optimization, the adjoint method solves an adjoint equation $K\lambda = z$ [30], [31]. Since K^{-1} is known when solving $KU = F$ during the forward pass, then the adjoint equation can be solved with a small computational cost without factorizing or calculating the inverse of matrix again. The ReverseDiff is also closely related to backpropagation [32] for training a neural network in machine learning. The power of ReverseDiff is that it takes the automatic differentiation into a higher level, where any operation (mathematical or not mathematical) can be assigned with a derivative rule (called pullback function). Then with the chain rule, we can combine the derivatives of every single operation and compute the gradient from end to end of any black-box function, i.e., physics simulation engine or computer program. To formulate the process for ReverseDiff, a composite function : $F: y = f_n(f_{n-1}(\cdots f_2(f_1(x_1))))$ is considered, where any intermediate step can be written as $x_{n+1} = f_n(x_n)$. A computation graph of the composite function is shown in Figure 4. For simplicity, it is assumed that any intermediate function inside the composite function is a single input and a single output.

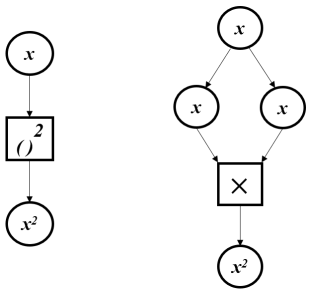


Figure 5: Computation Graphs of x^2 (left) and x*x (right)

The black arrows are the forward pass for function evaluations, and the red arrows are reverse differentiation with chain rule been applied. The idea can be generalized to multiple inputs and outputs as well. To ReverseDiff of the function from end to end in Figure 4, we will define and expand derivative using chain rule as:

$$\bar{x}_1 = \frac{dy}{dx_1} \tag{5}$$

$$= \frac{dy}{dy} \times \frac{dy}{dx_n} \times \ldots \times \frac{dx_3}{dx_2} \times \frac{dx_2}{dx_1}$$

On the right side of the (3.1), y is called seed, and its value equals 1. Starting from the last node, y, it essentially calculates $\bar{x}_i = dy/dx_i$ as going backward through the computation graph. It can be shown that

$$\bar{x}_i = \frac{dy}{dx_{i+1}} \times \frac{dx_{i+1}}{dx_i} = \bar{x}_{i+1} \times f_i' \qquad (6)$$

In other words, for any function $x_{i+1} = f(x_i)$, the derivative of the input x_i can be determined from the derivative of the output x_{i+1} multiplied by f_i. For ReverseDiff, (3.2) is known as the pullback function, as it pulls the output derivative in backward to calculate the derivative of the input. To evaluate the derivative using the pullback function, we need to know the output derivative and the value of x_i as well. Thus, a forward evaluation of all the intermediate values must be done and stored first before the reverse differentiation process. Theoretically, the computation time of the ReverseDiff is proportional to the number of outputs, whereas it scales linearly with the number of inputs for ForwardDiff. In practice, ReverseDiff also requires more overhead and memory to store the information during forward calculation. In the examples used in this paper, the input space dimension is on the order of 10^3 to 10^4, but the output space dimension is only 1.

The pullback function is the rule needed to implement for every single operation during the forward calculation. For example, suppose the function we want to evaluate is $f: y = \sin(x_2)$. Then we need to write a generic pullback function for $b_1 = a^n$ as $a_1 = b_1 \times n a^{n-1}$ and for $b_2 = \sin(a_2)$ as $a_2 = b_2 \times cos(a_2)$. For the sake of demonstration, we can then combine these two pullback functions as one (will not do in practice as chain rule will take care of) it as:

$$B^f(\bar{y}) = \bar{y}\cos(x^2) \times 2x \qquad (7)$$

Notice that the pullback function of the exponent x^2 is known from calculus. However, we can also treat it as a multiplication (a fundamental mathematical operation) of two numbers. The computation graph is shown in Figure 5. The pullback function for multiplication of two real numbers, $y = x_1 \times x_2$, is: $B^f(\bar{y}) = \bar{y} \times x_2, \bar{y} \times x_1$. Then the pullback function of the input x in Figure 5 is the summation of the two terms in (3.3), which is $y(x_1 + x_2) = y \times 2x$ when $x_1 = x_2 = x$. As can be seen from the example above, any high-level operations can be decomposed into a series of elementary operations such as addition, multiplication, *sin*(.), and *cos*(.). Then the pullback function of the high-level operations can always be inferred from the known pullback functions with the chain rule. However, it will be a timesaver if the rules for some high-level operations can be defined directly. Just like the exponent function, it takes much longer computationally to convert it into a series of multiplications, while using a given rule from calculus, such as $x_n = nx_{n-1}$, is much more efficient. In the structural topology examples demonstrated in Section 4, we will define a custom pullback rule for the backslash operator of the sparse matrix, which results in a much more efficient calculation of the gradient. The implementation details of this rule are discussed in Section 3.2.

3.2. Automatically Differentiate Finite Element Model

A standard forward calculation must be coded as illustrated in the dashed box in Figure 1 to construct a differentiable FEM solver. Then we need to make sure every operation in the forward calculation has a pullback function associated with ReverseDiff. Most elementary mathematical operations in linear algebra have pullback functions well-defined in Julia. However, the operation for solving $KU = F$, where K is a sparse matrix, has not been defined. Therefore, it is important to write an efficient custom rule for the backslash operator $U = K\backslash F$. There are two parts associated with the backslash operation. The first part is to construct a sparse matrix $K = sparse(I, J, V)$, where I and J are the vectors of row and column indices for non-zero entries and V is a vector of values associated with each entry. The pullback function of the operation is defined as $\bar{V}$ ($\bar{K}$) = NonZerosOf($\bar{K}$). The process is illustrated in Figure 6. The second part is for the backslash operation. For a symmetric dense matrix, K, the pullback function of $U = K \setminus F$ can be written as:

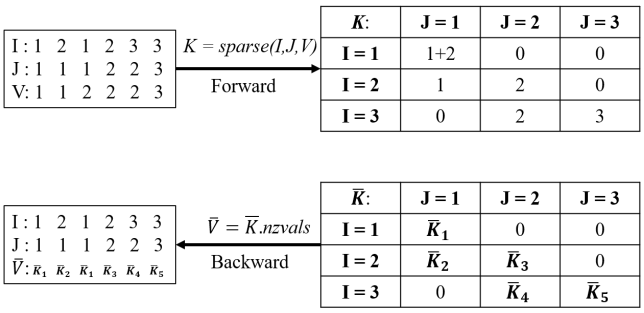


Figure 6: Forward and Backward Calculation of sparse(I,J,V) function

The $\bar{U}$ is the derivative of each element of U with respect to the downstream objective function. In (4.1), it requires two backslash operations, but in practice, the factorization or inverse of K can be reused, and only one expensive factorization is needed. This is why the ReverseDiff can "automatically" calculate the gradient by only evaluating the function in forward mode.

$$\bar{K}(\bar{U}) = -\bar{F}U', \quad where\ U = K\backslash F, \bar{F} = K\backslash\bar{U} \qquad (4.1)$$

When K is sparse, (4.1) can be done efficiently by only using the terms in $\bar{F}$ and U that correspond to the nonzero entries of the sparse matrix K. Otherwise, (4.1) will result in a dense K matrix that takes up significant memory.

3.3. Programming Language

Julia is used as the programming language. Julia has an excellent ecosystem for scientific computing and automatic differentiation. We use a Julia registered package called ChainRules.jl to define the custom pullback function of the finite element solver. Flux.jl was used to construct the neural network and for optimization.

4.1. 2D Density-based Topology Optimization

Figure 7 is a well-known MBB beam (simply supported beam) for the benchmark test in topology optimization. The objective is to minimize the compliance of the beam subjected to a constant point load applied in the center. Due to symmetry about the vertical axis, the design domain (Figure 7) only includes half of the original problem. The objective function minimizes the displacement at the load point. The equality constraint is that the overall mass fraction

of the optimized topology must be consistent with the target mass fraction at the input layer to the generator.

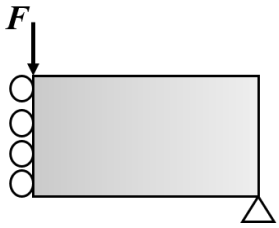


Figure 7: Design Domain and Boundary Conditions of MBB Beam

Figure 8 shows the convergence of MBB beam topology design within 100 iterations. The mass fraction is kept as 0.3. It is shown that the objective value in the vertical axis almost flats out at 40 iterations, where the design from the generator stays almost the same afterward.

When assembling the global stiffness matrix K from the element density values, x, the actual density value is penalized using x^p, where $p \geq 1$. The penalty factor eliminates the checkerboard patterns and creates a smooth boundary. Figure 9 shows side-by-side comparisons of the results from our proposed approach and SIMP 88-line code. The experiment ran combinations of three target mass fraction values and two penalty values for both methods. All designs in Figure 9 are generated after 100 iterations for the MBB beam. The number below each optimized design is the magnitude of the displacement at the applied load point. The proposed method works very well with a low mass fraction design. However, when the mass fraction increases, the details of the design are hard to capture, and a higher penalty value is required to make the design clearer. Although the details of designs from the two methods are different, both methods result in close displacement values. This means the optimized structures have equivalent overall stiffness given the mass fraction, loading, and boundary conditions.

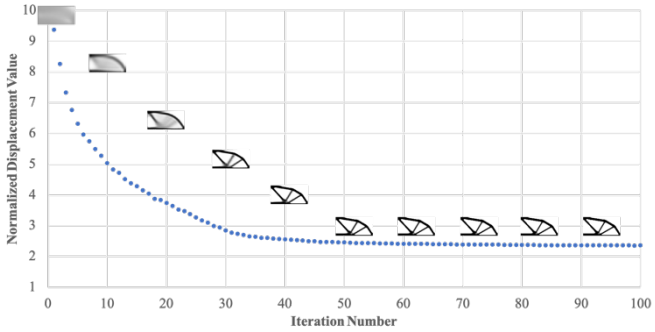


Figure 8: Convergence of the Objective Function of the Proposed Method

Table 1 shows the computation time of the proposed method on a 2015 Mac with 2.7GHz Intel i5 Dual Core and 8Gb memory. The time in the table is an average of 100 iterations. The actual time of each iteration varies due to the convergence of the mass constrain in the inner loop as shown in Figure 1.

Table 1: Computation Time of Proposed Method for Each Iteration

Mesh Size	Time (seconds) per Iteration
48*24	0.08
96*48	0.28
192*96	1.3

Figure 10 shows two more cases with different loading and boundary conditions (cantilever beam and bridge) on the density-based topology optimization using the proposed design framework. The cantilever beam on the left has fixed boundary conditions on the left side and a tip load at the midpoint on the right side. The bridge design on the right is simply supported at the lower left and restricted in the vertical direction at lower right. A point load is applied at the midpoint at the bottom surface. The optimized structure using the proposed method does not look exactly like the one using SIMP 88-line code. However, they show similar trends for most of the cases. In Figure 10, the displacement at the applied load point is less using the proposed method, which means the structure has a higher stiffness.

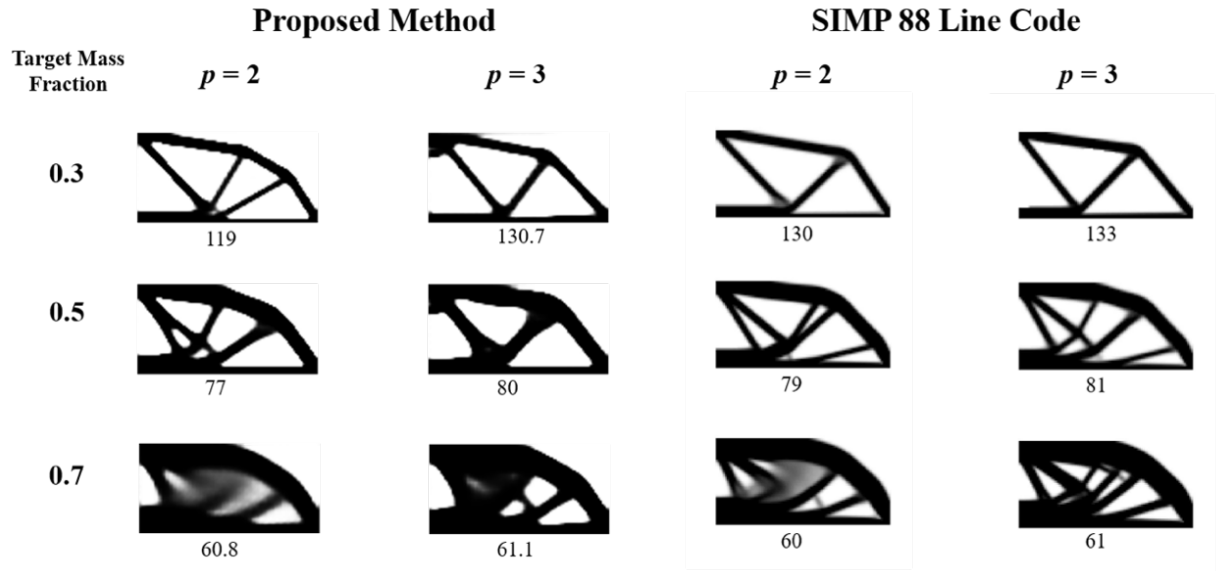


Figure 9: Comparisons of Results Between the Proposed Method and 88-line code (Value under each image is the displacement at load point)

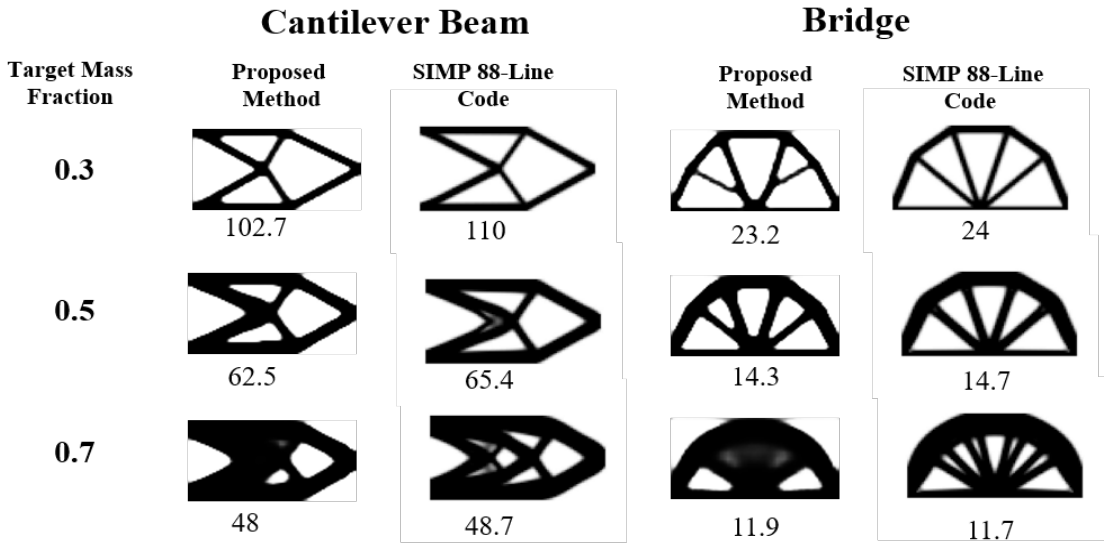


Figure 10: Comparisons of Cantilever Beam (Left) and Bridge (Right) Designs using Proposed and SIMP 88-line Method (Value under each image is the displacement at load point)

4.2. 2D Compliant Mechanism Optimization

To demonstrate the flexibility of our proposed method, we use this approach to perform optimization for compliant mechanism design. For a compliant mechanism, the structure must have low compliance at the applied load point but high flexibility at some other part of the structure for desired motion or force output. A force inverter, for example, requires the displacement at point 2 to be in the opposite direction to the applied load at point 1.

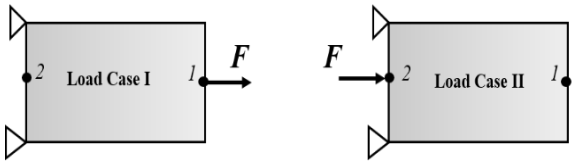


Figure 11: Two Load Cases for Optimizing a Compliant Mechanism

Therefore, the objective function has to be formulated as a combination of two parts:

$$L = \frac{U_2^{\mathrm{I}}}{U_1^{\mathrm{I}}} + w[U_2^{\mathrm{II}} + U_1^{\mathrm{I}}] \qquad (4.1)$$

The first term on the right-hand side is the geometry advantage, which is the ratio of the output displacement over input displacement for load case I. This term has to be minimized because the output displacement has to be negative, opposite the input force direction. Also, this term needs to be as negative as possible. However, only this term in the objective function will result in intermediate density and a fragile region around point 2. A second term is added to solve this problem, which measures the compliance of the overall structure, and we want to make sure the structure is stiff when the load is applied at either point 1 or 2. The superscripts *I* and *II* denote two different loading cases (Figure 11). For case *I*, the force is applied at point 1, and displacements are recorded at points 1 and 2. For case II, the force is applied at point 2, and only the displacements at point 2 are recorded. Both displacements at the loading points of the two cases must be minimized to achieve low compliance. Therefore, for each iteration, two load cases, as opposed to one in the previous example, have to be run, and the displacement values will then be fed to the objective functions. The weight coefficient of the second term in (4.2), w, has to be a small number (< 0.1). Otherwise, the design will be too stiff and result in minimal geometry advantage. The image on the left of Figure 12 is the optimized design of this force inverter where $w = 0.01$ and 0.3 volume ratio. It achieves a geometry advantage of $226/47$.

Figure 12: Optimized force inverter with different objective functions. Left: use geometry advantage; Right: use target displacement

5. Conclusions

In this work, an integrated physic-informed deep neural network and topology optimization approach is presented as an efficient and flexible way to solve the topology optimization problem. We can calculate the end-to-end gradient information of the entire computational graph by combining the differentiable physics model (FEM) with deep learning layers. The gradient information is used efficiently during the training process of the neural network. We demonstrated the proposed optimization framework on different test cases. Without explicitly specifying any hand-coded equations for gradient calculation and update rules, the neural network after training can learn and produce promising results. The generated optimized structure achieves the same level of overall stiffness as the well-known SIMP method.

The proposed framework is much simpler to implement as only the forward calculations and basic derivative rules are required. For the compliant mechanism design, only the objective function is required to be implemented, and the proposed method can achieve different designs that satisfy the design targets.

6. Declarations

6.1. Funding

No funding was received to assist with this study and the preparation of this manuscript.

6.2. Conflicts of interests/Competing interests

All authors have no conflicts of interest to disclose.

6.3. Availability of data and materials

All data generated or analyzed during the current study are available from the authors on reasonable request.

6.4. Code availability/Replication of results

The data and code will be made available at reader's request.

Reference

[1] S.S. Rao, Engineering optimization: theory and practice, John Wiley & Sons, 2019.

[2] A. Kentli, "Topology optimization applications on engineering structures," Truss and Frames—Recent Advances and New Perspectives, 1–23, 2020, doi: 10.5772/intechopen.90474.

[3] R.H. Myers, D.C. Montgomery, C.M. Anderson-Cook, Response surface methodology: process and product optimization using designed experiments, John Wiley & Sons, 2016.

[4] A.N. Gorban, D.C. Wunsch, "The general approximation theorem," in 1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98CH36227), 1271 - 1274, 1998, doi: 10.1109/ijcnn.1998.685957

[5] H. Lin, S. Jegelka, "Resnet with one-neuron hidden layers is a universal approximator," in Advances in neural information processing systems, 6169–6178, 2018.

[6] S. Rawat, M.H. Shen, "A novel topology design approach using an integrated deep learning network architecture," ArXiv Preprint ArXiv:1808.02334, 2018.

[7] M.-H.H. Shen, L. Chen, "A New CGAN Technique for Constrained Topology Design Optimization," ArXiv Preprint ArXiv:1901.07675, 2019.

[8] Z. Nie, T. Lin, H. Jiang, L.B. Kara, "Topologygan: Topology optimization using generative adversarial networks based on physical fields over the initial domain," Journal of Mechanical Design, **143**(3), 31715, 2021, doi: 10.1115/1.4049533

[9] W. Chen, K. Chiu, M.D. Fuge, "Airfoil Design Parameterization and Optimization Using Bézier Generative Adversarial Networks," AIAA Journal, **58**(11), 4723–4735, 2020, doi: 10.2514/1.j059317

[10] M. Innes, A. Edelman, K. Fischer, C. Rackauckus, E. Saba, V.B. Shah, W. Tebbutt, "Zygote: A differentiable programming system to bridge machine learning and scientific computing," ArXiv Preprint ArXiv:1907.07587, 140, 2019.

[11] M. Innes, "Don't unroll adjoint: differentiating SSA-Form programs," ArXiv Preprint ArXiv:1810.07951, 2018.

[12] S.A. Nørgaard, M. Sagebaum, N.R. Gauger, B.S. Lazarov, "Applications of automatic differentiation in topology optimization," Structural and Multidisciplinary Optimization, **56**(5), 1135–1146, 2017, doi: 10.1007/s00158-017-1708-2

[13] S.B. Dilgen, C.B. Dilgen, D.R. Fuhrman, O. Sigmund, B.S. Lazarov, "Density based topology optimization of turbulent flow heat transfer systems," Structural and Multidisciplinary Optimization, **57**(5), 1905–1918, 2018, doi: 10.1007/s00158-018-1967-6

[14] A. Vadakkepatt, S.R. Mathur, J.Y. Murthy, "Efficient automatic discrete adjoint sensitivity computation for topology optimization--heat conduction applications," International Journal of Numerical Methods for Heat & Fluid Flow, 2018, doi: 10.1108/hff-01-2017-0011

[15] M. Raissi, P. Perdikaris, G.E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," Journal of Computational Physics, **378**, 686–707, 2019, doi: 10.1016/j.jcp.2018.10.045

[16] Z. Mao, A.D. Jagtap, G.E. Karniadakis, "Physics-informed neural networks for high-speed flows," Computer Methods in Applied Mechanics and Engineering, **360**, 112789, 2020, doi: 10.1016/j.cma.2019.112789

[17] X.I.A. Yang, S. Zafar, J.-X. Wang, H. Xiao, "Predictive large-eddy-simulation wall modeling via physics-informed neural networks," Physical Review Fluids, **4**(3), 34602, 2019, doi: 10.1103/physrevfluids.4.034602

[18] X. Jin, S. Cai, H. Li, G.E. Karniadakis, "NSFnets (Navier-Stokes flow nets): Physics-informed neural networks for the incompressible Navier-Stokes equations," Journal of Computational Physics, **426**, 109951, 2021, doi: 10.1016/j.jcp.2020.109951

[19] M.P. Bendsoe, O. Sigmund, Topology optimization: theory, methods, and applications, Springer Science & Business Media, 2013.

[20] G. Allaire, "A review of adjoint methods for sensitivity analysis, uncertainty quantification and optimization in numerical codes," 2015.

[21] M.Y. Wang, X. Wang, D. Guo, "A level set method for structural topology optimization," Computer Methods in Applied Mechanics and Engineering, **192**(1–2), 227–246, 2003.

[22] X.Y. Yang, Y.M. Xie, G.P. Steven, O.M. Querin, "Bidirectional evolutionary method for stiffness optimization," AIAA Journal, **37**(11), 1483–1488, 1999,doi: 10.2514/2.626

[23] D.P. Kingma, J. Ba, "Adam: A method for stochastic optimization," ArXiv Preprint ArXiv:1412.6980, 2014.

[24] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative adversarial nets," in Advances in neural information processing systems, 2672–2680, 2014.

[25] E. Andreassen, A. Clausen, M. Schevenels, B.S. Lazarov, O. Sigmund, "Efficient topology optimization in MATLAB using 88 lines of code," Structural and Multidisciplinary Optimization, **43**(1), 1–16, 2011, doi: Andreassen_2010

[26] B. Bourdin, "Filters in topology optimization," International Journal for Numerical Methods in Engineering, **50**(9), 2143–2158, 2001, doi: 10.1002/nme.116

[27] J. Revels, M. Lubin, T. Papamarkou, "Forward-mode automatic differentiation in Julia," ArXiv Preprint ArXiv:1607.07892, 2016.

[28] G. Allaire, F. Jouve, A.-M. Toader, "Structural optimization using sensitivity analysis and a level-set method," Journal of Computational Physics, **194**(1), 363–393, 2004,doi: 10.1016/j.jcp.2003.09.032

[29] R.M. Errico, "What is an adjoint model?," Bulletin of the American Meteorological Society, 78(11), 2577–2592, 1997.

[30] M.A. Akgun, R.T. Haftka, K.C. Wu, J.L. Walsh, J.H. Garcelon, "Efficient structural optimization for multiple load cases using adjoint sensitivities," AIAA Journal, **39**(3), 511–516, 2001,doi: 10.2514/3.14760

[31] R.T. Haftka, Z. Gürdal, Elements of structural optimization, Springer Science & Business Media, 2012.

[32] R. Hecht-Nielsen, Theory of the backpropagation neural network, Elsevier: 65–93, 1992, doi: 10.1109/ijcnn.1989.118638

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 241-251 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Evaluation Studies of Motion Sickness Visually Induced by Stereoscopic Films

Yasuyuki Matsuura[1,2], Hiroki Takada[2,*]

[1]*Gifu City Women's College, Department of Cross-Cultural Studies, Gifu, 501-0192, Japan*

[2]*University of Fukui, Department of Human and Artificial Intelligent Systems, Graduate School of Engineering, Fukui, 910-8507, Japan*

ARTICLE INFO

Article history:
Received: 21 January, 2021
Accepted: 01 July, 2021
Online: 03 August, 2021

Keywords:
3D (3-dimantional) movie
Head Mounted Display (HMD)
Virtual Reality (VR)
Accommodation
Convergence
Visually Induced Motion Sickness (VIMS)
Stabilometry
Biological signal

ABSTRACT

Humans have experienced motion sickness and possessed the knowledge of stereopsis since classical antiquity. Knowledge of stereopsis dates back to approximately 300 B.C., when Euclid first recognized the concept of depth perception in human vision. Further, the motion sickness is including a sensation of wooziness and nausea that has been documented since approximately 400 B.C., when it was mentioned in the Aphorisms of Hippocrates. Stereoscopic images that utilize binocular stereopsis can frequently cause viewers to experience unpleasant symptoms including visual fatigue. Despite the increased use of three-dimensional (3D) display technologies and numerous studies on 3D vision, there is insufficient accumulation of researches to clarify the effects of 3D images on the human body. Therefore, the safety of viewing virtual 3D images is an important social issue. Inconsistency between convergence and lens accommodation is suspected as a cause of which motion sickness induced by stereoscopic viewing have not yet been identified. A system to simultaneously measure the convergence and lens accommodation is constructed to characterize the 3D vision. Fixation distances were compared between the convergence and lens accommodation while a subject repeatedly viewed 3D video clips. The results indicated that the accommodative power did not correspond to the distance of convergence after 90 s of continuously viewing 3D images. Presently, the relationship between this inconsistency and the unpleasant symptoms remains unclear. Therefore, we introduce empirical research on the motion sickness that can contribute to developments in the relevant fields of science and technology.

1. Introduction

In general, deviation occurs between the images formed on the two bilateral retinas when he/she gazes at a point with both eyes. This deviation is termed binocular parallax due to positional differences in the eyeballs. In humans, it plays an important role in perceiving three-dimensionality. Currently, most three-dimensional (3D) movies and 3D television (3DTVs) use binocular parallax to distribute 2D images to both eyes to achieve stereopsis. Principle of the stereopsis using the abovementioned method is described in a book written by Euclid in approximately 280 B.C. [1, 2]. In the early half of the 19th century, Charles Wheatstone started the stereoscopic photography when he invested the binocular stereoscopic image display method "Stereoscope", which converted a pair of stereo images [3, 4]. In recent years, various 3D video display systems such as mobile devices, free-viewpoint TV, and 3D cinema have been developed. A few recent displays can also present binocular and multi-aspect autostereoscopic images although 3D glasses are generally required. In either case, however, there are the following issues.

(1) unpleasant symptoms including headache, vomiting, and eye strain.

(2) lack of ambience and realism.

Especially in Japanese 3DTVs, dynamic movements cannot be fully expressed since the binocular disparity is set to one degree or less (See section 3.3). Excessive measures against visually-induced motion sickness (VIMS) have been implemented without an appropriate manufacturing standard for stereoscopic video clips (VPs) and their display systems since the eye strain induced by 3D video viewing does not have been still elucidated.

*Corresponding Author: Hiroki Takada, 3-9-1 Bunkyo Fukui, Fukui 910-8507, Japan, +81-776-27-8795 & takada@u-fukui.ac.jp

In 3D video viewing, it is generally understood that the lens is accommodated to the depth of the screen that displays the image, whereas the eyes converge at the position of the 3D object, which is a common idea of eye strain with 3D image viewing. As might be expected, the convergence and lens accommodation are consistent in natural vision. The discrepancy is considered to be cause of the eye strain and the motion sickness induced by stereoscopic viewing [5-6].

According to [7], if the viewing conditions are sufficiently bright, the depth-of-field (DOF) of a target has a mean difference in the order of 1.0 Diopter, and the accommodation-convergence conflict discussed above is a particular problem only in the case of proximity displays such as head-mounted displays (HMDs) and smart glasses [8]. Several methods are available to reduce motion sickness induced by the accommodation-convergence conflict, such as accommodation-invariant near-eye displays [9], light field [10], and rapid adjustment of the focal length to the congestion distance, "variable focus" [11, 12]. The first method displays images as if they are in focus even when the focal length and convergence distance are inconsistent. The second is a technology that generates images close to the visual sensation of the naked eye by taking photos and VPs from multiple viewpoints simultaneously. In addition, optical components in artificial lenses have been developed to change the focus of the human eye intentionally by changing the lens [13]. In addition, as a countermeasure that ignores the accommodation-convergence conflict, a notable method reduces the symptoms of motion sickness by providing the viewer with sound and vibration synchronized with the VPs [14].

Factors associated with the DOF include pupil diameter and resolution. Therefore, image viewing conditions definitely influence the pupil diameter. Most previous studies used high DOF to prevent blurriness, resulting in a measurement environment quite different from everyday conditions.

Moreover, the distribution of the convergence fusional limits in stereoscopic images was obtained in [15], where 84% subjects were able to see a stereoscopic image with a binocular disparity of two degrees. Notably, a single target without a surrounding image was used in the study. Generally, in absence of another parallax image, an accommodation-convergence process that merges double images into a single one functions as a positive feedback system [15].

Deviation between the eyeball positions causes differences in the formed retinal image because of the approximately 6 cm interpupillary distance between the bilateral eyes. The human ability to detect this difference in the retinal images between the bilateral eyes is nearly 10 times more accurate than normal visual acuity. When the deviation is too large, the image information from the bilateral eyes cannot be fused and a double image is formed, making the anteroposterior relationship unclear. A remarkable percentage of the population cannot perceive 3D vision by binocular parallax alone [16–17]. Stereopsis test methods include the T.S.T. (Titmus Stereo Test) and Lang (Lang Stereo Test).

However, it has been reported that there are some influence of stereoscopic viewing on health, which causes unpleasant symptoms, such as visual fatigue, headache, and the vertigo [18-19]. Severity of VIMS is not affected only by construction of the images but also by the viewing environment. It has also been reported that prolonged viewing of stereoscopic displays can cause several health hazards such as severe visual fatigue and headaches [20–22]. On the side notes, it has been reported that the elderly with the mild cognitive impairment (MCI) tends to have a strong interest in stimulation by stereoscopic images [23]. Pregnant women, children, the elderly, and those who consumed alcohol should refrain from stereopsis in a few cases since it is easy for them to be influenced by the health problems associated with stereopsis [24]. However, it is necessary to indicate further hygienic investigations because of little knowledge of the biological effects such as visual fatigue and the VIMS [25].

This report provides an outline of the principle of stereopsis for various displays and the biological effects involved in the stereoscopic vision. Based on the forefront research in this field their clinical significance is also stressed for the description of future prospects.

2. Stereopsis principle and presentation methods

In humans, the two eyeballs are aligned approximately 6 cm apart horizontally. There are always subtly differences between two images formed on each retinas when a person sees an object with the bilateral eyes. Although the image formed on the retina is two-dimensional (2D), the brain reconstructs the information from the bilateral eyes and identifies the condition and the positional relationship of objects occupying the 3D space. Perception of the space and positional relationship is achieved through the eyeballs (lens accommodation and binocular convergence), difference in the eyeball position (binocular parallax and monocular movement parallax), and experience (sizes of objects, perspective, overlapping objects, texture, and shadows). These items are described below.

2.1. Lens accommodation

The mean diameter of the eyeballs in adults is about 22-25 mm and the weight is about 6-8 g. The cornea, anterior chamber, and lens are present in the anterior region and refract light coming into the eye to form images on the retina in the posterior region of the eyeball. However, the eyes cannot simultaneously set the focus on near and distant objects, and the focus has to be adjusted corresponding to the distance of each object. This is termed lens accommodation in the ocular optical system. In humans, the lens is accommodated by changing the lens thickness and curvature, adjusting the focus.

The annular ciliary body is present around the lens. Zinn's zonule radiating from the lens connects to the ciliary body. The ciliary muscle contracts in near vision, and relaxation of Zinn's zonule thickens the lens, shortening the focal distance and forming an image on the retina. In distance vision, the ciliary muscle is relaxed and Zinn's zonule extends the lens radially, which thins the lens and lengthens the focal distance, forming an image on the retina.

The range of lens accommodation is limited, and the limits of near and distance visions are termed near and far points of accommodation, respectively. However, hyperopia or myopia markedly varies among individuals and with aging. The ranges of these represent the accommodation ability, and the accommodation range is wide in young people. Changes in

accommodation ability with aging are mainly due to shifting the near point of accommodation to a distant site because the lens loses elasticity with aging and become unable to readily increase the lens thickness.

2.2. Convergence (binocular convergent movement)

Our vision can perceive the target in detail within a narrow range of only about 1-2°, termed the central fovea. The central fovea contributes to vision in a high-definition field. To see an object, the visual axes of the bilateral eyes have to be set toward the object. The angle formed by these visual axes of the bilateral eyes is the convergence angle, and the distance to the intersection of the visual axes is termed the convergence distance (Figure 1). Convergence (binocular convergent movement) represents horizontal eyeball movement by simultaneous inward rotation of the bilateral eyes to gaze at an object and form an image in the central fovea on the retina, and convergent movement causes 'cross-eyes'[26]. Lens accommodation also has to change corresponding to the distance of the object. The voluntary muscles (ocular muscles) responsible for eyeball movement are roughly divided into the extra- and intra-ocular muscles, and a factor of the extraocular muscle, convergence, and that of the intraocular muscle, accommodation, are useful clues for the visual system to perceive distance [27-28]. Convergent movement occurs in response to the clue of depth direction, and deviation (parallax) of the image on the retina between the bilateral eyes is the typical stimulation. It is an eyeball movement occurring almost simultaneously with lens accommodation to enable 'seeing' in the depth direction in a 3-dimensional space [29]. Therefore, convergence is not conjugated movement of the bilateral eyeballs unlike optokinetic nystagmus which occurs to fix images formed on the retina and vestibulo-ocular reflex which maintains and stabilizes images on the retina even in the presence of body movement [29].

The visual system is capable of identifying the conditions of convergence and accommodation using an efferent copy, which is the command from the brain to the ocular muscles, or proprioceptor signals of the ocular muscles or both information [30]. Regarding the control system model of convergence and accommodation, there is a model in which the extra- and intra-ocular muscle are moved so as to fuse the images on the retina of the bilateral eyes regarding the distances of the target object of convergence and accommodation, as input signals, the focus (blurring) on the retina as 0, and convergence and accommodation as output signals [31-32]. In the control system, crosslinks from the convergence control system to the accommodation control system or vice versa are present, and convergence and accommodation influence each other. Changes in convergence induced by changes in accommodation are termed accommodative convergence, and changes in accommodation induced by changes in convergence are termed convergence accommodation.

2.3. Deviation between the eyeball positions

In binocular stereoscopic images, which are currently mainstream, the viewer makes sense that the displayed object is present from the depth to the front side of the screen utilizing convergent movement and binocular parallax described above to show the image 3-dimensionally (Figure 2). Medial and lateral movements of the eyeballs are termed convergence and divergence, respectively.

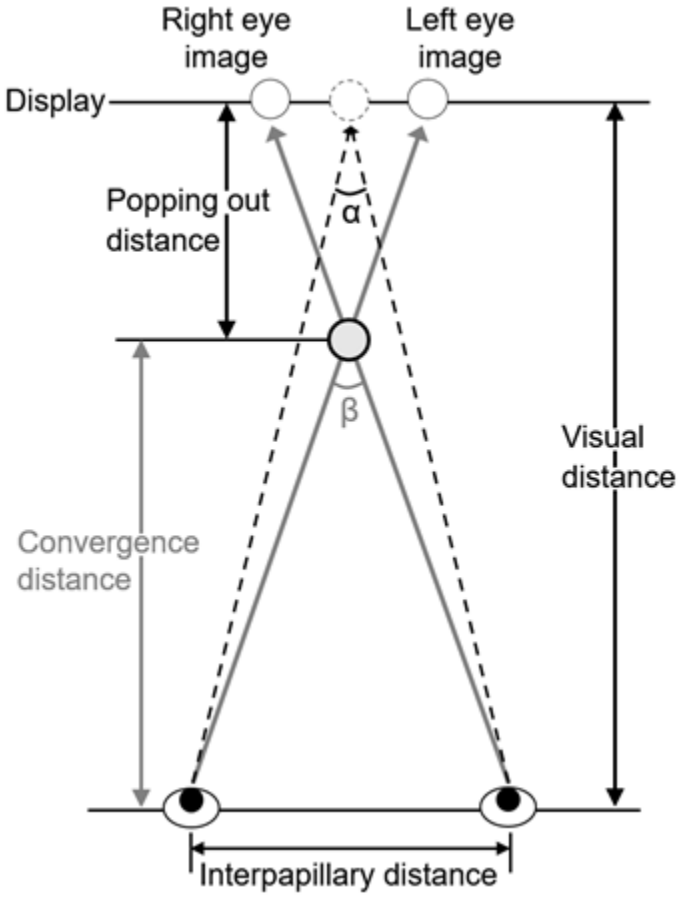


Figure 1: Eye convergence on seeing 3D display. α and β-α are defined as angle of convergence and parallactic angle, respectively.

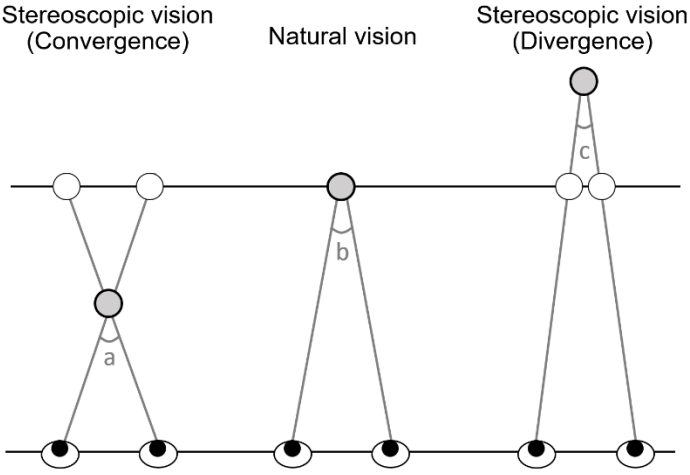


Figure 2: Binocular stereoscopic vision (Convergence and Divergence). a-b: Parallax angle while viewing an object/image in 3D popping out of the screen/display surface, c-b: that sinking in the surface.

Relative positions of the visual field and object change with movements of the object and observer, through which unevenness of the object and the anteroposterior relationship among several objects become perceptible. The direction and speed of visual target movement vary corresponding to the visual range and

interpupillary distance. For example, when looking at the landscape from a car or train, nearby trees and signal masts move backward and distant mountains slowly move backward, but the very far sun and moon appear to move in the same direction as the observer.

2.4. Experience

Humans can perceive the depth and three-dimensionality of space even from a picture written on a flat surface. Various kinds of visual features bring about depth perception (perspective perception), depending on experience and other factors [33-34]. A visual object forming a large retinal image is perceived as present in the front and that forming a small retinal image is perceived as present in the back. When the retinal image size of an object is remembered by experience, the distance to the object is sensed [35]. In contrast, when the size is not known, it is unclear whether the distance is different or the size is different. When objects overlap, the object covering part of the other is perceived as present in front. It may be unclear whether the objects are overlapping or a hole is open based on the presence of overlapping objects alone, but when shadows produced by an illumination light are added, the anteroposterior relationship between objects can be identified.

The presence of shadows also produces three-dimensionality. In daily life space, the shape is convex when the shadow is formed below, and concave when the shadow is formed above. Feelings of unevenness and the anteroposterior relationship may be reversed depending on the shadow position. High and low contrasts induce a feeling of forward and backward movements, respectively. Similarly, light and dark induce a feeling of forward movement and depth of backward movement, respectively.

Regarding gradients of density difference and texture, perspective and inclined planes are perceived corresponding to differences in the density among elements constituting texture and the state of change (gradient). The appearance of the size decreases with an increase in the distance from the observer, and the distance between objects decreases by 1/2 squared. Thus, it is easier to sense depth in the presence of a regular arrangement. This clue is closely related to perspective.

Perspective includes linear and aerial perspectives. In linear perspective, the feeling of depth is produced by composition on the assumption of a viewpoint position and distant vanishing point, and this is applied to perspective [36]. The Last Supper by Leonardo da Vinci is a typical example drawn by perspective. This work employed a one-point perspective setting the vanishing point in almost the center of the screen, which concentrates the line of sight to the theme and produces a sensation that a space extends in the back of the screen. In aerial perspective, near visual objects are clearly perceived while distant visual objects are dimly perceived, causing a feeling of depth. In real life, distant objects are blurred due to light scattering by air, giving a sense of perspective. When the size and shape are the same, clearly seen objects are felt near, and blurred objects are felt to be present at the back [37-38].

2.5. Stereoscopic image display technique

Stereoscopic image display methods include eye glass-type and naked-eye binocular stereoscopic display systems, multiview and depth information presentation systems, and wavefront reconstruction and space image systems. Of these, the binocular stereoscopic display system utilizes binocular parallax, in which 2 images with parallax in the horizontal direction are individually presented to the bilateral eyes, and it is generally used as a stereoscopic image display method [3, 39-40].

The image presentation methods employing the binocular stereoscopic display system include the side-by-side method in which the bilateral images are horizontally arranged, top-and-bottom method in which the bilateral images are vertically arranged (images for the left and right eyes are set at top and bottom, respectively), the line-by-line method in which the bilateral images are arranged alternately on each line horizontally or vertically, and Power 3D method [41-42].

3. Biological effects of image viewing

Regarding the biological effects of stereoscopic viewing, adding camera shake-like vibration to the entire image and dynamic changes that involve computer graphic images to induce a feeling of being present at the place are likely to cause VIMS. Stimulation with stereoscopic images induces 3D sickness whose symptoms are similar to motion sickness. This is not limited to stereoscopic images, and viewing images and rapidly moving screen that requires blinking may cause headache, vomiting, and vertigo. These and other similar events can be broadly defined as VIMS.

3.1. Biological effects of motion sickness induced by stereoscopic viewing

Motion sickness is including a sensation of wooziness and nausea known since approximately 400 B.C. In [43] , the author described that “When Hellebore has been taken, let the body be general kept in motion, enjoying less rest, and less sleep. For even sailing proves that motion disturbs the functions of the body.” In [44], it is written that “they felt sick due to vehicle sickness in a Japanese oxcart (Figure 3) and all appeared inverted to their blinking eyes.” in the first half of the 12th century. Currently, besides vehicle motion sickness and VIMS, space sickness has also been reported [45-46]. As the space motion sickness symptoms develop in zero gravity, vomiting in a space suit helmet during extravehicular activity may cause suffocation. People have suffered the unpleasant symptom of motion sickness for long.

From the viewpoint of preventive medicine, it is important to accumulate the basic studies on stereoscopic viewing because stereoscopic viewing involves both positive [47] and negative aspects. Contrary to what you might think, there are a few reports on the former. Firstly, cases demonstrating the effect of antisuppression exercise of intermittent exotropia and the pleoptics have been reported by using stereoscopic image techniques [48]. In [47], the authors reported that “the accommodation training using 3D movie had temporarily improved visual acuity and seemed to lead to a decrease in asthenopia in their experiment.”

Unpleasant symptoms such as visual fatigue, vomiting, and headache are often caused by stereoscopic images utilizing binocular stereopsis, which depends on the viewing conditions [49]. Integrating several sensations, such as those from the skin and somatic sources, the body perceives space. Since the space perception excessively depends on the visual function, visual sensation among others carries the major burden while viewing stereoscopic ones. In most cases after viewing those images, the

symptoms associated with the motion sickness disappear when you stop watching, however, it may last almost a day in severe cases of the VIMS, which is not caused only by images but also by simulators. Sensations other than the visual one can be given by the simulators in which there is deviation between their sensations and motions included in the video film. In the simulator sickness, the motion sickness is also amplified by flickering in the screen. Furthermore, it has been reported about ataxia in the simulator sickness. and the US navy prohibits the persons with experience of the simulator sickness from boarding within 24 hours after simulator operation [50]. Thus, these kinds of the knowledge may give a key consideration of the motion sickness including the car sickness.

Figure 3: Japanese oxcart

The VIMS is influenced by auditory [51], visual [52-53], olfactory [54], deep [55], and other sensations. The followings are case studies. A worker operating the remote control of a large power shovel has developed severe ophthalmalgia and headache one month after starting work owing to stereoscopic image viewing, and his quality of life subsequently deteriorated [56]. In addition, there is another report that a boy has developed acute internal strabismus due to 3D movie viewing [44].

Early research considered the overstimulation theory for explaining the mechanism behind the onset of motion sickness. According to this theory, the acceleration of a vehicle causes the overstimulation of the visceral and vestibular organs, leading to an excitement of the hypothalamus, which induces the vestibulo-vegetative reflex, causing various symptoms of the motion sickness. Instead of this overstimulation theory, it is necessary to develop some mechanism of the motion sickness because the motion sickness was found even in microgravity environments. According to the sensory conflict theory [18, 57], actual sensory information such as visual, vestibular, and somatosensory one is compared with that of the episodic memory in the central nervous system. The motion sickness would be induced only if a sensory information combination were different from that expected from the memory [58]. The vestibular stimulation is transmitted to the vomiting center in the medulla oblongata via the vestibulo-vegetative system. The vestibular and autonomous nerve systems are anatomically and electrophysiologically closely connected, strongly suggesting their relationship with the unpleasant symptoms of motion sickness [59]. When a motion sickness-inducing rotational load is provided to rats, it increases the histamine level in the brainstem and hypothalamus, which is related to vomiting during the motion sickness [60].

Severity of the motion sickness can be quantitatively evaluated in accordance with the analysis of the body sway, which is regarded as an output of the equilibrium system. In general, it is difficult to obtain significant difference from statokinesigrams and/or their area of sway, sway values, total length, and total locus length per unit area, with their eyes open because the visual information helps subjects to keep upright posture. In recent decades, numerical analysis of the mathematical model of the body sway shows the possibility to find significant differences with eyes open (See section 3.3) [61].

In case of movies and 3DTV utilizing systems to display binocular stereoscopic images, the biological effect of stereoscopic viewing cannot be ignored. Figure 1 shows binocular parallax determined by the positional relationship between bilateral eyes and the object in stereoscopic images, which depends on the distance between the centers of your two eyes (interpupillary distance: PD). Using binocular parallax, stereopsis is realized by individually distributing 2D images to observers' left and right eyes. It also depends on the viewing conditions (viewing position, darkness and light in the room, and physical conditions). Therefore, certain biological effects occurring during stereoscopic viewing can be attributed to individual differences in viewing conditions, interpupillary distance, and visual function of the viewers. Studies on these differences have been reported [56, 62].

3.2. Causes of motion sickness and visual fatigue by stereoscopic images

According to the investigation of the relationship between individual differences in visual function and parallax range of stereoscopic images, persons with high accommodation convergence have a narrow range of comfortable visual field, where he/she is prevented from suffering discomfort of popping out stereoscopic images [63]. A positive correlation has been found between the grade of phoria and subjective evaluation for degree of stereoscopic viewing-induced fatigue [64]. As the side notes, there is less visual fatigue in persons with slight exophoria while viewing stereoscopic images [65]. Hence, considering individual differences in visual function is necessary while discussing the biological effects of stereoscopic viewing.

Theories of "inconsistency between convergence and accommodation" and the "influence of excess parallax" have been suggested in order to explain other causes for stereoscopic viewing-induced fatigue besides differences in the individual visual function. The former states that the visual fatigue would be caused by inconsistency between the accommodation and convergence distances, which also increases the accommodation load. The later states that the binocular parallax increased to emphasize the stereoscopic effects.

The following is frequently described as an effect of the binocular stereoscopic display system upon our body; accommodation and convergence contradict each other during stereoscopic viewing because we just focus upon the surface of the display and simultaneously adjust the convergence to the stereoscopic object popping out from it [66, 67]. Most of

researchers believe that there is deviation between the accommodation and the convergence during stereoscopic vision, however, the authors reported that the accommodation was not fixed to the surface of the display during stereoscopic vision [68].

The range in which the anteroposterior regions of the focused region appear to be in focus is termed DOF, which is generally approximately ±0.2–0.3 Diopter in humans [69, 70]. We would not feel inconsistency between the accommodation and the convergence when a stereoscopic object is presented within the DOF [71]. This also resolved the inconsistency problem during stereoscopic viewing, inhibiting the unpleasant feeling observed in a few studies [31, 72], whereas fatigue was caused in presence of a temporary change of binocular parallax, although the change was within the DOF in another study [73, 74].

A theory suggests that because the lens accommodation and convergence are simultaneously changing in opposite directions, their functions become unstable, and efforts to fuse the images created by the bilateral eyes are assumed to be the cause of fatigue [75–77]. In addition, we are conducting preliminary experiments on the illuminance of the experimental environment and interpreting the experimental data so far. The pupil diameter is affected when the environmental illuminance changes. When the illuminance is lowered, the pupil diameter becomes larger, and the DOF becomes shallower. We considered that it is easy to be induced the motion sickness because there is a discrepancy between the lens accommodation and the surface of the display, which is not included in the range of the DOF [78].

3.3. The current trend of motion sickness and visual fatigue by stereoscopic images

Because both the convergence and lens accommodation fit the objects in natural stereopsis, they do not contradict each other. The following hypothesis as described in section 3.2 seems to have captured the hearts of many 3D image engineers and researchers, i.e. it is often considered that their inconsistency is the cause of the VIMS. In 2011, 3D Consortium (3DC) formulated safety guidelines in which the range of comfortable parallax angle for stereoscopic images was specified as ±1.0° [79–81]. This is also based on the hypothesis and studies serving as the basis for this comfortable parallax range reported by [79] and [80]. Setting the change in binocular parallax within approximately 1°, approximately 87% subjects could fuse and observe stereoscopic images under the experimental conditions of their studies [79–81]. After these reports were published, however, this finding about the fusion limit was refuted as follows.

It was reported that stereoscopic images at a parallax angle up to approximately 2°could be fused and observed by approximately 84% subjects [15]. It was concluded that the difference in the binocular parallax between [15] and [80] was due to premature processing deviating from elementary statistics. In recent decades, it has been discussed it is appropriate to revise the comfortable parallax range for stereoscopic images to 2° [82]. As observed during natural viewing, the convergence and lens accommodation in the young were synchronized in accordance with their observation. No inconsistency was observed while viewing stereoscopic video clips under medium illuminance [83].

It is important for researchers in this field to set the visual environment, which can affect the results of the experiment. Thus, the experimental results of [83] cannot be simply compared with those of previous studies because the environment of [83] is different from those of experiments under low illuminance in the previous studies. In addition, the stereoscopic images were drawn by using “power 3D method” [41]. In [25], the author states that “no study has clearly shown the association although it is generally assumed that long-term stereoscopic viewing leads to unpleasant feelings including the visual fatigue.” Several points presently remain unclear with regard to the accommodation-convergence mismatch during stereoscopic viewing and the discomforts.

4. Mathematical models of the body sway to evaluate the motion sickness

Severity of motion sickness is measured by stabilometry in accordance with the consideration of equilibrium function [78, 84]. In stabilometry, recording of stabilograms for 60 seconds begins when the standing posture stabilizes. Statokinesigrams are composed of each component of the stabilograms. Indices such as area of sway, total locus length, and total locus length per unit area are classically estimated to analyze statokinesigrams [84-85]. The latter is known as a parameter of the fine control of standing posture by the proprioceptive reflexes [86]. In addition to the abovementioned indices, an index termed sparse density was proposed in consideration of the non-linearity of the system to control upright position [87-88]. The following nonlinear system

$$\frac{\partial x}{dt} = -\frac{\partial}{\partial x}U_x(x) + \mu_x w_x(t), \tag{1}$$

$$\frac{\partial y}{dt} = -\frac{\partial}{\partial y}U_y(y) + \mu_y w_y(t) \tag{2}$$

has been proposed for the description of the body sway where $w_x(t)$ and $U_x(x)$ represent the white noise and the time-averaged potential function (TAPF) in the lateral component, respectively. In general, the first term on the right-hand side is regarded as a linear function for each component [89-91], i.e. the body sway has been described as a Brownian motion [92-93]. Based on [94] and [95], the lateral component of the SDE (1.1) is assumed to be independent of the anterior-posterior component (1.2). Also, we did not obtain remarkable significance in the cross correlation between those components from the stabilograms measured in our experiments [87].

5. Mathematical approach for the evaluation of body balance function

5.1. Improved deductive theory

It is difficult to describe the abnormality in the body sway during the alcoholic load [87-88, 96] or the motion sickness. Prolonged exposure to a stereoscopic video clip, the mathematical model (1) has been investigated [97], but interaction between anteroposterior and lateral components cannot be neglected, and a new mathematical model is being investigated.

Non-linear stochastic differential equations (SDEs) (1) were obtained from our deductive theory [61, 78, 84]. The Markov process without abnormal diffusion is required by the randomness in the body sway, which is based on our observation. In most cases,

$\mu_x \neq \mu_y \neq 1$ as to compare variations with the others in polygraphs that are measured and recorded in the experiments of the electrophysiology [61, 78, 84]. The TAPFs can be estimated as

$$U_x(x) = -\frac{\mu_x^2}{2}\ln G_x(x) + const., \quad (3)$$

$$U_y(y) = -\frac{\mu_y^2}{2}\ln G_y(y) + const., \quad (4)$$

where $G_x(x)$, $G_y(y)$, are expressed as distributions for each direction. Several minimum values of the TAPFs are often obtained from the stabilometry. In numerical simulations, local stability is also seen as motions with high-frequency near the minimal potential surface, where a high density of representation points is expected be generated by the SDEs. Inversely, degree of the local stability cannot be measured only by the total locus length per unit area, but also by the sparse density including more local information in the measurement [87].

5.2. Application of our deductive theory

The mathematical method in this chapter has already been applied to quantitate severity of the motion sickness induced by stereoscopic viewing and the blur of liquid crystal [94, 97-99]. Especially in [94], it has been discussed that peripheral vision contributed to an increase in the sway value with eyes closed after the exposure to a stereoscopic video clip. However, upright posture is stable with eyes open, in fact, the sway value is so small that the instability has been able to be evaluated while viewing video clips. As mentioned above, our deductive theory has been recently improved to enable comparison of variations among independent components. We have also succeeded in enhancing the accuracy of the evaluation during the exposure to stereoscopic video clips. In addition to the skewness, kurtosis, and standard deviation of the probability density distribution of the observed data, the translation error in the nonlinear analysis was herein used as an evaluation index for the numerical analysis of SDEs [61]. As a result, we constructed a new theoretical system to obtain the SDEs describing the equilibrium system from the measurement data of each subject.

6. Evaluation of stereoscopic image-induced motion sickness

In addition to the physiological methods involving autonomic nerve activity, subjective psychological methods have been well developed to measure the influence of the VIMS on the body. Simulator Sickness Questionnaire (SSQ) is a best-known measurement to assess the VIMS including the simulator sickness, which comprises 16 effective subjective items extracted from 1,119 paired data on the Motion Sickness Questionnaire (MSQ) measured before and after experiencing a simulator by using factor analysis [100]. The VIMS is also assessed using physiological measurements such as body sway, blood pressure, respiratory rate, electrocardiography, electrogastrography, perspiration, resistance value of the skin, and number of eye blinks [101–104]. In a study using the SSQ score, it was also reported by the group complaining of vibration load-induced motion sickness that the difference in stance width during upright makes a difference in the incidence of motion sickness [105].

6.1. Effect of background vision on the equilibrium system

In previous studies, compared with visual pursuit, higher sway values including, the area of sway, total locus length, total locus length per unit Area, and sparse density during the peripheral viewing of 3D images were observed [94]. Especially in case of backgrounds, the appearance of actual space that humans perceive and that of 3D VPs is different, which is considered a reason for the influence of peripheral vision on the equilibrium system. In this section, we verify whether 3D VPs viewed without backgrounds influences the equilibrium system, and develop a mathematical model.

The body sway was measured during 1 min of video viewing, and thereafter, 3 min of standing with eyes closed after the pre-rest. Before and after this stabilometry examination, we performed a subjective evaluation of motion sickness symptoms using the SSQ. The smart glass MOVERIO BT-200 (EPSON, Nagano) was used to view the VPs used in the experiment of [106]. This device facilitates augmented reality (AR); however, to remove any external stimuli in the experiment other than those provided by the videos, they were projected on a black screen for measurement. In the video, spheres were fixed at four corners while another sphere moved through the screen in a complex manner.

We performed a two-way analysis of variance (ANOVA) that uses the persistence of the influence of exposure to VPs as a factor for each analytical index calculated from a statokinesigram. In the ANOVA results, several non-interacting main effects were observed for each pair of factors (solidity/backgrounds). According to the statistical analysis of the total locus length per unit area, there were main effects (1) on solidity when viewing VPs with backgrounds, and (2) of backgrounds when viewing 2D VPs. In this connection, backgrounds exerted a main effect when viewing 3D VPs in accordance with the ANOVA for sparse density S3. In addition, there was a main effect on solidity when viewing videos without backgrounds.

The equilibrium system was affected 1–2 min after viewing the 3D VPs with backgrounds. In addition, the SSQ result indicates that motion sickness may be caused by viewing 3D VPs with backgrounds. In addition, the sway values in the control were compared with those obtained after viewing 3D VPs without backgrounds (with their eyes closed). The area of sway and S3, both measured 2–3 min after the viewing, were significantly larger than those in the control. At that time, the total locus length per unit area was significantly smaller than that in the control. Subjects were allowed to use their peripheral vision; however, it was easy to focus on the central sphere, which was the same as the pursuit viewing of images because subjects viewed the VPs without backgrounds. Therefore, the influence reduction was observed 1–2 min after the 3D viewing. Moreover, 3 min after viewing, the instability of the system may increase owing to physical fatigue from maintaining an upright posture, leading to increased body sway.

Table 1: Recent researches

Authors	M. Malińska et al [89]	A. M. Baranowski et al [90]	T. H. Cho et al [91]	Y. Sawada et al [14]
Year	2015	2016	2017	2020
Apparatus	3D: Screen with Shutter glasses HMD: HMD with gloves	Screen with 3D Shutter glasses	3DTV with Polarization glasses	HMD when sitting on a chassis of a scooter
Experiments	3D: Watching part of 'Avatar' HMD: Training in handling on the virtual workstation	3 genres (horror, action, and documentary) with three between-subjects viewing conditions (director's 3D, artificial 3D, and 2D)	3D and 2D films	3D video with Sound and/or vibration riding simulator motorcycle
Measuring (Objective)	Electrocardiogram		Ocular parameters (Accommodation, Convergence, Stereo-acuity, Tear break-up time)	
Measuring (Subjective)		Fast Motion Sickness scale (FMS) and Self-Assessment Manikin scale (SAM)		FMS
Evaluation methods	Statistical Analysis	Statistical Analysis	Statistical Analysis	Statistical Analysis

6.2. Effects of duration on the equilibrium system

In previous studies, it was shown that viewing 3D VPs affects body sway; however, there was no comment on the duration of viewing VPs. In this section, we examine the effects of duration on the body sway and introduce a mathematical model that describes the equilibrium system [107]. In addition, we succeeded in finding the temporal fluctuation in the mathematical model [108].

The experimental protocol of [107] used two patterns of measurements: following a standing pre-rest, 1 min with eyes open and 3 min with eyes closed; 2 min with eyes open and 3 min with eyes closed. A 3D VP with binocular disparity and a 2D VP for uniocular viewing were displayed on the 3D display KDL 40HX80R (SONY, Tokyo) installed 2 m from the subjects. The experiment considered the order effect, and we conducted the measurements for 2D and 3D VPs in random order. In addition, measurements for other durations were also performed on different days. Excluding the total locus length per unit area, we observed that the increase in sway values depended on the duration of the exposure to 2D VPs. In addition, sway values 1–2 min after viewing a 3D VP for 2 min were significantly greater than those while viewing it. Therefore, the influence of the viewing on the equilibrium system was seen even after the exposure to VPs. Regardless of the solidity of the VPs, the area of sway and the total locus length were significantly greater at 2–3 min after viewing than when viewing VP sway values. Similar results were observed in the control experiment. Therefore, in the experiment [107], an increase in the sway values at 2–3 min after viewing may not be caused by VIMS, but by fatigue in maintaining the upright posture. Hence, duration has an effect on the equilibrium function after viewing the VPs, and viewing the 3D VPs for 2 min continues to influence the equilibrium function for at least 2 min after viewing.

7. Problems and future prospects

Table 1 lists the recent research on this topic [14, 89-91]. The progression of TV technology to high image quality has facilitated the sales of naked-eye 3D displays for medical use and high-definition glassless 3DTV and enhancement in the image quality of 3D images. Furthermore, recently there has been a rapid progression in weight reduction and enhancement in the performance of eyeglass-type wearable devices [109]. A small projector is attached to the frames of eye glasses and sun glasses, and images are projected either on the inner side of the lens, or projected in front of the eyes using a semitransparent binocular HMD with an integrated lens and projector or non- and semitransparent monocular displays. These new image display methods have already appeared, increasing the opportunities for stereoscopic viewing in various fields, for not only amusement but also medical care and industrial use.

The characteristic of our study introduced in section 4, 5 and 6 is that it not only aggregated experimental studies on the influence of stereoscopic viewing on visual function, but also helps establish scientific techniques to quantify motion sickness. Previous studies discussed inconsistency between convergence and accommodation without simultaneously measuring them; however, we performed stabilometry, simultaneous measurement of convergence and accommodation, and evaluated body balance and visual functions for a basic investigation of stereoscopic viewing-induced motion sickness in the experimental study. By increasing the number of subjects, performing close investigation

with autonomic nerve evaluation using electrocardiography and electrogastrography, and evaluating the body balance function using body sway in a seated posture, this study facilitates safe stereoscopic viewing with less occurrence of VIMS.

Virtual Reality (VR) sickness can be caused by a visual illusion called vection [110-111] and movement on the screen while viewing 3D images. According to the sensory conflict theory, VR sickness can also be induced when passive movement creates a mismatch between information related to orientation and movement supplied by the visual and vestibular systems. This mismatch induces feelings of nausea. In particular, vection is easily caused in HMDs and on large-sized high-definition 4K/8K displays. In addition, understanding of changes in the bio-signals during the vection helps us to confirm the previous studies [112]. It might improve our knowledge in the concept the motion sickness [112]. Therefore, it is important to examine the influence of vection on human bodies in detail.

8. Conclusion

This report provides an outline of the principle of stereopsis for various displays and the biological effects involved in the stereoscopic vision. Based on the forefront research in this field their clinical significance is also stressed for the description of future prospects. The spread of 3D and 4K/8K TV cannot progress unless the safety of stereoscopic images is secured, resulting in elimination from market competition. Therefore, prevention and alleviation of motion sickness stereoscopic viewing and providing basic documents to establish the safety criteria of 3D not only secure the extensive application of stereoscopic images and safety and relief of viewers but also contribute to technological development in Japan.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research has been supported in part by JSPS KAKENHI Grant Number JP17K00715, JP18K11417, and JP26870490.

References

[1] H. E. Burton, “The optics of Euclid,” JOSA. **35**(5), 357-372, 1945.
[2] I. P. Howard, B. J. Rogers, Binocular Vision and Stereopsis. New York: Oxford University Press, 1995.
[3] W. Charles, “Contributions to the Physiology of Vision. Part the First. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision,” Philosophical Transactions of the Royal Society of London, **128**, 371-394, 1838.
[4] B. David, The stereoscope; its history, theory, and construction, with its Application to the fine and useful arts and to education. London: John Murray, 1856.
[5] F. M. Toates, F. M. “Vergence eye movements,” Documenta Ophthalmologica, **37**(1), 153–214, 1974.
[6] H. G. Hoffman, D. R. Patterson, E. Seibel, M. Soltani, L. Jewett-Leahy, S. R. Sharar, “Virtual reality pain control during burn wound debridement in the hydrotank,” Clin J Pain, **24**, 299-304, 2008.
[7] R. Patterson, “Human factors of stereo displays: An update,” Journal of the Society for Information Display, **17**, 987-996, 2009.
[8] S. Nagata, “The binocular fusion of human vision on stereoscopic displays - field of view and environment effects,” Ergonomics. **39**(11), 1273-1284, 1996.
[9] R. Konrad, N. Padmanaban, K. Molner, E. A. Cooper, G. Wetzstein, “Accommodation-invariant Computational Near-eye Displays. ACM Transactions on Graphics,” **36**(4), doi:10.1145/3072959.3073594, 2017.
[10] M. Broxton, J. Flynn, R. Overbeck, D. Erickson, P. Hedman, M. DuVall, J. Dourgarian, J. Busch, M. Whalen, P. Debevec, “Immersive light field video with a layered mesh representation,” ACM Transactions on Graphics, **39**(4), doi:10.1145/3386569.3392485, 2020.
[11] D. Dunn, C. Tippets, K. Torell, P. Kellnhofer, K. Akşit, P. Didyk, K. Myszkowski, D. Luebke, H. Fuchs, Wide Field Of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors,” IEEE Transactions on Visualization and Computer Graphics, 23(4), 1322-1331, 2017.
[12] K. Akşit, W. Lopes, J. Kim, P. Shirley, D. Patrick, “Near-eye varifocal augmented reality display using see-through screens,” ACM Transactions on Graphics, **36**(6), DOI:10.1145/3130800.3130892, 2017.
[13] W. Cui, L. Gao, “Optical mapping near-eye three-dimensional display with correct focus cues,” Optics Letters, **42**(13), 2475-2478, 2017.
[14] Y. Sawada, Y. Itaguchi, M. Hayashi, K. Aigou, T. Miyagi, M. Miki, T. Kimura, M. Miyazaki, “Effects of synchronised engine sound and vibration presentation on visually induced motion sickness,” Scientific Reports. **10**, DOI:10.1038/s41598-020-64302-y, 2020.
[15] S. Nagata, “Distributions of "Vergence Fusional Stereoscopic Limit (VFSL)" of Disparity in Stereoscopic Display,” Transactions of the Virtual Reality Society of Japan, **7**(2), 239-246, 2002.
[16] W. Richards, “Stereopsis and stereoblindness,” Experimental B rain Research, **10**, 380-388, 1970.
[17] W. Richards, “Anomalous stereoscopic depth perception,” JOSA, **61**, 410-419, l971.
[18] J. T. Reason, J. J. Brand, Motion Sickness, London: Academic Press Inc, 1975.
[19] C. D. Balaban, J. D. Porter, “Neuroanatomic substrates for vestibulo-autonomic interactions,” J Vestib Res, **8**, 7–16, 1998.
[20] K. Murata, S. Araki, K. Yokoyama, K. Yamashita, Y. Okumatsu, S. Sakoh, “Accumulation of VDT Work-Related Visual Fatigue Assessed by Visual Evoked Potential, near Point Distance and Critical Flicker Fusion,” Ind Health. **34**, 61-69, 1996.
[21] G. Hamagishi, “Ergonomics for 3D Displays and Their Standardization,” The Technical Report of the Institute of Image Information and Television Engineers. **33**(16), 9-12, 2009.
[22] K. Nishimura, T. Iwata, K. Murata, “Effects of 3-dimensional video games on visual nervous function,” Akita J Med. **37**, 85-91, 2010.
[23] T. Shibata, T. Kawai, K. Noro, A. Arimoto, T. Ohshima, S. Matsuoka, “A Study on the welfare application of stereoscopic 3D images for dementia patients,” Ergonomics. **36**(Suppl), 390-391, 2000.
[24] Samsung, Viewing TV using the 3D function http://www.samsung.com/ca/pdf/3D-tv-warning_en.pdf Accessed 30 November 2017.
[25] Ministry of Internal Affairs and Communications of Japan, “Final Report of Investigation Committee for 3D Television,” 2012.
[26] M. Ikeda, What are your eyes looking at? Tokyo: Heibonsha, 1988.
[27] S. K. Fisher, K. J. Ciuffreda, "Accomodation and Apparent Distance," Perception, **17**, 609-621, 1988.
[28] M. Mon-Williams, J. R. Tresilian, "Some Recent Study on the Extrarentinal Contribution to Distance perception," Perception, **28**(2), 167-181, 1999.
[29] K. Uchikawa, K. Shinomori, Visual I: The faculty of sight and Initial function Tokyo: Asakurashoten; 2007.
[30] H. M. Burian, G. K. Von Noorden, Binocular Vision and Ocular Motility: Theory and Management of Strabismus St Louis: Mosby; 1980.
[31] N. Hiruma, T. Fukuda, “Viewing Conditions for Binocular Stereoscopic Images base on Accommodation Response,” The IEICE transactions on information and systems D-2, **73**(12), 2047-2054, 1990.
[32] C. M. Schor, "A Dyanmic Model of Cross-Coupling between Accomodation and Convergence: Simulations of Step and Freqency Responses," Optometry and Vision Science, **69**, 258-269, 1992.
[33] H. Wallach, D. N. O'Connell, “The kinetic depth effect," Journal of Experimental Psychology,” **45**(4), 205–217, 1953.
[34] S. A. Linkenauger, M. Leyrer, H. H. Bülthoff, B. J. Mohler, "Welcome to wonderland: The influence of the size and shape of a virtual hand on the perceived size and shape of virtual objects,” PloS one, **8**(7), doi: 10.1371/journal.pone.0068594, 2013.
[35] G. Mother, Foundations of Perception, London: Taylor & Francis, 2006.
[36] N. D. Cook, Harmony, Perspective and Triadic Cognition. New York: Cambridge University Press, 2011.
[37] S. Nagata, “Visual Sensitivities to Cues for Depth Perception,” Journal of The Institute of Television Engineers of Japan, **31**(8), 649-655, 1977.
[38] J. E. Cutting, P. M. Vishton, “Perceiving Layout and Knowing Distances: The integration, Relative Potency, and Contextual Use of Different

Information About Depth, “Perception of Space and Motion, **22**(5), 69-117, 1995.
[39] Y. Takaki, “Basics of Three-dimensional Displays,” The Journal of The Institute of Image Information and Television Engineers, 67(11), 966-971, 2013.
[40] S. Aukstakalnis, Practical Augmented Reality: A Guide to the Technologies, Applications, and Human Factors for AR and VR, Boston: Addison-Wesley Professional, 2016.
[41] T. Fuji, M. Kosaka, T. Komuro, A, Shimotomai, “PC 3D Viewer Kit,” Journal of The Society of Photographic Science and Technology of Japan, **72**(4), 261-265, 2009.
[42] N. Fujiyoshi, M. Nagasawa, “Dawn of the Era of Three-dimensional Images,” Mitsubishi Denki giho, **85**(3), 2-6, 2011.
[43] Hippocrates, Coar T. The Aphorisms of Hippocrates with a translation into Latin, and English, London: Classics of Medicine Library, 1982.
[44] J. Ikegami, Konjaku Monogatari-Shu Honshobu (Last Part) Tokyo: Iwanamishoten, 2001.
[45] A. Graybiel, C. D. Wood, E. F. Miller, D. B. Cramer, “Diagnostic Criteria for Grading the Severity of Acute Motion Sickness,” Aerospace Med, **39**(5), 453-455, 1968.
[46] P. S. Cowings, K. H. Naifeh, W. B. Toscane, “The Stability of Individual Patterns of Autonomic Responses to Motion Sickness Stimulation,” Aviat Space and Environ Med, **61**, 399-405, 1990.
[47] A. Sugiura, M. Miyao, T. Yamamoto, H. Takada, “Effect of strategic accommodation training by wide stereoscopic movie presentation on myopic young people of visual acuity and asthenopia,” Displays, **32**(4), 219-224, 2011.
[48] T. Handa, “The present condition of three-dimensional films and utilization of three-dimensional technology in visual function tests and orthoptics procedures,” Journal of Japanese Association of Certified Orthoptist, **41**, 45-52, 2012.
[49] H. Ujike, “Report of ISO International Workshop on Human Safety on Image,” VISION. 17(2), 143-145, 2005.
[50] US Navy. OPNAVIST 3710.7T, 2004.
[51] S. Takane, Y. Suzuki, T. Sone, H-Y. Kim, “A study on control of distance perception by simulation of HRTF,” Proceedings of the Virtual Reality Society of Japan Annual Conference, **1**, 55-58, 1996.
[52] T. Inoue, “Eye Movement and Accommodation when Viewing 2D and 3D Images,” The Journal of the Institute of Television Engineers of Japan, **50-4**, 423-428, 1996.
[53] R. S. Kennedy, K. S. Berbaum, W. P. Dunlap, L. J. Hettinger, “Developing Automated Methods to Quantify the Visual Stimulus for Cybersickness,” Hum Factors Ergon Soc Annu Meet, 40-2,1126-1130, 1996.
[54] M. Ohsuga, T. Tatsuno, F. Shimono, K. Hirasawa, H. Oyama, H. Okamura, “Bedside Wellness - Development of a Virtual Forest Rehabilitation System,” Stud Health Technol Inform, **50**, 168-174, 1998.
[55] E. M. Kolasinski, Simulator sickness in virtual environments (ARI Technical Report 1027). Alexandria: U.S. Army Research Institute for the Behavioral and Social Sciences, 1995.
[56] H. Uchida, K. Hiwatashi, “Individual Difference of Stereoscopic Vision,” Proceedings of the ITE Annual Convention, **29**, 115-116, 1993.
[57] C. D. Balaban, J. D. Porter, “Neuroanatomic substrates for vestibulo−autonomic interactions,” J Vestibular Research, **8**, 7-16, 1998.
[58] K. Hirayanagi, “A present state and perspective of studies on motion sickness,” The Japanese Journal of Ergonomics, **42**(3), 200-211, 2006.
[59] N. H. Barmack, “Central vestibular system: vestibular nuclei and posterior cerebellum,” Brain Research Bulletin, **60**, 511-541, 2003.
[60] N. Takeda, M. Morita, T. Kubo, A. Yamatodani, T. Watanabe, H. Wada, T. Matsunaga, “Histaminergic Mechanism of Motion Sickness Neurochemical and Neuropharmacological Studies in Rats,” Acta Otolaryngologia, **101**, 416-421, 1986.
[61] F. Kinoshita, H. Takada, “Numerical analysis of SDEs as a model for body sway while viewing 3D video clips,” Mechatronic Systems and Control, **47**(2), 98-105, 2019.
[62] M. Sato, “Individual Differences in Stereopsis,” Journal of Japanese Society Opthaimpligical Optics, **35**(2), 33-37, 2014.
[63] H. Mizushina, H. Ando, “The Relationship between Disparity Range for Comfortable Viewing of Stereoscopic Images and Individual Differences in Visual Function,” The Technical Report of the Institute of Image Information and Television Engineers, **38**(11), 23-26, 2014.
[64] T. Shibata, J. Kim, D. M. Hoffman, M. S. Banks, “The Zone of Comfort: Predicting Visual Discomfort with Stereo Displays,” J Vis, **11**(8), 1-29, 2011.
[65] S. Kubota, K. Kudo, M. Takemoto, A. Shimada, Y. Nakamura, “Affect of Visual Acuity and Accommodation Speed on Visual Fatigue During Movie Viewing on 3D Television,” The Journal of the Institute of Image Information and Television Engineers, **67**(7), J262-269, 2013.
[66] F. M. Toates, “Vergence eye movements,” Doc Ophthal, **37**, 153-214, 1974.
[67] Ultra-Realistic Communications Forum, Ultra-experience Design and Evaluation Section, Working Group of Evaluation of 3D Images, “Report of Evaluation about Visual Fatigue of Stereoscopic Video Revised Edition,” 2013.
[68] M. Miyao, S. Ishihara, S. Saito, T. Kondo, H. Sakakibara, H. Toyoshima, “Visual accommodation and subject performance during a stereographic object task using liquid crystal shutters,” Ergonomics, **39**(11), 1294-1309, 1996.
[69] F. W. Cambell, “The Depth of Field of the Human Eye,” Journal of Modern Optics, **29**, 157-164, 1957.
[70] W. N. Charman, H. Whitefoot, “Pupil Diameter and Depth-of-Field of Human Eye as Measured by Laser Speckle,” Optica Acta, **24**, 1211-1216, 1977.
[71] T. Kawai, H. Morikawa, K. Ohta, N. Abe, Basic Principles and Production Technology of Stereoscopic Images. Tokyo: Ohmsha, 2010.
[72] Y. Nojiri, H. Yamanoue, A. Hanazato, F. Okano, "Measurement of Parallax Dis tribution and its Applacation to the Analysis Visual Comfort for Stereoscopic HDTV," Proc SPIE. **5006**, 195-205, 1993.
[73] S. Yano, M. Emoto, T. Mitsuhashi, “Two Factors in Visual Fatigue Caused by Stereoscopic HDTV Images,” Displays, **25**, 141-150, 2004.
[74] F. Speranza, W. J. Tam, R. Renaud, N. Hur, “Effect of Disparity and Motion on Visual Comfort of Stereoscopic Images,” Proc SPIE, **6055**, 94-103, 2006.
[75] M. Emoto, K. Masaoka, Y. Yamanoue, M. Sugawara, Y. Nojiri, “Horizontal Binocular Disparities and Visual Fatigue while Viewing Stereoscopic Display,” VISION, **17**(2), 101-112, 2005.
[76] K. Ukai, P. A. Howarth, “Visual Fatigue Caused by Viewing Stereoscopic Motion Images: Background, Theories and Observations,” Displays, **29**, 106-116, 2008.
[77] M. Lambooji, W. Ijsselstejin, M. Fortuin, I. Heynderickx, “Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review,” Journal of Imaging Science and Technology, **53**(3), 1-14, 2009.
[78] H. Takada, M. Miyao, F. Sina, Ed, Stereopsis and Hygiene, Singapore: Springer, 2019.
[79] S. Yano, “Size of Disparity for Binocular Fusion - A Study on Stimulus Target Properties -,” The transactions of the Institute of Electronics, Information and Communication Engineers, **75**(10), 1720-1728, 1991.
[80] M. Emoto, S. Yano, S. Nagata, “Distribution of Fusional Vergence Limit in Viewing Stereoscopic Image Systems,” The Journal of the Institute of Image Information and Television Engineers, **55**(5), 703-710, 2001.
[81] Safety Guidelines Sub-Committee of 3D Consortium. 3DC Safety Guidelines. 2011.
[82] T. Kojima, “Study of Visibility and Biomedical Effects of 3D Images,” Doctoral Dissertation of Graduate School of Information Science, Nagoya University, 2014.
[83] T. Shiomi, K. Uemoto, T. Kojima, S. Sano, H. Ishio, H. Takada, M. Omori, T. Watanabe, M. Miyao, “Simultaneous measurement of lens accommodation and convergence in natural and artificial 3D vision,” Journal of the Society for Information Display, **21**(3), 120–128, 2013.
[84] H. Takada, K. Yokoyama, Ed, Bio-information for Hygiene, Singapore: Springer, 2021.
[85] J. Suzuki, T. Matsunaga, K. Tokumatsu, K. Taguchi, Y. Watanabe, "Q&A on Stabilometry GuideBook (1995)," Equil Res, **55**, 64-77, 1996.
[86] T. Okawa, T. Tokita, Y. Shibata, T. Ogawa, H. Miyata, "Stabilometry: significance of locus length per unit area (L/A) in patients with equilibrium disturbances," Equil Res, **54**, 283–293, 1995.
[87] H. Takada, Y. Kitaoka, Y. Shimizu, “Mathematical Index and Model in Stabirometly,” Forma, **16**(1), 17-46, 2001.
[88] H. Takada, Y. Kitaoka, M. Ichikawa, M. Miyao, "Physical meaning of geometrical index for stabilometry," Equil Res, **62**(3), 168–80, 2003.
[89] M. Malińska, K. Zużewicz, J. Bugajska, A. Grabowski, “Heart rate variability (HRV) during virtual reality immersion,” Int J Occup Saf Ergon, **21**(1), 47-54, 2015.
[90] A. M. Baranowski, K. Keller, J. Neumann, H. Hecht, “Genre-dependent effects of 3D film on presence, motion sickness, and protagonist perception,” Displays, **44**, 53-59, 2016.
[91] T. H. Cho, C. Y. Chen, P. J. Wu, K. S. Chen, L. T. Yin, “The comparison of accommodative response and ocular movements in viewing 3D and 2D displays,” Displays, **49**, 59-64, 2017.
[92] A. Einstein, “Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen,” Annalen der Physik, **322**(8), 549-560, 1905.
[93] A. Einstein, “Zur Theorie der Brownschen Bewegung,” Annalen der Physik, **324**(2), 371–381, 1905.

[94] M. Takada, Y. Fukui, Y. Matsuura, M. Sato, H. Takada, "Peripheral viewing during exposure to a 2D/3D video clip: effects on the human body," Environ Health Prev Med, **20**(2), 79-89, 2015.
[95] P. A. Goldie, T. M. Bach, O. M. Evans, "Force platform measures for evaluating postural control: reliability and validity," Arch Phys Med Rehabil, 70, 510–517, 1989.
[96] H. Takada, Y. Shimizu, Y. Matsuura, T. Shiomi, M. Miyao, "Non-linear analysis of stabilograms with alcoholic intake," in Annu Int Conf IEEE Eng Med Biol Soc 2012, 4208-4211, 2012.
[97] K. Yoshikawa, H. Takada, M. Miyao, "Effect of Display Size on Body Sway in Seated Posture While Viewing an Hour-Long Stereoscopic Film," in Universal Access in Human-Computer Interaction. User and Context Diversity 2013 Part II, C. Stephanidis, M. Antona, Ed, Heidelberg: Springer-Verlag, 336-341, 2013.
[98] K. Fujikake, M. Miyao, R. Honda, M. Omori, Y. Matsuura, H. Takada, "Evaluation of High-Quality LCDs Displaying Moving Pictures, on the Basis of the Form Obtained from Statokinesigrams," Forma, **22**(2), 199-206, 2007.
[99] K. Fujikake, H. Takada, M. Omori, M. Miyao, "Evaluation of High-Quality LCDs Displaying Moving Pictures by Use of the Form Obtained from Statokinesigrams and the Dynamics," Forma, **22**(3), 217-229, 2007.
[100] R. S. Kennedy, L. E. Lane, K. S. Berbaum, M. G. Lilienthal, "A simulator sickness questionnaire (SSQ): A new method for quantifying simulator sickness," International J Aviation Psychology, **3**, 203-220, 1993.
[101] S. R. Holomes, M. J. Griffin, "Correlation Between Heart Rate and the Severity of Motion Sickness Caused by Optokinetic Stimulation," J Psychophysiology, **15**, 35-42, 2001.
[102] N. Himi, T. Koga, E. Nakamura, M. Kobashi, M. Yamane, K. Tsujioka, "Differences in autonomic responses between subjects with and without nausea while watching an irregularly oscillating video," Autonomic Neuroscience: Basic and Clinical, **116**, 46-53, 2004.
[103] Y. Yokota, M. Aoki, K. Mizuta, Y. Ito, N. Isu, "Motion sickness susceptibility associated with visually induced postural instability and cardiac autonomic responses in healthy subjects," Acta Otolaryngologia, **125**, 280-285, 2005.
[104] Y. Matsuura, H. Kato, Y. Mori, F. Kinoshita, T. Takaishi, H. Takada, "Evaluation of An Hour-Long Stereoscopic Film on Human Body by using Functional Test of Autonomic Nervous System," Bulletin of Society for Science on Form, **30**(1), 66, 2015.
[105] L. M. Scibora, S. Villard, B. Bardy, T. A. Stoffregen, Wider stance reduces body sway and motion sickness," in Proceedings of VIMS 2007, 18-23, 2007.
[106] H. Takada, Y. Mori, T. Miyakoshi, "Effect of Background Viewing on Equilibrium Systems," in Universal Access in Human-Computer Interaction. Access to Interaction, M. Antona, C. Stephanidis, Ed, Heidelberg: Springer-Verlag, 2015, 255-263.
[107] K. Yoshikawa, F. Kinoshita, K. Miyashita, A. Sugiura, T. Kojima, H. Takada, M. Miyao, "Effects of Two-Minute Stereoscopic Viewing on Human Balance Function," in Universal Access in Human-Computer Interaction. Access to Interaction, M. Antona, C. Stephanidis, Ed, Heidelberg: Springer-Verlag, 2015, 297-304.
[108] F. Kinoshita, Y. Mori, M. Miyao, H. Takada, "On mathematical models of two-minute stereoscopic viewing on human balance function," Forma, **32**(S), 11-17, 2017.
[109] List of Officials of Commerce and Information Policy Bureau, Ministry of Economy, Trade and Industry of Japan, Digital Content White Paper 2017, Tokyo: Digital Content Association of Japan, 2017.
[110] M. H. Fischer, "Optokinetic ausgeloste Bewegungs-wahrnehmungen und optokinetinetisher Nystagmus," Journal of Psychological Neurology, **41**, 273-308, 1930.
[111] T. Brandt, "Differential effects of central versus peripheral vision on egocentric and exocentric motion perception," Experimental Brain Research, **5**, 476-491, 1973.
[112] A. Koohestani, D. Nahavandi, H. Asadi, P. M. Kebria, A. Khosravi, R. Alizadehsani, S. Nahavandi, "A Knowledge Discovery in Motion Sickness: A Comprehensive Literature Review," IEEE Access, **7**, 85755-85770, 2019.

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 252-258 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Combustion Flame Temperature Considering Fuel and Air Species and Optimization Process

Prosper Ndizihiwe[1,*], Burnet Mkandawire[2], Kayibanda Venant[3]

[1]*University of Rwanda, Renewable Energy, Kigali, 4285, Rwanda*

[2]*Malawi University of Business and Applied Sciences, Mechanical Engineering, Blantyre, Private Bag 303, Malawi*

[3]*University of Rwanda, Electrical Engineering, Kigali, 4285, Rwanda*

ARTICLE INFO

Article history:
Received: 05 May, 2021
Accepted: 24 June, 2021
Online: 03 August, 2021

Keywords:
Air and fuel
Equivalence ratio
Carbon dioxide
Stoichiometry
Species of fuel

ABSTRACT

Estimation of optimal Air or oxygen is important for the combustion process to be efficient and produce more energy. This is to be based on each component of the fuel and the air, considering their respective pressure and density. At first, this research investigates the role of N_2, O_2, CO_2 present in combination with CH_4, and the air on the flame temperature; using simulation with Cantera 2.4. Results have been compared and calibrated with field data from KivuWatt company. It then demonstrates the way to achieve optimum Air Fuel Ratio (AFR) for the various species of the fuel. The results estimated the flame temperature by means of the percentages of all species of the fuel and the air, as well as various conditions of pressure and temperature. Finally, it combines all to show different values of optimum AFR at various species percentages; and uses a python program to create an AFR calculator available online through the link provided.

Nomenclature:

AFR: Air-Fuel Ratio
υ_F, υ_O: balancing constants
A_{act}: Actual quantity of oxygen
ϕ: Equivalence ratio
M: Mass
H: Enthalpy
υ: The number of moles
ω: The atomic weight
T_o: Reference temperature
c_p: Heat capacity
ρ: Density
P: Pressure
T: The temperature
O: Oxygen
H: Hydrogen
C: Carbon
S: Sulphur
R: Perfect gas constant
a_1 to a_6, coefficients of the thermodynamic system
Subscript F, i, in, OT, f, : fuel, any species, inlet, oxygen total, and formation respectively

1. Introduction

The combustion within the boiler burns fuel to create heat energy. The burning of fuel is the reaction of fuel with oxygen present in the air. The amount of fuel that can be burnt is limited by the oxygen present [1]. When all the fuel is not burnt, a part of it stays in the boiler and the other quantity goes to the atmosphere. This is the loss that reduces efficiency, and tends to pollute our environment [2]. Most of the fuels used in the boiler are hydrocarbons which release hydrogen and carbon as residuals, along with heat and pressure when burnt [3].

The quantity of these residuals and their temperatures impact the performance of the plant including the AFR [4]. The quantity of the exhaust depends both on the composition of the fuel, the composition of the air, and the effectiveness of the combustion [5]. In general, the global reaction of combustion is like

$$\upsilon_F F + \upsilon_O O_2 \xrightarrow{\boldsymbol{gives}} P \quad (1)$$

Let us see the combustion by taking into account the residuals within the fuel and the air.

1.1. Consideration of fuel and its impurities of the field and application

The general fuel formula is given by its composition of carbon, hydrogen, sulfur, oxygen, and nitrogen. So it is $C_a H_b S_c O_d N_e$ [6]

Combustion equation is

*Corresponding Author: Prosper NDIZIHIWE, Kigali, +25 0783058498 & ndizihiweprosper@gmail.com

$$C_aH_bS_cO_dN_e + O_2 \xrightarrow{gives} CO_2 + H_20 + SO_2 + N_2 \quad (2)$$

From this composition, the mass of the fuel can be computed as

$$M_f = aM_C + bM_H + cM_S + dM_O + eM_N \quad (3)$$

To achieve effective combustion, each element needs a determined quantity of oxygen as follows [7]:

a) a moles of O_2 are required to change C to CO_2
b) $b/4$ moles of O_2 are required to change H_b to H_2O
c) c moles of O_2 are required to change S_c to SO_2
d) The quantity of oxygen present in the fuel is subtracted from the quantity of oxygen required for complete combustion. That is, $d/2$ moles of oxygen are subtracted.
e) Nitrogen is present in the fuel however it doesn't undergo the combustion process (except at very high temperatures when some of it is converted to nitrogen oxides); hence it is not considered.

Therefore, the stoichiometric value of oxygen (A) is

$$A = a + \frac{b}{4} + c - \frac{d}{2} \quad (4)$$

In reality, the A_{act} is different from the stoichiometric value.

$$A_{act} = \phi A \quad (5)$$

Now let's consider air instead, If all elements of the air are involved in the combustion process equation (2) becomes,

1.2. Combustion with all species of the air

Let us define all the proportion (r) of each element compared with oxygen as $r = \frac{mk}{mO_2}$, i stands for any element. This gives $r_{O_2} = \frac{mO_2}{mO_2} = 1$, $r_{N_2} = \frac{mN_2}{mO_2} = \frac{78.96}{21} = 3.76$.

1.3. Brief on Cantera models

Cantera 2.4 is an open-source simulation software embedded in Matlab and Python used to solving dynamic chemical reactions [8]. In this paper the researchers used Python.

In the current work, the authors assume that all metallic impurities are omitted from the fuel. So, as to use Cantera simulation-based model summarized in equations (7), (8) and (9), taking methane as case. At stoichiometry, the equation is as follows:

$$CH_4 + 2(O_2 + 3.76N_2) = CO_2 + H_2O + 7.52N_2 \quad (7)$$

At rich combustion (when oxygen is lower), there is the formation of carbon monoxide as follows

$$CH_4 + (2/\phi)(O_2 + 3.76N_2) = (\frac{4}{\phi} - 3)CO_2 + 2H_2O + (4 - 4/\phi) + (7.52/\phi)N_2 \quad (8)$$

At lean combustion (when oxygen is higher), there is formation of oxygen in the product as follows.

$$CH_4 + (2/\phi)(O_2 + 3.76N_2) = CO_2 + 2H_2O + (4 - 4/\phi) + (7.52/\phi)N_2 \quad (9)$$

For both equations (8) and (9) above, if $\phi = 1$, they give (7). These models are connected with the AFR by:

$$AFR = \frac{\alpha A \sum_{i=1}^{n} r_i M_i}{\phi M_F} \quad (10)$$

where $s = ({}^{M_{air}}/_{M_F})_s$. At stoichiometry $\phi = 1$. Equation (10) shows that the equivalence ratio is higher when the air lessens, oppositely for the air fuel ratio, and there is an impact of the other species of the fuel and the air on the value of the AFR. The study has been done using the equivalence value instead of the AFR. By definition, the equivalence ratio is the ratio of actual fuel/ai (FAR) to the stoichiometric fuel/air [9]. The stoichiometric value occurs only when all elements and respective quantities are considered in computation [10].

Cantera uses those combustion principles and conservation of enthalpy in the combustion equation at constant pressure [11] to find the value of the final temperature. That is, the enthalpy of the reactant is equal to the enthalpy of the product. Writing the described global combustion equation in the way that allows quantifying masses the reactant is at the temperature T_1 and the product at T_2.

$$C_aH_bS_cO_dN_e + \phi A[O_2 + r_{N_2}N_2 + r_{CO_2}CO_2 + r_{H_2O}H_2O + r_{Ar}Ar] \xrightarrow{gives} (a + \alpha Ar_{CO_2})CO_2 + \left(\frac{b}{2} + \phi Ar_{H_2O}\right)H_20 + cSO_2 + (\phi - 1)AO_2 + \left(\frac{e}{2} + \phi A\right)N_2 + \phi AAr \quad (6)$$

$$\underbrace{\sum_{i=1}^{n} v_i M_i}_{T_1} \xrightarrow{gives} \underbrace{\sum_{i=1}^{n} v'_i M_i}_{T_2} \quad (11)$$

Now the conservation principle gives

$$H(T_1) = H(T_2) \quad (12)$$

$$H(T_1) = \sum_{i=1}^{n} v_i(\Delta H_{fi}^o + \int_{T_o}^{T_1} c_{pi}\, dT) \quad (13)$$

$$H(T_2) = \sum_{i=1}^{n} v'_i(\Delta H_{fi}^o + \int_{T_1}^{T_2} c_{pi}\, dT) \quad (14)$$

Using equation (12) yields

$$\sum_{i=1}^{n} v'_i(\Delta H_{fi}^o + \int_{T_1}^{T_2} c_{pi}\, dT) = \sum_{i=1}^{n} v_i(\Delta H_{fi}^o + \int_{T_o}^{T_1} c_{pi}\, dT) \quad (15)$$

Enthalpies of formation of molecular products are taken from thermodynamic table present in [12], so T_2 is the only unknown of equation (15). With Cantera, computation to deduce the value of T_2 is performed for all (16), and (8) cases, at different values of ϕ.

The enthalpy is calculated by [13], [14]

$$H = RT(a_1 + a_2T/2 + a_3T^2/3 + a_4T^3/4 + a_5T^4/5 + a_6/T, \quad (16)$$

[15] generated by NASA at standard pressure; which indicates that to have higher flame is important in view of yielding more energy.

1.4. Algorithm for optimum AFR

The current section shows the algorithm for realizing optimum AFR based on the results from chapters 3 and 4 taking into account the fact that each species present in the fuel is to undergo complete combustion by a specified quantity of air.

The composition of the species in a hydrocarbon is provided in Table 1 considering most present composition species [16], [17]

Table 1: Species composition

Species	Composition range(%) [18]
C	$\%_C$: [48, 68]
H	$\%_H$: [25, 47]
S	$\%_S$: [0, 8]
O	$\%_O$:[8, 18]

$$\%_C + \%_H + \%_S + \%_O = 100 \quad (17)$$

For complete combustion, the stoichiometric value ($S = AFR_s$) is computed following (11) by

$$S = \frac{v_o\omega_o}{v_F\omega_F} \quad (18)$$

[19].

The value of the mass of oxygen to make combustion of each species i will be

$$M_{o(i)} = S_i M_F \%_i \quad (19)$$

$\%_i$ is the percentage of species i. In practice, the carbon present in the fuel is the source of carbon dioxide; hydrogen is the source of water, sulphur the source of sulphur dioxide [20].

$$\begin{cases} C + O_2 \rightarrow CO_2 \\ H_2 + 0.5O_2 \rightarrow H_2O \\ S + O_2 \rightarrow SO_2 \end{cases} \quad (20)$$

Using (18) and (19) gives the total mass of oxygen required to burn each element

$$\begin{cases} M_{o(H)} = \frac{0.5 * 32}{1 * (1 * 2)} * M_F * \%_H \\ M_{o(C)} = \frac{1 * 32}{1 * 12} * M_F * \%_s \\ M_{o(S)} = \frac{1 * 32}{1 * 32} * M_F * \%_s \end{cases} \quad (21)$$

$$M_{0(C)} + M_{0(H)} + M_0 + M_{0(S)} = M_{oT} \quad (22)$$

Because oxygen is 21% of the air, the mass of air (M_{air}) is computed by (23).

$$M_{air} = {M_{oT}}/{0.21} \quad (23)$$

Since oxygen composes 21% of the air.

$$M_{air} = 4.762 * M_{oT} \quad (24)$$

$$AFR = {M_{air}}/{M_{fuel}} \quad (25)$$

The algorithm of the air-fuel ratio and mass of the air is simply represented by Figure 1

This work deals with the estimation of the flame temperature at different compositions of the fuel and the air for various values of the air-fuel ratio and equivalence ratio. It also presents the method of reaching the optimum value of the air-fuel ratio and the mass of the air, taking into account initial pressure and temperature.

It has four sections: Section 1 is the introduction; section 2 for methodology, section 3 presents the result and its interpretation and finally concludes in section 4.

2. Methodology and process

Referring to models described above, numerical simulation is done with Cantera codes present in python following the equations (7) to (9) and (15) then the results are compared with KivuWatt field data. KivuWatt: Is a thermal power plant built in Rwanda/Karongi district. This is part of Contour Global plc, is producing 26 MW since 2010, and is using Methane gas from lake Kivu [21] [22].

The value of the nitrogen/air ratio, carbon/air ratio, Nitrogen/fuel ratio, and Oxygen/fuel ratio is varied from zero to one at specified constants equivalence ratio under standard temperature and pressure. The value of the enthalpy is estimated by using formula (16), where the final/flame temperature used is of result from the simulation. The value of the enthalpy of formation used is $52MJ/kg$ [23], and the heat capacity is $35.07\ (J/molK)$ at 300K [24].

The algorithm is based on the results of recent publications, explaining the role of the pressure, temperature, and density on the AFR has been demonstrated. Putting this together with results from Cantera simulation gives the procedure summarized by Figure 1 to come up with calculation and online calculator.

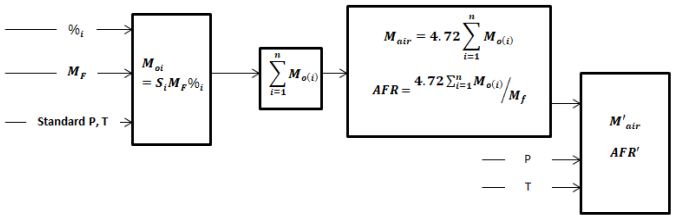


Figure 1: Summary of optimum AFR process

The variation concerning the density has been analyzed from the results of pressure and temperature using the state equation since the is for pressure and temperature are very high [25].

$$\rho = \frac{P}{rT} \tag{26}$$

$r = {}^{R}/_{\omega}$, $R = 8.31$ is the constant of a perfect gas.

3. Results and Discussion

3.1. Simulation and its comparison with field data

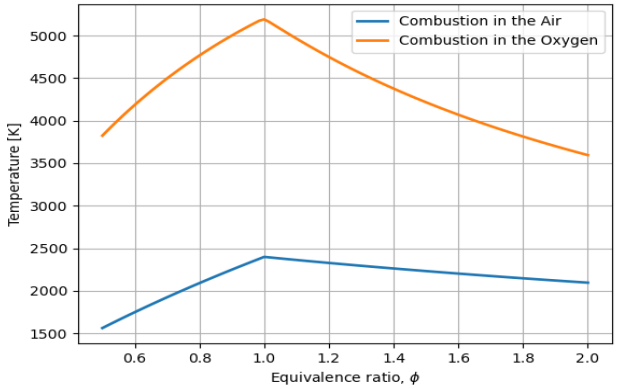


Figure 2: Comparison of Flame temperature for Oxygen and Air

Figure 2 indicates that combustion is much more efficient when it is done with oxygen.

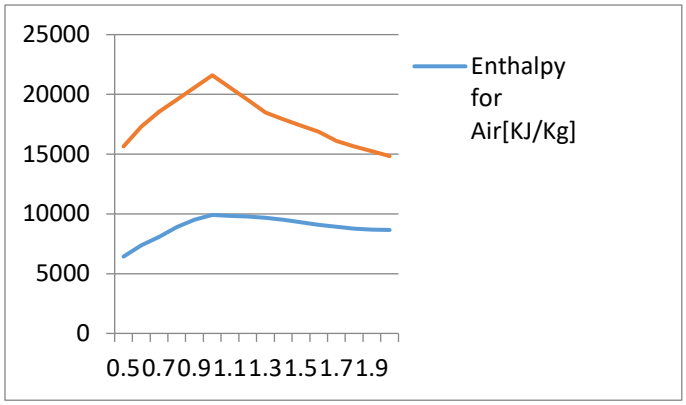


Figure 3: Comparison of Combustion enthalpies for oxygen and air

Figure 3 estimates the value of the enthalpy, computed by using the result of Figure 2 for both cases of combustion in oxygen and air at different values of the equivalence ratio.

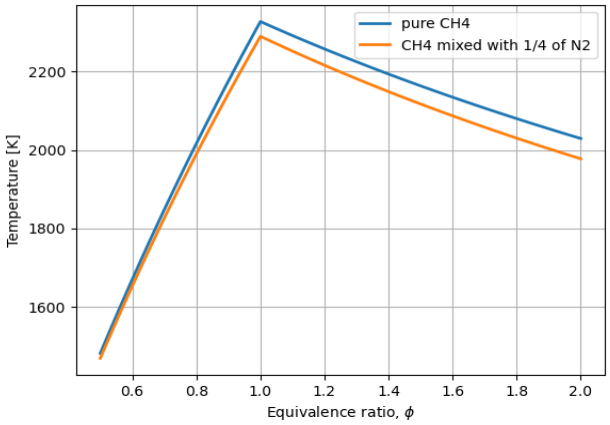


Figure 4: Role of the presence of nitrogen in the fuel

Figure 4 quantifies the resulting flame temperature in a case where a quarter of the fuel is nitrogen. It is visible that the temperature is lowered when the fuel contains nitrogen as an impurity.

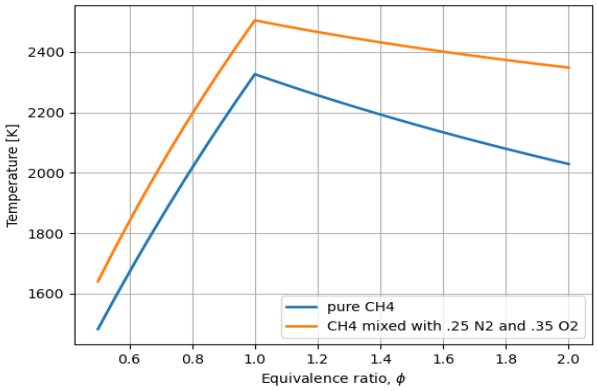


Figure 5: Role of oxygen and nitrogen in the fuel

Figure 5 shows how much flame temperature is affected by the presence of oxygen and nitrogen in the fuel. They lower the temperature and comparing with Figure 4, oxygen itself does not negative effect.

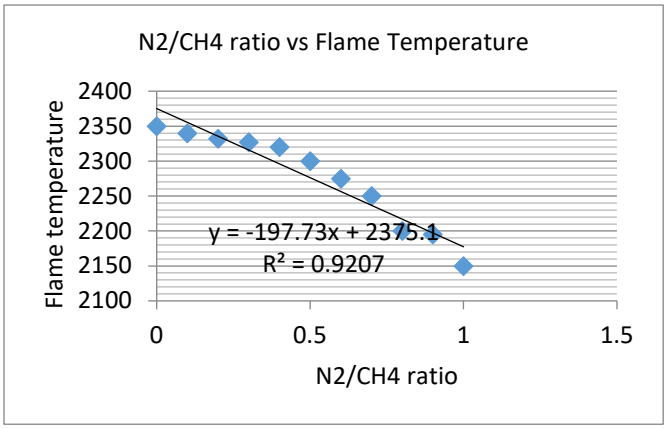


Figure 6: Influence of portion of nitrogen

Figure 6 is the results from the analysis of field data. It shows that as nitrogen in the fuel goes up, the flame temperature comes down. Maximum flame temperature is achieved for a case where there is no nitrogen ($\frac{N_2}{CH_4} = 0$).

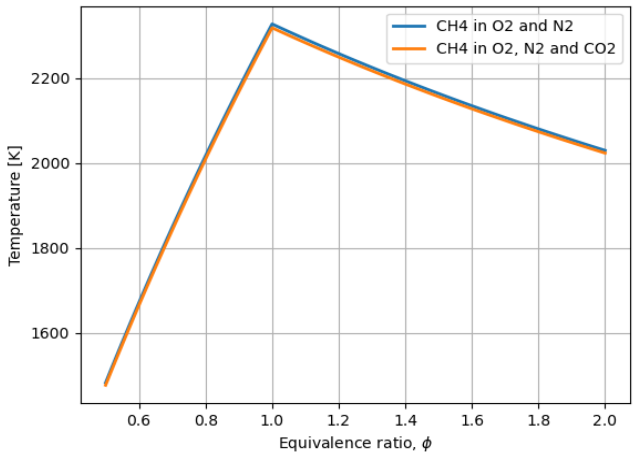


Figure 7: Role of the presence of carbon dioxide

Figure 7 analyses the impact of carbon dioxide present in the air. This shows that carbon dioxide has a very small negative impact on the flame temperature.

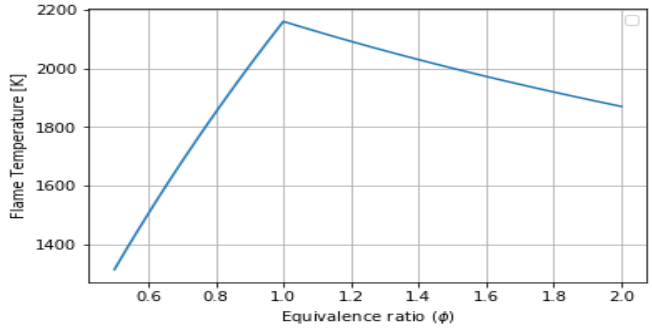


Figure 8: Equivalence Ratio vs Flame Temperature for Air-Fuel combustion, T in: $100K$

Figure 8 indicates the values of flame temperature when T_{in} is very small ($100K$). The comparison with Figure 2 (graph in blue), shows that inlet temperature is to be increased to have more flame temperature.

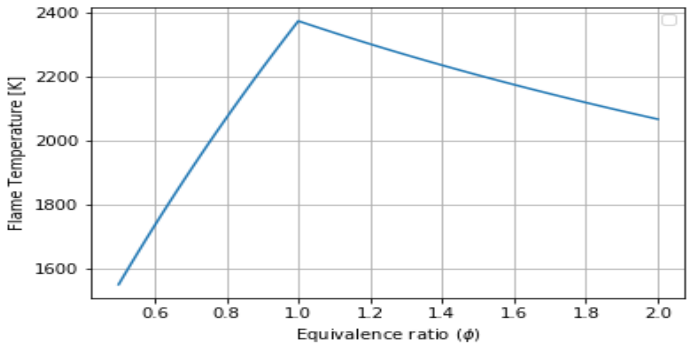


Figure 9: Equivalence Ratio vs Flame Temperature for Air-Fuel combustion, Tin: $400K$

Comparison of Figure 2 (graph in blue), Figure 8, and Figure 9 show the increase of T_{in} from 100K to 293K then to 400K, but the flame temperature has increased from 2150K to 2400K, then to 2350K, respectively, at $\phi = 1$ tells that inlet temperature would be improved, but when it becomes higher the flame temperature becomes very low.

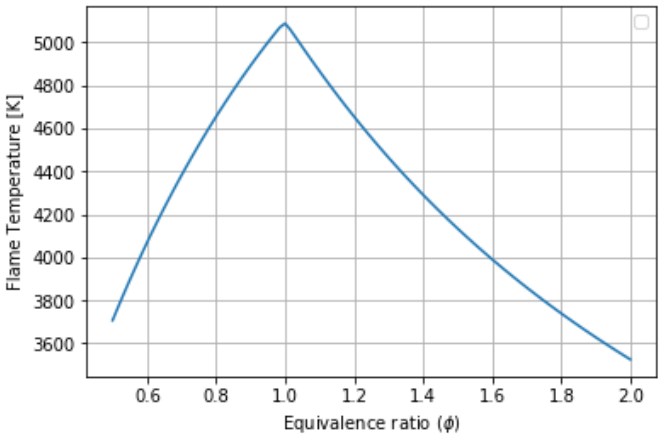


Figure 10: Equivalence Ratio vs Flame Temperature for O_2 and CO_2-Fuel combustion, Tin: 300K

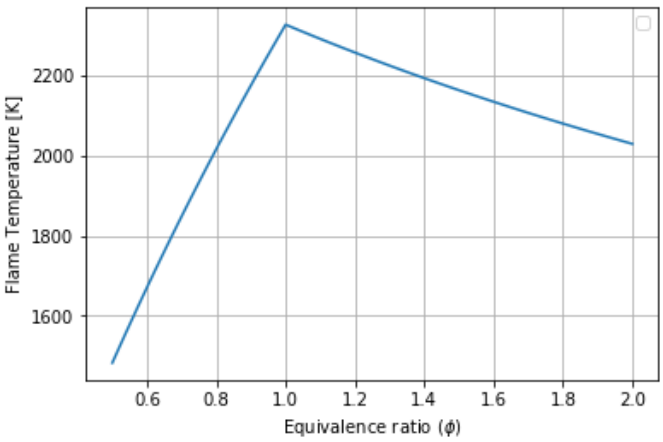


Figure 11: Equivalence Ratio vs Flame Temperature for O2 and N2-Fuel combustion, Tin: 300K

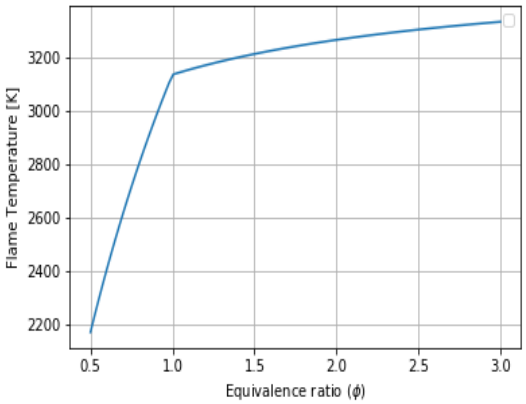


Figure 12: the flame temperature for O_2 present in the fuel

Comparison of Figure 10 with Figure 10 emphasizes what has previously been demonstrated in Figure 7 at a bit increase of inlet temperature from ambient (293K) to 300K.

In Figure 12, the flame temperature is higher when $\phi > 1$. So, it informs that when there is oxygen in the fuel, the air would be reduced, thus the AFR is to be smaller than the stoichiometric value.

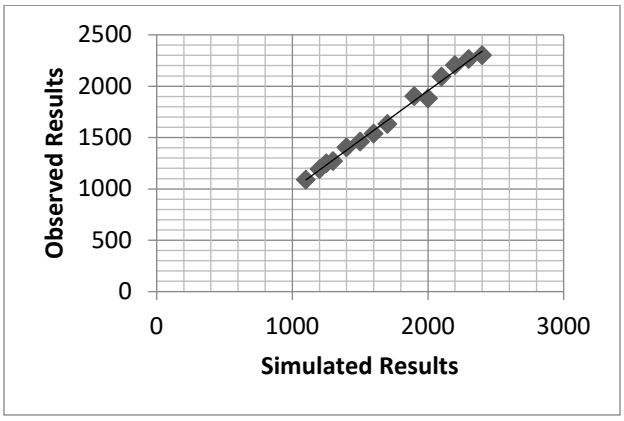


Figure 13: Calibration of flame temperature [^{0}C]

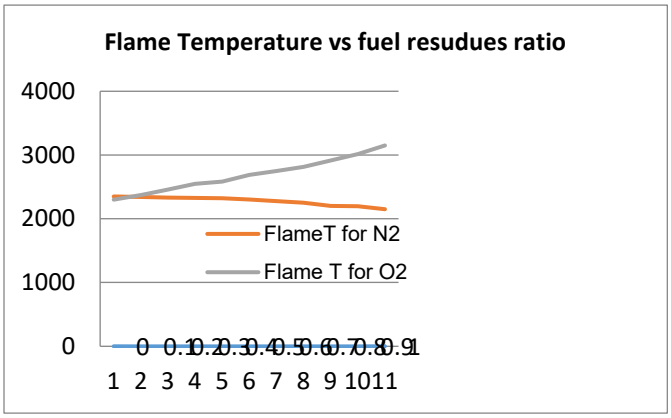


Figure 14: Flame Temperature vs fuel residues

Figure 14, resulting from the analysis of field data, emphasizes the results in Figure 12. It indicates that the presence of oxygen in the fuel is positive but nitrogen is negative.

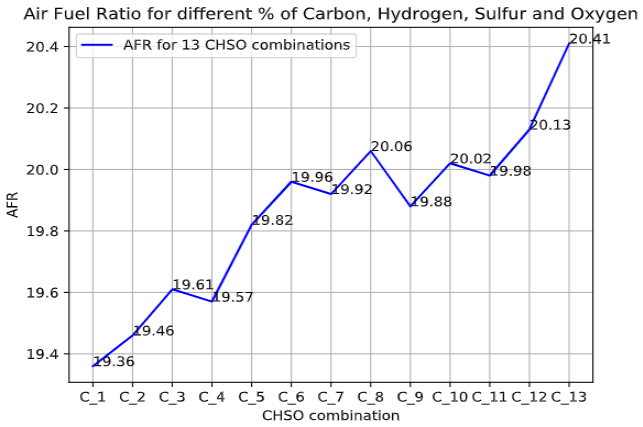


Figure 15: AFR for different species percentages at standard condition of pressure and temperature

3.2. *Algorithm results*

For different combinations of the fuel species percentages at the standard condition of pressure and temperature, specific values of the AFR are plotted in Figure 15. It shows that changes in species percentages (from C_1 to C_13) correspond to different values of AFR at constant pressure, temperature, and density (standard condition).

Various values of AFR for different values of pressure, temperature and density are for fixed species percentages of the fuel plotted in Figure 16.

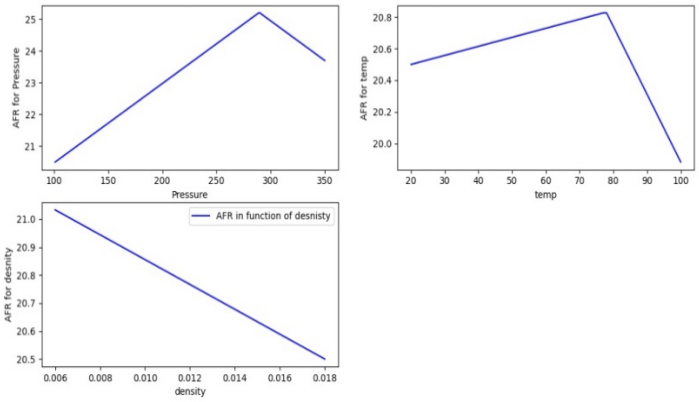


Figure 16: Variation of AFR with Pressure [KPa], Temperature [°C] and Density [kg/m^3]

Variation of both species' percentages and pressure-temperature states are considered in the program to get specific values. The method to compute the AFR producing optimum flame temperature is built following the algorithm and accessible online through the link;

http://ndipros.pythonanywhere.com/airfuelratio/

4. Conclusion

This research quantified the flame temperature at specific values of the fuel and air species. Analysis also showcases that it is feasible to employ a measured quantity of air for combustion efficiency. Again, it contains the method to calculate and model-based calculator to compute the AFR and mass of the air to be used for optimum power output and reduce exhausts. Practical feasibility requires a method to measure percentages of all chemical species within the fuel and the air, and the controller of the boiler combustion process, this will be the next research. A built calculator is hosted online for accessibility. The study has shown that presence of oxygen in the fuel is positive in this case the air is to be reduced proportionally. The combustion within the oxygen has a more remarkable positive impact than in the air. It is better to separate oxygen from the air before combustion which is not easy. Preheating the fuel is also an advantage, however, this should be done only up to a point where a good viscosity and density are reached, since uncontrolled preheating reduces the output temperature and requires some time and cost.

References

[1] R. Pradhan, P. Ramkumar, and M. Sreenivasan, "Air-Fuel Ratio (Afr) Calculations In An Internal Combustion Engine Based On The Cylinder

Pressure Measurements," Int. J. Eng. Res. Apllication, **2**(6), 1378–1385, 2012.
[2] B. Abbas Al-Himyari, A. Yasin, and H. Gitano, "Review of Air-Fuel Ratio Prediction and Control Methods," Asian J. Appl. Sci., **2**(4), 471–478, 2014, [Online]. Available: www.ajouronline.com.
[3] A. Marjanovi, "Control of Thermal Power Plant Combustion Distribution Using Extremum Seeking," in IEEE Transactions on Control Systems Technology, 2017, **25**(5), 1670–1682.
[4] T. K. Ibrahim, M. M. Rahman, and A. N. Abdalla, "Optimum gas turbine configuration for improving the performance of combined cycle power plant," Procedia Eng., 15(September 2015), 4216–4223, 2011, doi: 10.1016/j.proeng.2011.08.791.
[5] N. Stuban and A. Torok, "Utilization of exhaust gas thermal energy – theoretical investigation," June 2010, 2014, doi: 10.1109/ISSE.2010.5547301.
[6] K. O. Povarov, "Distribution of impurities and gases between steam and water phases of the geothermal fluid in the low pressure zone," Sci. Train. Res. Cent. Geotherm. Energy, **14**(5), 1–12, 1996.
[7] J. Colannino, "Introduction to Combustion Analysis," Model. Combust. Syst., 2(13), 101–189, 2006, doi: 10.1201/9781420005035.ch2.
[8] A. Felden, "Cantera Tutorials - A series of tutorials to get started with the python interface of cantera.," Cerfacs(November 2015, [Online]. Available: https://www.cerfacs.fr/cantera/docs/tutorials/CANTERA_HandsOn.pdf.
[9] R. T. Vedula, R. Song, T. Stuecken, G. G. Zhu, and H. Schock, "Thermal efficiency of a dual-mode turbulent jet ignition engine under lean and near-stoichiometric operation," Int. J. Engine Res., **18**(10), 1055–1066, 2017, doi: 10.1177/1468087417699979.
[10] P. K. R. B. W. K. G. Wise, "Exploring Engineering," Elsevier, 5(3), 1–656, 2020.
[11] S. Londerville, J. Colannino, and C. E. Baukal, "Combustion fundamentals," John Zink Hamworthy Combust. Handbook, Second Ed. 1 - Fundam.), 79–124, 2012, doi: 10.1201/b11619.
[12] M. J. Simpson, Two Studies in Gas-Phase Ion Spectroscopy, 53(9. Springer Verlag Berlin Heidelberg, 2013.
[13] B. Mcbride, S. Gordon, and M. Reno, "Coefficients for Calculating Thermodynamic and Transport Properties of Individual Species," Nasa Tech. Memo., 4513(NASA-TM-4513, 98, 1993, [Online]. Available: http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19940013151_1994013151.pdf.
[14] B. Franzelli, E. Riber, M. Sanjosé, and T. Poinsot, "A two-step chemical scheme for kerosene-air premixed flames," Combust. Flame, **157**(7), 1364–1373, 2010, doi: 10.1016/j.combustflame.2010.03.014.
[15] H. Pitsch, "Thermodynamics , Flame Temperature and Equilibrium," 2018. https://cefrc.princeton.edu/sites/cefrc/files/2018_pitsch_lecture2.pdf (accessed Apr. 30, 2021).
[16] M. Bajus, "Sulfur Compounds in Hydrocarbon Pyrolysis," Sulfur reports, **9**(1), 25–66, 1989, doi: 10.1080/01961778908047982.
[17] D. Wu, W. Pisula, M. C. Haberecht, X. Feng, and K. Müllen, "Oxygen- and sulfur-containing positively charged polycyclic aromatic hydrocarbons," Org. Lett., **11**(24), 5686–5689, 2009, doi: 10.1021/ol902366y.
[18] A. M. Dadile, O. A. Sotannde, B. D. Zira, M. Garba, and I. Yakubu, "Evaluation of Elemental and Chemical Compositions of Some Fuelwood Species for Energy Value," Int. J. For. Res., **2020**(6), 1–8, 2020, doi: 10.1155/2020/3457396.
[19] D. V. Thierry Poinsot, "Theoretical and Numerical Combustion, Second Edition," Decis. Support Syst., **38**(4), 557–573, 2005.
[20] W. H. Green, "Combustion Chemistry," Princeton-CEFRC Summer Sch. Combust.(June, 1–110, 2014.
[21] E. Plant, "KivuWatt Project," Power Technology, 2021. https://www.power-technology.com/projects/kivuwatt-project-lake-kivu-kibuye/ (accessed Feb. 14, 2021).
[22] S. Nagarajan et al., "PROJECT : KIVUWATT POWER PLANT," Kigali, 2006.
[23] B. Atakan, "Compression – Expansion Processes for Chemical Energy Storage : Thermodynamic Optimization for," Energies, **5**(11), 1–21, 2019, [Online]. Available: http://dx.doi.org/10.3390/en12173332.
[24] P. J. Linstrom and W. G. Mallard, "Gas Phase Thermochemistry Data," NIST Chemistry WebBook, NIST Standard Reference Database Number 69, 2010. http://webbook.nist.gov (accessed Apr. 25, 2021).
[25] G. De Montricher and J. M. Stuchly, "Equation of state," PSIG Annu. Meet. PSIG 1985, **5**(4), 1–17, 1985, doi: 10.1201/9781351034227-3.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 259-277 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Software Development Lifecycle for Survivable Mobile Telecommunication Systems

Mykoniati Maria*, Lambrinoudakis Costas

Department of Digital Systems, University of Piraeus, Piraeus, 18532, Greece

ARTICLE INFO

Article history:
Received: 21 March, 2021
Accepted: 19 July, 2021
Online: 03 August, 2021

Keywords:
Telecommunication Systems
Software Development Lifecycle
Survivability

ABSTRACT

Survivability of systems is a very important system property and consists major concern for organizations and companies. Survivable systems should maintain their critical services functional in a timely manner. There are several approaches, proposed in the literature, on how to develop survivable telecommunication systems, but the majority is based on node outages or path failures, missing the main scope of survivability which is service failure. The contribution of this paper is that it presents a SDLC (Software Development Life Cycle) for developing survivable mobile telecommunication systems. Additionally, the main characteristic of a mobile telecommunication system is that it consists of different types of nodes (ex. MME, SGSN, etc.) that are connected to systems (ex. 5G, 4G, 3G, 2G etc.) and thus form an intersystem that provides services to end users. This interconnection and interoperability of network nodes is of high complexity constituting a threat to system survivability. Thus, another contribution of the current research work is that it provides a systematic approach for handling this complexity.

1. Introduction

Availability and continuity of critical IT infrastructures is a matter of concern in many of scientific fields like security, robustness, fault tolerance etc. In fact, the unavailability and failure of such infrastructures causes severe financial losses to many organizations.

Survival of IT infrastructures, like information systems or network systems is a matter of concern for any company that develops and maintains network systems. That means that such systems should continue to support the critical services even during attacks, failures or accidents. A definition of survivability is: "survivability is the capability of a system to fulfil its mission, in a timely manner, in the presence of threats such as attacks or large-scale natural disasters" [1], with security, robustness, fault-tolerance and recovery of systems to be among survivability's main disciplines.

It is important to highlight that survivability focusses on the survival of the mission of the system and not of the system itself. This is the core principle of survivability.

There is much research on survivability measures and approaches that should be adopted by a system to be survivable. But how can we be sure that a system is survivable? What are those capabilities that should be tested in order for a system to be characterized as survivable and against which threats? Additionally, what are the interconnections and interoperability threats that should be considered when survivability of large complex system of systems, like mobile telecommunication systems, is examined and how could these be analysed at everyday work when building such systems?

Through literature review, a detailed research on survivability approaches is presented highlighting that most of them address survivability of telecommunication networks by handling node or path outages. However, survivability should be based on service failure and not on system failure. In fact, even if the entire network is performing as expected, there could be failures in services for many other reasons. For example, a software bug could result in a specific service failure, or delays caused by excessive load in specific network nodes could result in random service failures. Another reason could be that robustness requirements are not considered during system design. A very representative example is the handling of collision scenarios, where two messages requesting a service arrive at the same node simultaneously. Robust system design could resolve this conflict.

To conclude, the contributions of the current paper are:

- The solution proposed by the current paper is Survivability by Design, meaning that survivability should be part of the software development lifecycle (SDLC) of the telecommunication system. The idea comes [2] which is a

*Corresponding Author: Mykoniati Maria, mmykoniati@gmail.com

paper titled as "Life-Cycle Models for Survivable Systems", that proposes survivability to be part of the SDLC phases and describes how this could be achieved. This is the theory that the current research is based on to describe how survivable telecommunication systems shall be developed.

- Another contribution of the current paper is that it addresses the risk of service failures arising from the increased complexity of interconnection and interoperability of mobile telecommunication network nodes. This is a major concern since most of the times development teams tend to focus only on the node under development, when new features are to be developed, without taking into consideration requirements or threats coming from connectivity with the other nodes. More specifically, even if the entire system is tested end-to-end, when a mobile telecommunication network node is operating in the provider's environment, it may be connected to nodes developed by other companies. The behaviour of that node is unpredictable, and this should be considered during SDLC phases, by setting appropriate survivability requirements and design practices, and by testing without ignoring specific failure scenarios.

During the next chapter, survivability as a term is examined in order to present the main principles and requirements of survivability. Following this literature review, the general framework in the form of a software development lifecycle (SDLC) is presented. Finally, the paper closes with overall conclusions.

2. Literature Review

2.1. Survivability as a term

As described in [1], survivability is the ability of a system to maintain its critical services that serve system's mission in a timely manner in case of attacks, failures or disasters. As a result, survivability itself is a system property that the system should emerge and should be considered as a requirement during the design phase and not as an ad-on characteristic [2]. Additionally, since the focus is on critical services and system mission, survivability should be considered as a different set of characteristics for each system, based on system's scope. For example, for a telecommunication network, survivability as a requirement may include, define and implement mechanisms that would allow the system to feature robustness, fault-tolerance, interoperability, restorability, security, safety, resilience, dependability etc, for its critical services in order to provide uninterrupted communication to end users. For an e-shop, usability or secure transactions would also be key principles for the survival of the mission of the system. There is much research on gathering these characteristics to a general set for systems' design, with the most representative one being the research described in [2]. They argue that for any system survivability is succeeded if it has the ability to provide Resistance, Recognition and Recovery (3Rs) from attacks or failures. In extend the system should provide Adaptation and Evolution by improving system survivability and increasing its resistance by knowledge gained from previous attacks or failures.

Threat for the survivability of a system, according to [3], is anything that may prevent the system from providing its essential services under the "minimum acceptable level of service", or affecting the provision of its essential services for more time than the one predefined as acceptable. As a result, the threat against a system's survivability is unknown and not always predictable through a risk analysis. Therefore, it is critical for survivability to gather, analyse and deal with the impact threat incident may cause, rather than focussing on predicting all possible threats. For instance, from the "survivability point of view", it is more important to focus on how a network node would behave under a Denial of Service attack and how it could recover rather than identifying measures that would prevent this attack.

2.2. Survivable Systems

Having defined survivability, a brief description of different approaches that have been adopted for designing and implementing a system that satisfies the survivability requirements follows.

Starting with the Survivability Analysis Framework (SAF) [4], survivability is considered as a set of peoples' capabilities, a set of actions and of technology working together to achieve operational effectiveness. The focus is on interoperability of organizational components and how to cope with complexity arising from this interoperability in order to analyse potential failure conditions, likelihood of error conditions, impact of occurrences, or recovery strategies. This analysis yields requirements for the design and implementation of the system.

The second approach considers survivability as part of the system's development life cycle. It is described by research [2] and claims that "survivability goals and methods must be addressed for each action of the life-cycle", as survivability should be integrated into the primary development phase of system and not treated as an add-on property of an already implemented system. Starting with requirements specification, the system should be able to monitor itself in order to recognise attacks or failures, resist and recover from attacks and failures and reconfigure to adapt to attacks and failures. Additionally, after mission definition, essential services of system should be depicted, and the system should be designed in such a way so that to maintain these services when it is under attack or failure. Continuing with requirements, intrusion requirements should be defined, in order for the performance of the system under attack or failure to be defined, in order to ensure that acceptable levels of quality of service are always reached. What is important here is that intrusion scenarios are considered as usage scenarios to be handled. The testing of these requirements should include three attack phases, the penetration phase, where the intruder attempts to gain access to the system, the exploration phase, where the intruder has gained access and is exploring the integral system organization and capabilities to find possible exploitation targets, and the exploitation phase where the intruder performs attacks against system facilities. According to these phases, survivability strategies for resistance, recognition, recovery, adaptation and evolution must be enforced. By considering these requirements, the system may be designed and implemented as survivable.

The third approach is presented in [5], and it is based on analysing the different states of quality of service, that the system may fall into during a failure, and on estimating the probability of

the essential services being available during the failure. After changes to the environment or attacks to the system, the system may degrade to the next quality of service level. When failure is restored, the system may return to the higher QoS level. Acceptable QoS levels for the system and transitions between them, may be modelled with the use of a transition matrix.

Another approach for providing survivability is the one proposed by [6]. Contrary to the security approaches that try to prevent an attacker to gain access, the assumption here is that the attacker has gained access and the objective is to try to find ways to prevent him from interfering with systems' critical services. Methods of prevention are based on frustrating the attacker to believe that he or she has gained access to essential services.

A fifth approach is presented in [7] known as the WILLOW architecture. It is a proposal that focuses on proactive and reactive reconfiguration of a system in order to achieve survivability for its services. During proactive reconfiguration, it is possible to add, remove and replace components and interconnections of the system, as well as to adjust their mode of operation. This is called posturing and is used to minimize the system's vulnerabilities that can be exploited by various threats. For instance, such a reconfiguration may be to turn-off non-essential services and networking links as well as to strengthen the cryptographic keys if a virus has infected the system. The reactive configuration does the same actions, aiming to restore a system from damage or intrusions, in specific time intervals. In fact, as proposed, the most appropriate approach for reacting is fault tolerance. An example, of reactive reconfiguration against an attack or damage is the activation of applications' copies.

A similar approach of reconfiguring the system and switching to different level of quality of service is also provided in [8], where the authors claim that QoS and survivability are firmly connected. As a result, if QoS is to be measured, reconfiguration approaches may be triggered under certain measurements to provide survivability for the system. Firstly, as "survivable system", may be characterized, any system that may repair itself or degrade in such a way that will provide as much functionality as possible. This may be done if the system is able to switch between alternatives of acceptable predefined levels of functionality. Secondly, a survivable system is a system that may adapt threats in its environment and environmental changes and reallocate essential processing to most robust resources. All these may be achieved through dynamic reconfiguration. Such reconfiguration may be "process/host restart, migration of objects to alternate hosts, replication, transparent rebinding of clients and servers, use of service alternatives, and approximate services". [8] These reconfigurations may be based on several metrics like "available battery power, varying communication bandwidth, available memory or faults in software components" [8] and must be done in predetermined time and based on QoS service levels. Then a survivable system must provide a minimum level of QoS under changing environments. For that purpose, the best-suited elements are to be chosen at each time, based on these QoS factors.

2.3. Evaluation of System Survivability

According to related literature, evaluation of systems' survivability, is mainly based on defining different acceptance levels of system performance and on evaluating the impact by measuring the key properties like number of outages, time needed for system recovery etc. Though, these evaluation models are mostly based on node failures or link failures, but they are not giving the whole idea about the quality of service the system provides to end users. As a result, they seem to be based on system availability and continuity and not on critical services or system's mission availability. Of course, system's availability is of vital importance for supporting system's mission and providing end to end functionality. So, system availability should be part of any survivability analysis and evaluation plan. Thus, the purpose of this paper is to provide an entire evaluation framework of all survivability aspects and not only providing system - centric evaluation methods. As a result, many of these evaluation models could be very useful to pinpoint any possible network failures and include these in a test suite that would test if the system could recover from them or if it could function as expected while the system is suffering from these failures. But it is very important to provide guidance for testing or evaluating systems' survivability from the requirements specification step of a SDLC, up to the release of new product.

Starting with [9], the authors use a Markov model to map the possibility of a failure. They base survivability measurements on the frequency of failure events, on the duration of outages and on the impact of failure. Since the research is conducted through a case study with wireless networks, as a failure is considered node failure, power faults and link failures. A similar approach is proposed by [10], where the authors are using a semi-Markov survivability evaluation model for intrusion tolerant database systems. As key attributes for quantification of a database's survivability, integrity and availability are proposed. Much focus is paid on system's functionality under failure and how system performs against these attributes.

To continue with quantification of system's survivability, the author in [11], proposed network condition metrics which are density (based on topology and its changes), mobility (speed of node, predictability etc.), channel (bit error rate, capacity distribution etc.), node resources (memory, computing power etc.), network traffic (QoS, packet size, distribution etc.), derived properties (degree of connectivity, queueing delay, propagation delay etc.). In addition to those metrics, service requirements are also defined. Again, every adverse event, transits system's performance to another state which is quantified by these measurements (based on network and service performance) in order to be marked as acceptable or not. Another approach based again on a Markov model is being presented in [12]. It is focused on call losses of a telecommunication switching system because of various system failures like hardware/software faults, human errors, impairment damage from adverse environments etc. As key survivability metrics, system performance, availability and performability are used and the measurements proposed are measurements that can be used to describe system survivability such as the number of functional units, the number of connected nodes, the maximum traffic capacity, blocking probability, throughput/goodput, and the service restoration time.

To continue with evaluation methods, in [13], authors propose a testing survivability framework, focusing again on the recovery part of the survivability attributes. They firstly present the idea of 5-step phases of survivability of a system under failure, normal

phase, resistance phase, destroyed phase, recovery phase and adaptation phase. Then they propose a scheme for representing the different stages of system performance against time during these phases. For quantification of network performance, two factors are proposed to be used, the Node Connectivity Factor (NCF) and the Link Connectivity Factor (LCF). Practically though, they try to focus on the availability of an end-to-end activity for the end user which is what really matters. This is why their research focuses on source-destination pairs "SD-pairs", to describe connectivity and service quality "SD-quality" and test these factors by applying different failures in order to calculate SD Recovery time for each pair. Finally, NRD metric is calculated to give an overall idea about the entire system's survivability.

Another very important research on evaluation of survivability has been conducted by authors in [14]. The framework proposed, is based on developing a general measurement model, which may be specified based on specific domain requirements, a network survivability testing model, which is based on testing network performance against survivability metrics during different steps of system performance (resistance, destroy, recovery), and the network survivability evaluation, which includes measurement of the entire system's survivability based on different metrics, evaluation models or algorithms. The method concludes to a mechanism which if applied to the system under test, may provide all possible combinations of test schemes to test failures of a network and to measure them in order to extract conclusions on the overall system's survivability.

In [15] the authors propose measuring survivability through four attributes, Process-Weighted Average Availability (PWAA), Process-Weighted Average Controllability (PWAC), Process-Weighted Average Robustness (PWAR), Process -Weighted Average Adaptability (PWAD). These depict the state of the system through survivability life cycle, which is normal state, resistance state, destroy state, recovery state and adaptation phase.

Finally, another important approach for quantifying survivability is coming from authors in [16], who propose to base quantification, on system's reaction to specific attacks and vulnerabilities modelled by attack graph. The attack graph represents the nodes that the attacker may exploit, while the way chosen to transverse these nodes in order to cover all possible system functionality states is forward-search, breadth-first and depth-limited.

To conclude, what may be observed is that most approaches on quantifying survivability are based on measuring availability and robustness characteristics of the system. Though, survivability is a more complex attribute that the system as a whole should emerge and should be based on the ability of the system to continue serving critical services. As a result, the approach proposed in this paper for evaluating survivability, is a testing framework focussing on testing services available against systems failures, attacks or accidents.

2.4. Survivability and Telecommunication Systems

Before concentrating on the proposed SDLC for mobile telecommunication systems, we conclude the current literature review with a brief presentation of a few representative approaches for designing and implementing a survivable telecommunication system. It becomes clear that all these approaches are focussing on outages and path failures and not on service failures as survivability preserves.

In [17], the authors investigate the impact of possible failure scenarios and possible survivability strategies to contend with spatial and temporal network behaviour in mobile cellular networks. The failures for this paper are restricted to loss of BS, BSC-MSC or VLR. In [18], the authors analyse architectural principles for achieving minimization of services loss and service restoration through certain disaster recovery plans. The failure scenarios that are considered are central office switch fires, earthquakes, flooding, large-scale power outages, signalling network outages, fiber cuts, and terrorism. The result of these scenarios are outages to network devices for which the paper introduces a four-phase methodology to handle such cases. Another approach for providing survivability to Universal Mobile Telecommunication Systems (UMTS) networks is based on Markov chains, semi-Markov process, reliability block diagrams and Markov reward models [19].

What we may observe from these approaches is that the designs proposed are based on fault tolerance techniques and on how to mitigate the failure of network nodes. There are many other approaches in literature that indicate various techniques to handle the impact of the failure of a node or a link. Though, survivability is far more than that. Survivability should be part of every step of the SDLC. The current research focuses on providing survivability requirements for mobile telecommunication systems that should be taken into consideration during the requirements elicitation phase of the SDLC, and on how to validate the satisfaction of these requirements during the testing or development phases.

To sum up this literature review on survivability as a term and on approaches for providing survivability to a system the following requirements should be adopted:

- Survivability is a mission driven attribute which means that the mission of the system is what should survive at the end, and not the system itself. Additionally, the majority of approaches discriminate and mark system services at essential and non-essential services with the essential services being the ones that should survive, and perform at an acceptable level of QoS, when a system is under attack or failure.
- Threat against survivability is any failure that may affect its critical services. So, the system should be able to react to any failure even if the root cause is unknown.
- A system must be designed as survivable and for this to be succeeded, survivability requirements, based each time on system's nature, must be defined during requirements specification of every development life cycle. These requirements may be organized to 3Rs (recognition, resistance, recovery and adaptation methodology) Additionally, survivability requirements should be considered during all stages of system's development lifecycle and as part of the everyday work.
- For a system to be compliant with survivability requirements specification, a monitoring system that monitors and evaluates system's survivability is of vital importance. Additionally, if a monitoring system is available, the state of the system may be

known each time and preventive or corrective actions, like re-configuration or other system's self-healing processes, may be applied for providing survivability to the system, even when unplanned threats are realised.

- Finally, testing and evaluation of system's survivability should contain investigation of intrusion scenarios and failure incidents in order survivability requirements to be raised. This could be very useful if test driven development methodologies are used.

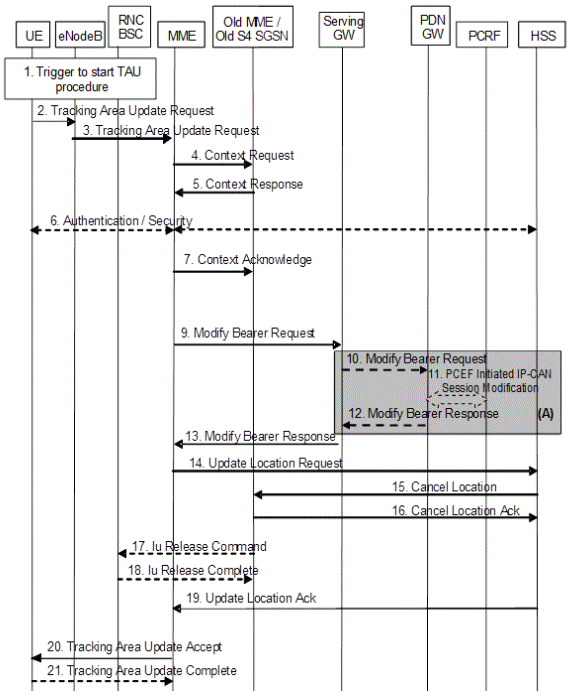
2.5. Mobile Telecommunication Systems

Before closing literature review, we will present some information on mobile telecommunication networks. Nowadays mobile telecommunication networks consist of a combination of 2G, 3G 4G and 5G mobile networks. Each network consists of the radio access network and the core network, which is finally connected to various networks like internet, IP Multimedia Subsystem (IMS) etc, to serve system's main mission which is to facilitate voice and data communications. Among network nodes, the communication in control-plane layer and user-plane layer is being established through specific interfaces.

Each of these systems has several nodes connected to each other. The particularity of mobile systems compared to other systems, like the internet, is that all services need an exchange of messages between a set of nodes to be established and performed. This significantly increases the risk of failure since problems may occur at any time during the exchange of the aforementioned messages. An example of such a message flow and possible failures that may occur, can be found in [20] or in the 3rd Generation Partnership Project (3GPP) standards. To continue with this logic, the network nodes that are connected to realize a service may be part of the same or a different network. For example, in 3G to 4G intersystem Tracking Area Update service, the nodes that may participate are from 4G, network nodes eNodeB, Mobility Management Entity (MME), Packet Gateway (P-GW), Serving Gateway (S-GW), Home Subscriber Server (HSS) and radio network controller (RNC), and Serving GPRS Support Node (SGSN) network nodes from the 3G network. This scenario is depicted in figure (1) bellow. Additionally, nodes may be manufactured from different organizations, a fact that increases the risk of interoperability failures. As a result, with various nodes interconnected, new networks are formed adding new system and survivability requirements that must be considered through the development of any new feature. The whole picture of a mobile network is shown in figure (2) bellow. This figure depicts the interconnection between 2G, 3G and 4G mobile networks through relevant interfaces. Though, the 5G network and the way it is connected with the rest of the mobile networks is missing. For this purpose, we utilize another picture from [21] that depicts the connection of the 5G network with the 4G network. This is figure 3 below.

The view of such interconnected systems adopted by the current work for all stages of the software development lifecycle is a multi-layered logic with the following levels:

- **Node level**: Any node of a mobile telecommunication network for which a new functionality or feature is to be developed. For example, MME should be considered to perform in node level.
- **System Level**: 2G, 3G, 4G and 5G, or any other that follow, are considered as systems. Nodes forming a system could be part of different PLMN operators. Any development task for a service that includes network nodes from the same system should be considered in system level.
- **Intersystem Level**: The entire telecommunication system may be considered as an intersystem. Nodes forming a network for serving an inter-system scenario may be considered as an intersystem. For example, in the scenario below, an Intersystem Tracking Area Update is depicted. The scenario includes nodes from 3G and 4G systems.



].

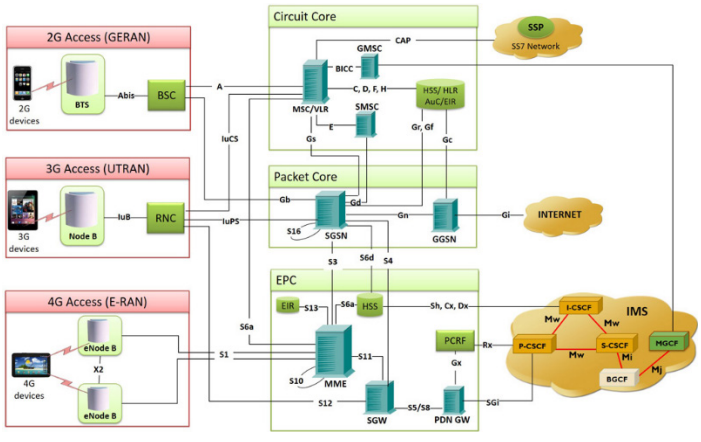


Figure 2: Common telecommunication network – The whole image (http://www.gl.com/telecom-test-solutions/communications-networking-2G-3G-4G-lab.html)

What is also important is that nodes supporting system or intersystem scenarios could even be part of different public switched telephone network (PLMN) operators. This means that when developing a new feature, the behaviour of nodes should not be considered as "known". Any possibility of receiving an unexpected message should be taken into consideration and the system should be able to resist to such a threat and recover from failure.

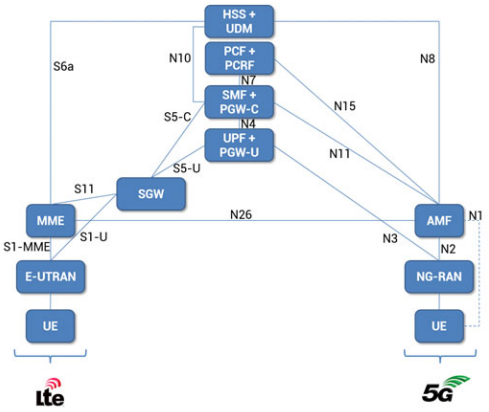


Figure 3: Common telecommunication network – 5G system added to the whole image (https://www.rfglobalnet.com/doc/g-core-network-architecture-network-functions-and-interworking-0001)

3. SDLC of Survivable Telecommunication Systems

Nowadays, systems development is mostly based on iterative models, or spiral models, in order to support continuous delivery of new functionality with certain predefined criteria. At the end of all iterations, an updated system, or a new release, is tested against its overall functionality in order to be delivered to the telecommunication operators.

Current research aims to improve this process by considering the survivability of critical services as the main requirement of the system under development. The main idea is to consider the whole (inter)system as a deliverable of any new release, instead of just focussing on a small part of the network. In this way, all survivability requirements at all system levels are considered and tested. The contribution of the current research is that it provides a complete proposal on how to handle survivability requirements and quality assurance of developed telecommunication system based on these requirements. The requirements are categorized to those related to service and those related to network since without it the system will not be available to perform any service. Additionally, the methodology proposed takes into consideration any arising requirement from the complicated interconnections of the telecommunication subsystems. All these requirements are gathered and grouped into 3Rs categories as described in literature; recognition, resistance, recovery and adaptation. In other words, requirements are enriched to include the whole network's survivability requirements. The result of not taking into consideration system and node inter-operability is a very important increase on the number of defects. Additionally, the testing methodology proposed by the current paper, considers all possible service failure scenarios and possible impact of any new functionality to the legacy code for critical services already developed.

The inputs to the aforementioned methodology are new features that will be developed or/and possible defects. When a new feature or a defect is planned to be developed, a new SDLC starts.

According to related literature, any methodology for designing survivable systems should start by defining the system's mission and the critical services that serve that mission. These should be documented and dealt with as requirements to any new functionality.

For mobile systems, critical are all services related to voice or data transmission from user perspective, and charging services from operator's perspective. This is also depicted in table (1) below, with service level requirements. So, for example, a voice bearer may be considered as critical service. A handover to such a bearer is critical also.

After definition of the mission and critical services that should survive, the general software development lifecycle (SDLC), is modified and used, with respect to special characteristics of the developed system, in such a way that at the end of the cycle the delivered (inter)system to emerge survivability. The SDLC that is proposed is depicted in figure (4).

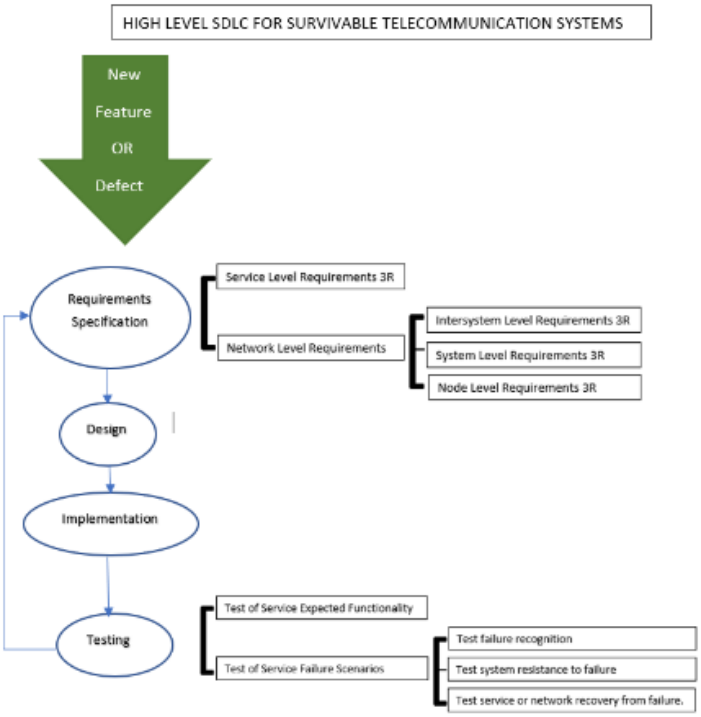


Figure 4: SDLC of Survivable Telecommunication Systems.

3.1. Requirement's specification

Requirements for extending the system's functionality are predefined and described in 3GPP documents. Survivability requirements should be based on a risk analysis study and detailed examination of the potential threats. As already explained, threats against survivability of the system are those that can directly affect the critical services of the system. This is the most effective way to protect critical services as such a service should survive even if the root cause of the failure is unknown. Thus, requirements are grouped to service level requirements that are related to services and network level requirements that are related to network

availability in order to support the operation of the services. For each group, requirements related to 3Rs (recognition, resistance, recovery, adaptation) methodology are presented.

In the tables below, high-level requirements related to survivability and defined by 3GPP are depicted. All these requirements are related to survivability and should be considered additionally to any requirement related to a new functionality or to any maintenance task. Additionally, any requirement that is an outcome of our research may also be depicted in service level survivability requirements table under columns titled "Our contribution". These requirements are related to **failure recognition** and **resistance** and are presented to previous papers [20], [22] related with survivability on telecommunication systems. Furthermore, the error handling requirements proposed from the current paper may be summarized to the following ones:

1. The system should be able to resist to failures related with **loss of messages**.

2. The system should be able to react to messages **arriving later or earlier** than expected. This should not have any impact to the service or to any other following services.

3. The system should be able to resist to failures related with **duplicate messages** sent to the nodes.

4. Any new functionality should be considered as a **threat to the critical services** already developed and any possible failure should be handled.

5. "**Hanging processes**" should also be considered as possible causes of failure.

Table 1: Service Level Requirements

3GPP Title	**3GPP Doc Num**	**3GPP Service Survivability Requirements related to failure Recognition**
UMTS Terrestrial Radio Access (UTRA) system	2101-301	*"*Set of attributes to describe UMTS bearer service (**delay variation tolerance, maximum transfer delay, maximum bit error rate**) information transfer rate attributes **(peak bit rate, mean bit rate, occupancy)**."* [23] *"*Performance: inherent transmission delay and level of traffic blocking"* [23]
Performance Management (PM)	32401	*"Data gathered through telecommunication management system are gathered to support performance evaluation on:* *- **Quality of Service** (e.g. delays during call set-up, packet throughput, etc) QoS can indicate the network performance expected to be experienced by the user."* [24]
Found across multiple 3GPP documents		
Error Causes **Please refer to certain interface 3GPP document for more details**		Specific error causes may be returned to the request message each time indicating a certain failure. For example, in GPRS Tunnelling Protocol (GTP) messages error cause "Mandatory IE incorrect" may be returned. From this the root cause of failure may be depicted and corrected by development team in case it can be corrected. Otherwise, there may be causes like "network failure" with root cause some failure to the network where all connections of the node with the node that returned this value, should be deleted.
Our contribution		
"Self-Diagnosis Framework for Mobile Network Services" [20]		Using the management reference mode of 32.101 we have proposed a self-diagnosis framework that may recognize and report different kinds of failure of service flow between nodes. Using this framework, the root cause of failure may also be depicted. Failures that have been analyzed are any possible failures that may occur when a message of a flow leaves a node to reach the neighboring node. The contribution of the paper is that focuses on diagnosis of service failure and not of system failure opposed to other proposals and to telecommunication management standard.
3GPP Title	**3GPP Doc Num**	**3GPP Service Survivability Requirements related to failure Resistance**
UMTS	2101-301	*"Handover should be transparent. In case of speech call loss of information may be tolerated but handover should be quick to **avoid connection break.** In case of data service temporary break is tolerable but not loss of information. Handover between terrestrial environments should be seamless within the same network"* [23]
		"Handovers should not increase the load on the fixed network significantly" [23]

		"The level of security should not be affected by handovers" [23]
		"Bearer services cannot be handed over between two environments if they are not supported in both. However, handover to an alternative bearer ***offering reduced capabilities*** *should be possible where this is supported by the service in use. The radio interface should have the capability to provide for handover and roaming between networks run by different operators"* [23]
Services and System Aspects;	22 101	*"Any handover required to maintain an active service while a user is mobile within the coverage area of a given network, shall be seamless from the user's perspective."* [25]
		"The 3GPP system shall be able to provide ***continuity between CS voice services*** *and the full duplex speech component of IMS multimedia telephony service with no negative impact upon the user's experience of the voice service. The same should be true for IMS Services."* [25]
		"The system shall support either *- transparent relay of the IP signaling and traffic;* *- service aware interconnection"* [25]
3G security; Security threats and requirements	21 133	*"Service Integrity: "It shall be possible to protect against unauthorized modification of user traffic"* *Service availability: It shall be possible to prevent intruders from restricting the availability of services by logical means"* [26]
Security Objectives and Principles	33 120	*"Security Objectives:* *1. to ensure that the security features standardized are compatible with world-wide availability* *2. to ensure that the security features are adequately standardized to ensure world-wide interoperability and* *roaming between different serving networks;"* [27]
Security architecture	33 401	The standard presents: *"- user identities confidentiality: MSIN, the IMEI, and the IMEISV should be confidentiality protected* *- user data signaling confidentiality: All S1 and X2 messages carried between RN and eNB shall be confidentiality-protected. Synchronization of the input parameters for integrity protection shall be ensured for the protocols involved in the integrity protection.* *- Integrity protection, and replay protection, shall be provided to NAS and RRC-signaling.* *- authentication and key agreement procedure between the mobile device and the core network,* *- security interworking of mobile networks (EUTRAN-UTRAN-GERAN)"* [28]
Technical Specification Group Services and System Aspects;	23 401	*"Authentication: NAS security mode control procedure is to take an EPS security context into use, and initialize and start NAS signaling security between the UE and the MME with the corresponding EPS NAS keys and EPS security algorithms"* [21]
5G; Security architecture and procedures for 5G System	33.501	The standard presents: *"- network access security: enable a UE to authenticate and access services via the network securely, including the 3GPP access and on-3GPP access, and in particularly, to protect against attacks on the (radio) interfaces* *-network domain security secure exchange of signaling and user plane data between networks.* *- User domain security: user access to mobile equipment.* *- Application domain security: enable applications in the user domain and in the provider domain to exchange messages securely"* [29]
		As it is presented to the current standard part of network life-cycle includes: *"the PLMN network is being adjusted to meet the long-term requirements of the network operator and the customer, e.g. with regard to performance, capacity and customer satisfaction through the enhancement of the network or equipment up-grade"* [29]
Found across multiple 3GPP documents		

Error Handiling		Some error causes indicate failures that can be handled in order to avoid dropping the service. Sometimes these handlings may be found across 3GPP documents or there may be implementation specific approaches that each organization implements during development of the device. To the example above "Mandatory IE incorrect" if we assume that the mandatory IE that is not correct is bearer ID. And the message causing this error is an answer to a previous message, then we may conclude which is the correct bearer id and ignore the error instead of dropping the service. The same may happen with network errors if we use relocation through selection functions to relocate the service that may be dropped in case it is critical (voice bearer for example)
Collision Handling		Collision is the case where two messages requesting a service arrive at a network and at the same time or one request arrives before the whole process of messages of the previous one has been completed. Then a handing of these requests should take place. This handling may be for example to serve both requests by a priority sequence, or to drop one of the two. For example, in case a request arrives for a UE that is already in process of a handover there is no meaning in processing it since the UE will leave from current Tracking area. Though there are cases that the service should continue to the Tracking area the UE will move to.
Our conrtibution		
"Fault Prediction Model for Node Selection Function of Mobile Networks" [22]		Our proposal regarding service resistance to failure is the fault prediction model proposed. This model takes into consideration DPMO (Defects per million opportunities) value which is a value that may be used to evaluate the operational performance of a node against 6sigma value. Then this value is used as a parameter in selection algorithm of mobile systems. This function is used to select a node which will be used to successfully complete a service flow.
Error Handiling		Apart from error causes defined by 3GPP documents and robust measurements that should be developed in order such cases to be handled, here we introduce some other error handline requirements: 1. The system should be able to resist to failures related to loss of messages. The failure should be ignored if this is possible. For example, if an acknowledgement message has not arrived, the service could be considered as established to avoid dropping it. If it could not be ignored, then the system should consider if there is a failure of neighboring node. In this case, the node should inform network management system and release any connection associated with this node. 2. The system should be able to react to messages arriving later or earlier than expected. This should not have any impact to the service or to any other following services. 3. The system should be able to resist to failures related with duplicate messages sent to the nodes. 4. Any new functionality should be considered as a threat to the **critical services** already developed and any possible failure should be handled.
Hanging Processes		As "hanging processes" we mean a service that fails, and leaves resources reserved causing failure to future services. For example, if a PDN Connection fails to be released and it is found as "already established" when a new PDN Connection is requested. This PDN Connection may be a critical service like voice bearer.
3GPP Title	**3GPP Doc Num**	**3GPP Service Survivability Requirements related to service Recovery from failure and adaptation.**
Restoration procedures	23 007	*"The data stored in location registers are automatically updated in normal operation; the main information stored in a location register defines the location of each mobile station and the subscriber data required to handle traffic for each mobile subscriber. The loss or corruption of these data will seriously degrade the service offered to mobile subscribers; it is therefore necessary to define procedures to limit the effects of failure of a location register, and to restore the location register data automatically"* [30]

Services and Systems Aspects;	22 101	*"The voice call continuity user's experience shall be such that, to the greatest degree possible, a consistency of service is provided regardless of the underlying communication infrastructure and technology"* [25]
UMTS	2101-301	*"Flexibility: Negotiation of bearer service attributes (bearer type, bit rate, delay, BER, up/down link symmetry, protection including none or unequal protection), parallel bearer services (service mix), real-time / non-real-time communication modes, adaptation of bearer service bit rate"* [23]
		"UTRA should adapt flexibly into changes and should have the capability to serve a variety of traffic densities (up to very high densities) and a variety of traffic mixes in an economical way." [23]
		"Flexibility and dynamic reconfiguration: minimum set of bearer capabilities, operating modes and features to ensure that inter-operability is always possible; continuity of operation during dynamic updating of terminal capabilities." [23]
Self-Organizing Networks (SON); Self-healing concepts and requirements	32541	*"In the case of software faults, the recovery actions may be :* *a) system initializations (at different levels),* *b) reload of a backup of software,* *c) activation of a fallback software load,* *d) download of a software unit,* *e) reconfiguration, etc.* *In the case of hardware faults, the recovery actions depend on the existence and type of redundant (i.e. back-up) resources."* [31] *'[If the faulty resource has no redundancy, the recovery actions may be:* *a) Isolate and remove the faulty resource from service so that it does not disturb other working resources;* *b) Remove the physical and functional resources (if any) from the service, which are dependent on the faulty one. This prevents the propagation of the fault effects to other fault-free resources;* *c) State management related activities for the faulty resource and other affected/dependent resources;* *d) Reset the faulty resource;"* [31] *e) Other reconfiguration actions, etc.* *"If the faulty resource has redundancy, the recovery action shall be changeover, which includes the action a), c) and d) above and a specific recovery sequence. The detail of the specific recovery sequence is out of the scope of the present document"* [31]

Table 2: Network Level Requirements

3GPP Title	**3GPP Doc Num**	**3GPP Network Survivability Requirements related to failure Recognition**		
		Node Level	**System Level**	**Intersystem Level**
Telecommunication management; Principles and high-level requirements	32.101	*"Telecommunication management system consists of an architectural framework or management reference model, that is used to collect measurements for management functions. Some of which are related to survivability like performance management, fraud management, fault management, security management, etc. With the use of performance measurements, configuration of system due to load needs may be executed. Additionally, for fault management, alarms or events may also imply a needed re-configuration for avoiding failures. Failure may be detected; isolated and root cause may be depicted."* [32]		

Performance Management (PM)	32401	*"Data sent at node level are gathered through telecommunication management system to support performance evaluation on:* *- traffic levels within the network, including the level of both the user traffic and the signaling traffic* *- verification of the network configuration: evaluation of effectiveness of changes of network plan related to traffic levels.* *- resource access measurements* *- resource availability (e.g. the recording of begin and end times of service unavailability)*" [24]	*"Network Operators are informed of PM - related events through alarms and may act accordingly."* [24]	
Fault Management;	32.111-1	"*If the faulty resource has no redundancy, the recovery actions shall be:* *- Generate and forward appropriate notifications to inform the **OS** about all the changes performed*." [33]		
3GPP Title	**3GPP Doc Num**	**3GPP Network Survivability Requirements related to system Recovery from failure and adaptation**		
		Node Level	**System Level**	**Intersystem Level**
Restoration procedures	23 007	*"The data stored in location registers are automatically updated in normal operation; the main information stored in a location register defines the location of each mobile station and the subscriber data required to handle traffic for each mobile subscriber. The loss or corruption of these data will seriously degrade the service offered to mobile subscribers; it is therefore necessary to define procedures to limit the effects of failure of a location register, and to restore the location register data automatically. The document describes data restoration procedures for VLR, HLR, HSS, GGSN, SGSN, MME. Triggering point is receiving a request for unknown IMSI in cases when the failing node has not detected the failure or receiving a message with restoration indicator set to not confirmed. These indicators show data corruption and procedure for restoring of these data through message exchange follows."* [30]		
		"Node restart. If a node restarts it sends a reset indicator to the neighboring nodes. Upon receiving such an indicator, the neighboring node shall inform its neighbors about the failure and release and re-initiate any PDN connection associated with failing node." [30]		

Fault Management	**32.111-1**	*"After a fault has been detected and the replaceable faulty units have been identified, some management functions are necessary in order to perform system recovery and/or restoration, either automatically by the NE and/or the EM, or manually by the operator. If the faulty resource has no redundancy, the recovery actions shall be:* *a) Isolate and remove from service the faulty resource so that it cannot disturb other working resources;* *b) Remove from service the physical and functional resources (if any) which are dependent on the faulty one. This prevents the propagation of the fault effects to other fault-free resources;* *c) State management related activities for the faulty resource and other affected/dependent resources."* [33]		
Self-Organizing Networks (SON); Self-healing concepts and requirements	32541	*"In the case of **software faults**, the recovery actions may be :* *a) system initializations (at different levels),* *b) reload of a backup of software,* *c) activation of a fallback software load,* *d) download of a software unit,* *e) reconfiguration, etc.* *In the case of **hardware faults**, the same as line of fault management above plus this:* *a) **Reset** the faulty resource;* *b) Other reconfiguration actions**, etc.* *If the faulty resource has redundancy, the recovery action shall be changeover.* ***Here we see that reconfiguration is something proposed by 3GPP but not a "must have" attribute."* [31]		

3GPP Title	**3GPP Doc Num**	**3GPP Network Survivability Requirements related to failure Resistance**		
		Node Level	**System Level**	**Intersystem Level**
(UMTS); protocol description and error handling	25.921	*"The error handling shall be specified in the protocol for the cases when the requirement for presence or absence of an IE indicated by the condition is not followed."* [34]		
Technical Specification Group Services and System Aspects;	23401	*"SGW-MME / SGW-PGW* ***GTP-C Load Control*** *feature is an optional feature which allows a GTP control plane node to send its Load Control Information to a peer GTP control plane node which the receiving GTP control plane peer node uses to augment existing GW selection procedure"* [21]	***"APN level load control*** *may be supported and activated in the network. If this feature is activated, the PDN GW may convey the Load Control Information at APN level (reflecting the operating status of the resources at the APN level), besides at node level."* [21]	
		"SGW-MME / SGW-PGW ***GTP-C Overload Contro****l feature is an optional feature. Nodes using GTP control plane signaling may support communication of Overload Control Information in order to mitigate overload situation for the overloaded node through actions taken by the peer node(s)"* [21]	***"NAS Level Congestion control****: The MME may detect the NAS signaling congestion associated with the APN and start and stop performing the APN based congestion control based on criteria: (max number of EPS bearers and EPS bearer activation per APN, one or multiple PDN GWs of an APN are not reachable or indicated congestion to the MME, Maximum rate of MM signaling requests associated with the devices with a particular subscribed APN, Setting in network management)"* [21]	
		"MME-Enb The ***MME Load Balancing*** *functionality permits UEs that are entering into an MME Pool Area to be directed to an appropriate MME in a manner that achieves load balancing between MMEs".* [21]	*"PDN* ***GW control of overload*** *by rejection of PDN connection requests from UE."* [21]	
		"*MME-Enb The* ***MME Load Re-balancing*** *functionality permits UEs that are registered on an MME (within an MME Pool Area) to be moved to another MME*" [21]		

		"MME *The MME shall contain mechanisms for avoiding and handling overload situations"* [21]		
		"SGW-MME ***Throttling of Downlink Data Notification*** *Requests. MME may restrict the signaling load that its SGWs are generating on it, if configured to do so."* [21]		
		"MME-UE *UE Level* ***NAS congestion****: The MME may detect the NAS signaling congestion associated with the UEs belonging to a particular group. The MME may start and stop performing the group specific NAS level congestion control based on criteria (maximum rate of MM and SM signaling requests associated with the devices of a particular group, Setting in network management)"* [21]		
Configuration Management (CM);	32.600	*"Configuration Management (CM), in general, provides the operator with the ability to assure correct and effective operation of the PLMN network as it evolves. CM actions have the objective to control and monitor the actual configuration on the Network Elements (NEs) and network resources, and they may be initiated by the operator or by functions in the Operations Systems (OSs) or NEs. CM actions may be requested as part of an implementation program (e.g. additions and deletions), as part of an optimization program (e.g. modifications), and to* ***maintain the overall Quality of Service (QoS)****. The CM actions are initiated either as single actions on single NEs of the PLMN network, or as part of a complex procedure involving actions on many resources/objects in one or several NEs."* [35]		

3.2. Design and Implementation

After requirements specification, design and implementation phases follow which are not worth analysing further since they are organization specific. **Robust and secure code design** techniques should be part of this phase. Additionally, **risks related to survivability** should be part of risk assessment which is usually conducted through the design phase.

3.3. Testing or Evaluation of System's Survivability

To continue, the testing phase of the proposed SDLC is presented. Testing is the way to evaluate a system's survivability. Testing phase should also follow the same model and test cases should be designed for node, system and intersystem level. In this way the whole system will be tested each time. Additionally, test cases should include tests against services' correct functionality, and they should be extended to also test any resistance, recognition or recovery survivability requirement to all testing levels (node, system, intersystem). For this to be achieved test-driven development is the most appropriate approach. Modern SDLC approaches are test-driven which is what is also proposed for the current SDLC.

Test-driven means that the tests are designed according to the requirements and are constructed even before the development of new features or maintenance tasks like bug fixing. Additionally, through this work we propose another approach that is related to test-driven development and has to do with failure impact evaluation. In other words, testing may be also used to evaluate the impact of any failure to critical services, and having this information available, new tasks may be extracted for the next iteration cycle regarding failure recognition, resistance or recovery. So, in this case tests are indeed driving the development and are a tool to discover many issues that may occur from any combination of services. So, any time a new service is to be

developed or updated, testing any possible combination of it with critical services will reveal any threats to critical services from the newly inserted code.

Impact analysis could be applied in any iteration of SDLC providing new requirements related to survivability requirements. Tests related to impact analysis may be:

1. Executing critical services before and after newly developed or modified service.

2. Executing critical services after failure of newly developed or modified service.

3. Executing critical services in collision with newly developed or modified service.

Additionally, another proposal is to test all survivability requirements for each new or modified functionality. So apart from just testing failure scenarios, recognition of failure and recovery from failure or resistance to failure should be also tested in order testing procedure to be considered complete.

All tests related to survivability evaluation and corresponding test approaches that could be used, are depicted in the following table (4) below. Test scenarios are also related to corresponding threat to survivability and impact of realization of this theat. Finally, any test case should be added to regression testing in order to ensure that future changes will not affect the existing functionality.

Table 3: Evaluation of Survivability Requirements through testing.

Survivability Threats		**Root Cause of Failure**	**Failure Impact in Node Level**	**Test Scenarios**	**Testing Methods**	**Examples from 4G network**
3GPP Fault Management; 32.111-1 Categories of faults for which an NE (network element) may	**Hardware failures**, i.e. the malfunction of some physical resource within a NE.	Device damage	Messages sent from one node to neighboring node may not be answered.	Testing of scenario where device is forced out - no information of the event to management system. The NE should be able to track the issue and report to management system. The impact from this failure to service under development and the restoration time should be defined.	Functional testing	Unplug of the device.
				Testing of scenario where device fails and sends alarm to management system. Service under development should be released or served by alternative resources after system re-configuration.	Functional testing	Enforcement of the NE to send a failure alarm to the management system
		CPU / Memory Overload		Testing of scenarios that		

raise alarms are:		System missconfiguration	Messages sent from one node to neighboring node may not be answered or answered with delay.	the message is not answered from neighboring NE in all phases of service establishment and test requirements related to handing of this situation.	Functional testing Unit or Module Testing Static Analysis,	Test scenarios where message is not answered.
			Faulty messages may arrive to NEs.	Testing of scenarios that the message arrives with wrong configuration information.	Functional testing Unit or Module testing Static Analysis Fuzzy-testing Fault-injection testing	Test message with wrong information about MMEs capability of supporting IOT devices.
	Software problems, e.g. software bugs, database inconsistencies	Any S/W bug that results in wrong functionality of service or non-compliance with standards	Service rejection or faulty service establishment.	Test-driven development with tests that are designed due to 3GPP standards requirements.	Functional testing Unit or Module Testing Static Analysis Fault-injection testing	Test all scenarios that reflect 3GPP requirements.
		S/W Bug lead to hanging processes	Future service requests may be rejected.	Enforce processes to be hanged and see if system reacts according to requirements. Test critical services impact if attempted.	Functional testing Unit or Module Testing Static Analysis,	Test if after deletion of a voice bearer it can re-established.
3GPP Fault Management; 32.111-1		Missing of robustness measurements like handling collision scenarios or handling of wrong Information Elements in messages	Service rejection or faulty service establishment.	Testing of all possible collision combination, especially with critical services, and test scenarios during which messages have wrong IEs that could be	Functional testing Unit or Module Testing Static Analysis,	PDN connection consists of a series of messages. A test case could include the modification of bearer id to a wrong one and see

				handled by robustness measurements.		if system is robust enough to handle this error.
		S/W bug that may lead to unanswered messages	Service rejection.	Testing of scenarios where messages of process under development are not answered.	Functional testing Unit or Module Testing Static Analysis,	A test case could be the PDN establishment and testing if service is properly rejected.
				Test the impact to critical services. Test cases with critical services already established and the above scenario following should be tested. The opposite is also valid scenario and should be tested. In this case failures from hanging processes will also be tested.	Functional testing Unit or Module Testing Static Analysis,	Testing of the above scenario after and before voice bearer handover.
		S/W bug that may lead to message sent twice	Service may be re-established if there is no mechanism for ignoring repeated messages	Testing of scenarios where messages of service under development are sent twice.	Functional testing Unit or Module Testing Static Analysis,	A test case could be sending PDN request twice for the same bearer.
	Functional faults, i.e. a failure of some functional resource in a NE and no hardware component can be found responsible for the problem.	Any other failure that may lead to service unavailability	Service Failure	Any related test	Functional testing Unit or Module Testing Static Analysis,	Any related test

	Loss of some or all of the NE's specified capability due to overload situations.	System Overload of requests	Messages sent from one node to neighboring node may not be answered or answered with delay.	Testing of service impact after increasing system load. Testing service impact after increasing load of service.	Stress testing Load testing Stability testing	Try to establish a voice bearer in a loaded system and an overloaded system. And try to see the impact to the system and voice bearer when system is loaded by voice bearer requests.
	Communication failures between two NEs, or between NE and OS, or between two OSs.	S/W failures, H/W failure, Overload situations, Path / Link failures, Network timing issues.	Messages sent from one node to neighboring node may not be answered or answered with delay.	Testing of scenarios of scenarios that the message is not answered from neighboring NE in all phases of service establishment and test requirements related to handing of this situation.	Functional testing Unit or Module Testing Static Analysis,	Testing of scenarios of scenarios that the message is not answered from neighboring NE.
Security Testing	Any security threat should be considered and tested. Details on security testing will not be provided to current document.					
Failure Recognition	In all possible errors, network management system should be tested. Network management system should be informed about any kind failure and should be able to trigger system resistance or recovery mechanisms. So, any NE that is under development should be tested against this functionality also.					
System Recovery	In all possible failure scenarios, recovery mechanisms following should also be tested.					

4. Conclusions and Future Work

To sum up, during the current paper, a development framework of a survivable mobile telecommunication system, based on system's mission and critical services, has been presented and proposed. This framework was based on the available survivability approaches through literature review with its main contribution to be that it provides a solution that is more focussed on interconnection and interoperation of systems forming larger intersystem. By this any survivability requirement from any level of service is considered through everyday development work and the focus is not only based on correct system functionality. Additionally, by this any interoperability and interconnection requirements and threats related to survivability may be examined through development life cycle.

Contrary to other approaches for evaluation of survivability, the one proposed is a more practical guide for testing the critical services of systems and evaluating measurements correlated to survivability of (inter)system, end to end from the requirements specification phase of the system and it does not only focus on node or link failure as most of proposals of literature review. This approach has been adopted because survivability is a built-on and not an add-on characteristic.

To sum up, the major outcomes of the current research are:

- The current research improves the traditional SDLC process, by enriching requirements analysis and testing phases with approaches related to survivability. The resulting proposed methodology is the Survivability Software Development Lifecycle presented in Chapter 3 that may be applied to telecommunication systems.
- The current research provides a systematic approach for handling the complexity arising from the interconnection of different network nodes of a telecommunication system.

Finally, as future work, we are planning to apply the proposed methodology in order gather and analyse metrics related to overall system survivability.

Conflict of Interest

The authors declare no conflict of interest

Acknowledgment

This work has been partly supported by the University of Piraeus Research Center.

References

[1] R.J. Ellison, D.A. Fisher, R.C. Linger, H.F. Lipson, T.A. Longstaff, N.R. Mead, "Survivable network systems: An emerging discipline", Carnegie Mellon University Digital Library, 1997, doi:10.21236/ada341963

[2] R.C. Linger, et al., "Life-Cycle models for survivable systems", Carnegie Mellon University Digital Library, 2002, doi: 10.1184/R1/6575138.v1

[3] V. R. Westmark, "A definition for information system survivability", in 37th Annual Hawaii International Conference on System Sciences, 10-19, 2004 doi: 10.1109/HICSS.2004.1265710.

[4] R.J. Ellison, "Survivability analysis framework", Carnegie Mellon University Digital Library. 2010, doi: 10.1184/R1/6584474.v1

[5] J. C. Knight, E. A. Strunk and K. J. Sullivan, "Towards a rigorous definition of information system survivability" in DARPA Information Survivability Conference and Exposition, 78-89, 2003, doi: 10.1109/DISCEX.2003.1194874.

[6] P. Pal, "Survival by defense – enabling" , in 2001 workshop on New security paradigms, 71–78, 2001, doi: 10.1145/508171.508183

[7] J. Knight, Dennis Heimbigner , Alexander Wolf, Antoinio Carzaniga, Jonathan Hill, Premkumar Devanbu, Michael Gertz, "The WILLOW survivability architecture", in Fourth Information Survivability Workshop, 2001, doi: 10.1.1.96.5316

[8] W. Li, L. Shu and Y. Feng, "A Dynamic Survivability reconfiguration framework based on QoS", in 2009 International Conference on Advanced Computer Control, 103-106, 2009, doi: 10.1109/ICACC.2009.107.

[9] D. Chen et al., "Network survivability performance evaluation: a quantitative approach with applications in wireless ad-hoc networks", in 5th International Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems, 61-68, 2002, doi: 10.1145/570758.570769.

[10] A. H. Wang, S. Yan and P. Liu, "A semi-Markov survivability evaluation model for intrusion tolerant database systems," in 2010 International Conference on Availability, Reliability and Security, 104-111, 2010, doi: 10.1109/ARES.2010.90..

[11] A.J. Mohammad, "Towards quantifying metrics for resilient and survivable networks", in 14th IEEE International Conference on Network Protocols (ICNP 2006), 2006, doi: 10.1.1.1.6900

[12] L.Y. Trivedi, "A general framework for network survivability quantification.", in 12th GI/ITG Conference on Measuring, Modeling, and Evaluation of Computer and Communication Systems, 369-378, 2004, doi: 10.1.1.94.3404

[13] M. Liang et al., "A novel method for survivability test based on end nodes in large scale network", KSII Transactions On Internet Ans Information Systems 9(2), 620-636, 2015, doi:10.3837/tiis.2015.02.008.

[14] C. Wang et al., "A general framework for network survivability testing and evaluation", Journal of Networks **6**(6), 831-841, 2004, doi: 10.4304/jnw.6.6.831-841

[15] M. Liang et al., "Research on survivability metrics based on survivable process of network system", in 4th international conference on security of information and networks, 247-250, 2011, doi: 10.1145/2070425.2070470 .

[16] L. Zhang, W. Wang, L. Guo, W. Yang and Y. Yang, "A survivability quantitative analysis model for network system based on attack graph", in 2007 International Conference on Machine Learning and Cybernetics, 3211-3216, 2007, doi: 10.1109/ICMLC.2007.4370701.

[17] D.W. Tipper et al., "Survivability analysis for mobile cellular networks", in Communication Networks and Distributed Systems Modeling and Simulation Conference, 2731-2738, 2002, doi: 10.1.1.361.411

[18] M. C. Baker, C. A. Witschorik, J. C. Tuch, W. Hagey-Espie and V. B. Mendiratta, "Architectures and disaster recovery strategies for survivable telecommunications services", Bell Labs Technical Journal, **9**(2), 125-145, 2004, doi: 10.1002/bltj.20030.

[19] S. Dharmaraja, V. Jindal and U. Varshney, "Reliability and survivability analysis for UMTS networks: an analytical approach", IEEE Transactions on Network and Service Management, **5**(3), 132-142, 2008, doi: 10.1109/TNSM.2009.031101. 8

[20] M. Mykoniati et al., "Self-Diagnosis framework for mobile network services", JACN, **7**(2), 2019, doi: 10.18178/JACN.2019.7.2.268

[21] "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access", Version 16, Technical specification 3GPP org The Mobile Broadband Standard 23401, 2021.

[22] M. Mykoniati et al., Lambrinoudakis, "Fault prediction model for node selection function of mobile networks", in 9th International Conference on Information Communication and Management ICICM 2019, 153-159, 2019, doi: doi.org/10.1145/3357419.3357452.

[23] "Requirements for the UMTS Terrestrial Radio Access (UTRA) system", Version 3.0.1, Technical specification 3GPP org The Mobile Broadband Standard, 21.01U, 1997

[24] "Telecommunication management;Performance Management (PM);Concept and requirements", Version 16.0.0., Technical specification 3GPP org The Mobile Broadband Standard 32401, 2020

[25] "Service aspects; Service principle", Version 18.1.1, Technical specification 3GPP org The Mobile Broadband Standard 22101, 2021

[26] "3G security; Security threats and requirements", Version 4.1.0, Technical specification 3GPP org The Mobile Broadband Standard 21133, 2002

[27] "Security objectives and principles", Version 4.0.0, Technical specification 3GPP org The Mobile Broadband Standard 33120, 2001

[28] "System Architecture Evolution (SAE); Security architecture", Version 16.3.0, Technical specification 3GPP org The Mobile Broadband Standard 33.401, 2020

[29] "Security architecture and procedures for 5G System", Version 17.2.0, Technical specification 3GPP org The Mobile Broadband Standard 33.501, 2021

[30] "Restoration procedures", Version 17.1.0, Technical specification 3GPP org The Mobile Broadband Standard 23.007, 2021

[31] "Telecommunication management; Self-Organizing Networks (SON); Self-healing concepts and requirements", Version 16.0.0, Technical specification 3GPP org The Mobile Broadband Standard 23.541, 2020

[32] "Telecommunication management; Principles and high level requirements", Version 16.0.0, Technical specification 3GPP org The Mobile Broadband Standard 32101, 2020

[33] "Telecommunication management; Fault Management; Part 1: 3G fault management requirements", Version 16.0.0, Technical specification 3GPP org The Mobile Broadband Standard 32111-1, 2020

[34] "Guidelines and principles for protocol description and error handling", Version 7.0.0, Technical specification 3GPP org The Mobile Broadband Standard 25.921, 2007

[35] "Telecommunication management; Configuration Management (CM); Concept and high-level requirements", Version 16.0.0, Technical specification 3GPP org The Mobile Broadband Standard 32600, 2020

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 278-291 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

An Alternative Approach for Thai Automatic Speech Recognition Based on the CNN-based Keyword Spotting with Real-World Application

Kanjanapan Sukvichai*, Chaitat Utintu

Department of Electrical Engineering, Faculty of Engineering, Kasetsart University, 10900, Thailand

ARTICLE INFO

Article history:
Received: 02 April, 2021
Accepted: 26 July, 2021
Online: 03 August, 2021

Keywords:
Thai ASR
MFCC
KWS
CNNs

ABSTRACT

An automatic speech recognition (ASR) is a key technology for preventing an ongoing global coronavirus epidemic. Due to the limited corpus database and the morphological diversity of the Thai language, Thai speech recognition is still difficult. In this research, the automatic speech recognition model was built differently from the traditional Thai NLP systems by using an alternative approach based on the keyword spotting (KWS) method using the Mel-frequency cepstral coefficient (MFCC) and convolutional neural network (CNN). MFCC was used in the speech feature extraction process which could convert the voice input signals into the voice feature images. Keywords on these images could then be treated as ordinary objects in the object detection domain. The YOLOv3, which is the popular CNN object detector, was proposed to localize and classify Thai keywords. The keyword spotting method was applied to categorize the Thai spontaneous spoken sentence based on the detected keywords. In order to find out the proposed technique's performance, real-world tests were carried out with three connected airport tasks. The Tiny-YOLOv3 showed the comparative results with the standard YOLOv3, thus our method could be implemented on the low-resource platform with low latency and a small memory footprint.

1. Introduction

Automatic speech recognition (ASR) is still an intriguing and demanding subject among researchers around the world. It is the mechanism that enables human-machine interaction through speech orders. In the present of the Coronavirus outbreak, this technology could decrease the propagation of the infection because no physical interaction such as touchscreen is required. Moreover, the automatic speech recognition system could also provide the hand-free experience that produces the huge advantage among the handicapped and visually impaired people. This manuscript is an extension of work that was initially presented in the 2021 Second International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP) [1]. In Thailand, according to the computers and the Thai language research [2], the first development of Thai ASR system concerned about isolated word recognition, which could be used for short commands [3]. The other improved method, which might be employed for continuous voice in little jobs, such as e-mail access or telephone banking, was also offered. More researches were publishing on huge vocabulary continuous speech recognition, or LVCSR [4], [5]. Inadequate speech databases and Thai-language complexity such as letter-to-sound mapping, phoneme set selection, segmentation, and tonality, were the key problems when developed a Thai automatic speech recognition system as mentioned in [6]. Nevertheless, many Thai organizations put effort in developing Thai speech corpus such as the LOTUS corpus by NECTEC [7]. For the standard automatic speech recognition system in natural language processing (NLP), it consisted of three basic models: Acoustic model, Lexicon, and Language model. For Thai language, the limitation of available large speech database and the complexity of Thai language could contribute to delayed speech recognition improvement [2].

In an effort to solve the problem, the automatic speech recognition was constructed using the different approach. The proposed method used the keyword spotting methodology to recognize the Thai sentence linked to the categorized keywords instead of using the traditional ASR system. This approach could ignore the presence of unpredictable words in the utterances such as unprecedented words and non-speech sounds. In such a case,

*Corresponding Author: Kanjanapan Sukvichai, Faculty of Engineering, Kasetsart University, 10900, Thailand, : +662-797-0999, fengkpsc@ku.ac.th

https://dx.doi.org/10.25046/aj060431

the typical method must include the large vocabularies, as well as grammars.

More details about the keyword spotting were clearly mentioned in the following section. For investigating more corresponding Thai research, the use of acoustic modeling based on Hidden Markov Models (HMMs) method [8], [9]. The models consisted of filler models, syllable models, and keyword models, which used for handling out-of-vocabulary sound elements in the speech recognition using the keyword spotting algorithm. Another paper proposed the nondestructive determination of maturity of the Monthong Durian based on Mel-Frequency Ceptral Coefficients (MFCCs) and Neural Network. Then we have further investigated more excellent works that related to ours like SpeechYOLO paper [10], this paper used first version of YOLO with keyword spotting algorithm for localizing boundaries of utterances within input signal by considering audio as objects and use Short-Time Fourier Transform (STFT) technique to extract the audio feature. In our research we have identified a similar approach in order to solve the limitations in Thai speech recognition field, but we use the improved version of YOLO which was YOLOv3 and we also use Mel-Frequency Cepstral Coefficients (MFCC) technique which could artificially implement the behavior of the human auditory perception due to the fact that it is capable of using the non-linear Mel scale which relies on the non-linear frequency scale of the real human.

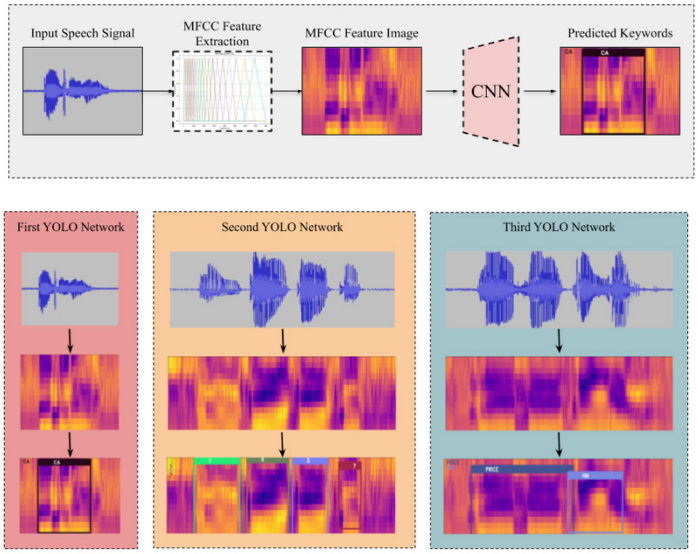


Figure 1: **Top:** the core methodology pipeline composed of the Mel-Frequency Cepstral Coefficient (MFCC) as the feature extraction method and the YOLOv3 as the popular convolutional neural network (CNN) based object detector which was then applied to localize and classify keywords. **Bottom:** our proposed method for the real-world application, which consisted of three separated YOLO networks including airline names, numbers (0-9), and frequently asked questions (FAQs). Both full and tiny version of YOLOv3 were proposed.

This manuscript started with explaining how Keyword Spotting (KWS) to clarify the speaker propose and also investigated the KWS prior works. Then, the system pipeline will be explained which started from the speech feature analysis extraction method. This process was crucial since it could represent the voice signal information in the form of a feature image. Consequently, in the next section, the convolutional neural network was explained including the history of object detection techniques, the novel CNN-based object detectors, and the major improvement in YOLO family. After that, the voice dataset collection is described. Then, the highlights of our core methodology are explained. This manuscript also explained about the YOLO annotation and network training. Finally, the proposed method consisted of three networks was experimented. These three networks were responsible for three real-world tasks that would then introduce in the following section. In the last section of this manuscript, the comparison between the regular and the light-weight version of YOLOv3 performances was discussed for the additional usage in the low-resource platform. The overall design of our work was displayed in Figure 1.

2. Keyword Spotting for the Spoken Sentence

In general, most people tend to struggle to understand the spoken sentences especially with the native speakers. Unlike a written sentence, the spoken sentence is more difficult and variety in terms of speed, accent, style, or tone depending on the morphological richness of each language such as Thai language as previously mentioned. The most reasonable answer why the spoken sentence is difficult is that we don't experience enough vocabularies. Many scientists also say that listening requires more vocabulary knowledge than reading. In addition, it is difficult to tell whether the sentence contains unknown terms or is quickly spoken. Eventually, the spoken words are not always finished or carefully delivered with the native speakers like syllable stressing or skipping some words. Just like the human, the machine must transcribe the entire sentence with a huge model of vocabulary. To address this issue, the keywords are employed to involve the main idea rather than to use every single word in the sentence. This idea was motivated by the scanning strategies used in the reading comprehension test. Many skillful readers used this technique to reduce time to search for some specific information by just looking for the related words instead of reading the whole paragraph line by line. This scanning technique also seemed to help the system to come across the complexity and the variety of the sentence structures.

For the ASR system, there were certain technologies known as wake word, the particular word or phrase which would enable some additional functions. These voice-related systems, such as "Hey Siri" by Apple's Siri or "OK Google" by Google Now, exploited the keyword spotting (KWS) method which would listen until the specific keywords could be recognized [11]. The prior work for KWS introduced the Keyword/Filler Hidden Markov Model (HMM) [12]. Each keyword was trained with the HMM model, then the non-keyword segments, so-called fillers, was trained separately with a filler model HMM. Due to the decoding need of Viterbi, this strategy ended with considerable computational complexity. Then, the following enhanced approaches concerned the use of the discriminative model on the keyword spotting task, namely, large-margin formulation [13] and recurrent neural networks [14]. Although it could produce many significant improvements over the traditional HMM method, the high computational time was still occurred due to the entire utterance processing in order to locate an optimal area of keywords. These limitations were still challenging until Google proposed the novel deep learning keyword spotting approach base on the deep neural network, also called DeepKWS. Since this method could provide high detection performance with smaller

memory consumption and shorter runtime computation, then it was appropriate for the mobile device usage. Unlike the HMM approach, this process did not require the excessive sequence search algorithm. After that, Google explored a small-footprint keyword spotting (KWS) system using the Convolutional Neural Networks (CNNs) [15]. The CNN approach could outperform the DeepKWS in many ways. For instance, it provided more robustness over different speaking styles, reduced model size and more practical for spectral representation input. Thus, in this research, we chose the CNN-based object detection for localizing and classify Thai keywords from the spoken sentence.

3. Speech Feature Analysis and Extraction

Automatic Speech Recognition (ASR) system is influenced by feature analysis and extraction, since the high-quality features provide an acceptable and dependable results in the localization and classification process. The main purpose of feature extraction is to reveal acoustic information in terms of a sequence of feature vectors, which can successfully characterize a given speech data. Several feature extraction techniques are available to extract the parametric representation of the audio signal, such as perceptual linear prediction (PLP), linear prediction coding (LPC) and Mel-frequency Cepstral Coefficients (MFCC). MFCC has been proven to be the most prevalent and powerful technique [16], [17].

Mel-Frequency Cepstral Coefficients (MFCC) is the transformation from the speech waveform to frequency domain mathematical features, which are considered to be much more accurate than time domain features. This technique artificially implements the behavior of the human auditory perception due to the fact that it uses the non-linear Mel scale which relies on the non-linear frequency scale of the real human, so it results in the parametrically resemblances between the extracted vectors and human sense of hearing. The output of the MFCC is a short-term power spectrum coefficient of a windowed signal produced from the original signal Fast Fourier Transform (FFT). This transformation is done using the logarithmic power spectrum through a linear transformation, Discrete Cosine Transform (DCT). MFCC for speech can reveal more compact spectrum since Mel-scale coefficients are countable.

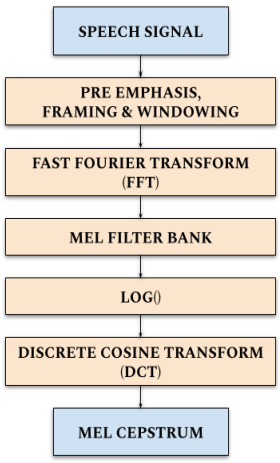


Figure 2: The overall MFCC derivation

In detail, the overall MFCC feature extraction technique is shown in Figure 2. It consists of windowing the speech signal, applying the Fast Fourier Transform (FFT), taking the log of the magnitude, warping the frequencies on a Mel scale, applying the Discrete Cosine Transform (DCT). In order to remove the MFCC from the speech signal, pre-emphasis begins. Compensate filtering for the high frequency area disappearing during the mechanism of voice creation are considered in the pre-emphasis. In addition, the significant of the high-frequency component is also strengthened. This step is therefore followed by the high-pass filter as explained in (1).

$$S_y(n) = S(n) - \alpha S(n-1) \tag{1}$$

where,

$S(n)$ is the input signal.

$S_y(n)$ is the output signal.

α is a control slope of the filter ranged from 0.9 to 1.0

The z-transform of the filter is defined as (2).

$$H(z) = 1 - \alpha z^{-1} \tag{2}$$

After the pre-emphasis process, the spectrum of the signal is balanced and some glottal effects from the vocal tract parameters are removed. Then the speech signal needs to be analyzed over a short period of time by capturing into a discrete frame, with some overlapping between frames. The purpose of the overlapping analysis is to approximately center each speech signal at some frame and avoid significant information loss. On each frame, the Hamming window is multiplied individually to maintain the continuity of the first and the last points in a frame. If the signal in a frame is denoted by $S(n)$ then the signal after windowing is applied will be $S(n)*w(n,\alpha)$ where $w(n,\alpha)$ represented the Hamming windowing defined as (3)

$$w(n,\alpha) = (1-\alpha) - \alpha\cos\left(\frac{2\pi n}{N-1}\right) \tag{3}$$

where,

$w(n,\alpha)$ is the Hamming window.

N is the total number of samples in a frame.

n is a sample number.

Each short-time frame was then transformed into the spectral features by applying the Fast Fourier Transform (FFT), which was the accelerated version of the Discrete Fourier Transform (DFT). This technique converted each frame of N samples from the time domain into frequency domain and was shown in (4).

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-\frac{(2\pi nk)j}{N}} \tag{4}$$

where,

N is the total number of points in the FFT computation.

Next, the Fourier transformed output was passed through a set of band-pass filters, so called Logarithmic Mel-filter bank, in order to cover from a real frequency, estimated in the Hertz unit,

f m, by the equation (5). The Mel scale was about a linear frequency spacing of less than 1 kHz and a logarithmic spacing of more than 1 kHz, respectively.

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \tag{5}$$

The frequency of Mel-scale is proportional to the linear frequency logarithm. Due to the fact that the behavior of human's aural perception is non-linear, this concept can be implemented using Mel-scale filter bank, which is commonly the combination of the K triangular filters. The example of a filter bank with K=20 is illustrated in Figure 3. The higher frequency filters contain more bandwidth than the lower frequency filters, but they share similar temporal constraints. Each triangular filter is centered with a maximum amplitude of 1 and is reduced linearly to zero until it approaches the central frequency of the two neighboring filters, where zero response is present and can be derived as (6).

$$F_m(k) = \begin{cases} \dfrac{k - f(m-1)}{f(m) - f(m-1)}, & f(m-1) < k < f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)}, & f(m) < k < f(m+1) \\ 0, & \text{Otherwise} \end{cases} \tag{6}$$

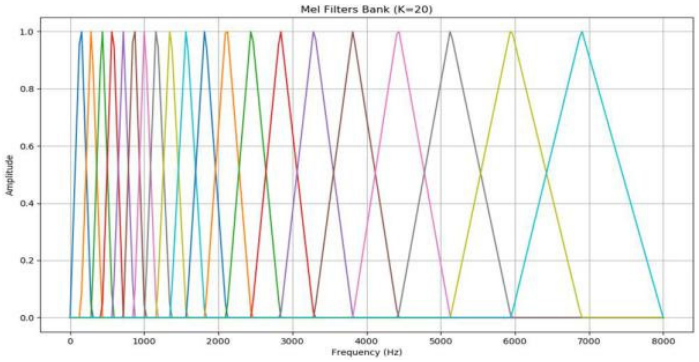


Figure 3: Mel filters bank when K=20

where,

$F_m(k)$ is the filter number m at k frequency.

m is the filter number in Mel filters bank.

Therefore, the Mel spectrum of the frequency spectrum $X(k)$ was calculated by multiplying the spectrum by each of the triangular Mel weighting filters as shown in (7).

$$S(m) = \sum_{k=0}^{N-1}\left(\left|X(k)\right|^2 F_m(k)\right) \tag{7}$$

where,

M is the total number of triangular Mel weighting filters.

The Discrete Cosine Transformation (DCT) was then applied on the transformed Mel frequency coefficients in order to produce a set of twelve cepstral coefficients. Since this algorithm, which was described in (8), resulted in a signal with a quefrency peak in the time-like domain, so called cepstral domain. Thus, Mel-Frequency Cepstral Coefficients, or MFCC, was nominated from the final features which were similar to cepstrum. The zeroth coefficient was excluded because it contained the unreliable information.

$$C(n) = \sum_{m=0}^{M-1} \log_{10}\left(S(m)\right) \cos\left(\frac{\pi n (m - 0.5)}{M}\right) \tag{8}$$

where,

$S(m)$ is the frequency coefficient.

$C(n)$ is the cepstral coefficient.

L is the number of MFCCs.

M is the number of triangular bandpass filters.

$n = 0, 1, \ldots, L$

Following this phase, a speech signal in the form of an MFCC might be employed to perform machine learning technology or treated as an ordinary image. The MFCC for each speech frame was stacked and converted into an RGB image with Plasma color map. Figure 4 shows examples of MFCC's Thai keyword images which include check-in, price and what time is it respectively.

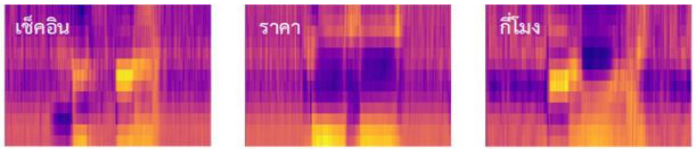


Figure 4: The examples of Thai keyword MFCCs

4. Convolutional Neural Networks for Voice Image

The sequence of the cepstral vectors or MFCC characteristics was received from the recorded voice as an input as previously discussed in the last section. The next step is to derive the corresponding keywords. In this section, computer vision technique for the Natural Language Processing (NLP) task is detailed. In order to resolve this challenge, we opted to employ a state-of-the-art object detection algorithm because it can be used to estimate the location (object localization) and category (object classification) of each object in a given image. Therefore, this technique should be able to extract the important sort of information for better semantic understanding of images [18]. From the preceding progress, as a result of considering MFCC features as the normal image, then keywords can be comparative to objects. Thus, the object detection is well-suited and also helpful for solving this situation as shown in Figure 5. The object detection technique that we chose in our work is the novel deep learning approach based on Convolutional neural network (CNN) [18][19]. CNN is a specialized type of neural network model modified especially for dealing with two-dimensional input data.

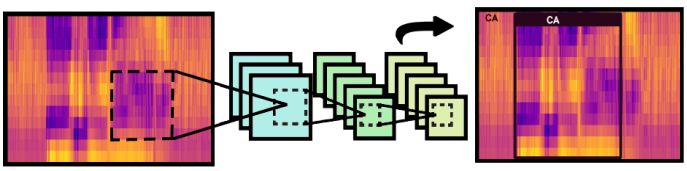


Figure 5: Object detection on MFCCs.

It is typically composed of three fundamental types of layer consisting of convolutional, pooling and fully connected layers. Convolutional and pooling layers contain filters, whose depth increases from left to right while width and height have the inverse proportion, have an interactive role in extracting features. The Convolutional layer is denominated from the important linear mathematical operation in the layer, so-called convolution, or filtering. The convolution function is described in (9).

$$S(i,j) = \sum_m \sum_n I(m,n)F(i-m, j-n) \tag{9}$$

where,

I is the two-dimensional array such as Image.

F is filter or kernel.

$S(i,j)$ is the output or feature map.

This operation involves the dot product between the input array and the filter or kernel, by convolving across its width and height, then extended throughout its depth. Next, the result from convolution is passed through an activation unit called Rectified Linear Unit (ReLU) calculated by (10). It is a piecewise function that will return the input directly if the input value is positive, otherwise, it will return zero.

$$R(z) = \max(0, z) \tag{10}$$

Subsequently, it is down sampled or normalized by the pooling operation, e.g. max pooling, in the pooling layer. This operation summarizes the initial activation feature map to become more robust. After that, fully connected layers then map them into output via activation function, which normally uses the Softmax activation defined in (11). It provides the probability distribution for each neuron, in the same way as a traditional neural network.

$$\sigma(y)_i = \frac{e^{y_i}}{\sum_{j=1}^{n} e^{y_j}} \tag{11}$$

where,

y_i are the element of the input vector.

n is the total number of classes.

y is the input vector.

The learning parameters can be optimized via the Stochastic gradient descent (SGD) method and the back propagation, called the training, in each layer across the models. The algorithms aim to minimize the difference between prediction and ground truth. Therefore, in the training process, the dataset images are fed forward with initial kernels and weights, so-called forward propagation. Then a model's accuracy is calculated by a loss function. The error value is used for updating kernels and weights later in the back propagation using gradient descent optimization. The overview of convolutional neural network architecture and the parameter optimization process are illustrated in Figure 6.

The convolutional neural network has been shown to be outperformed to previous methodologies in many ways. Firstly, Spatial Hierarchical feature representation can be learned automatically and the multiple nonlinear mappings can reveal the hidden patterns. Then deeper architecture, which significantly increases the model competency, allows related task optimization. Finally, due to the uprising of CNN learning performance, some challenging computer vision problems might be solved from a new perspective. CNN method has also been popular among various research topics such as facial recognition [22] and pedestrian detection [23]. CNN-based object detection technique aims to identify the location and the class of every object existing in the input image. Nowadays, its frameworks can be divided into two main categories, which are two-stage and one-stage approaches. A two-stage model is a technique that is constructed based on the standard object detection pipeline. It proposes two separate consecutive methods which are region proposal and object classification. In detail, region proposals are generated first, then each proposal will be classified according to the object classes. The second methodology, which is a one-stage model, considers transforming the traditional object detection problem to the regression problem by merging all separated tasks, consisting of localization and classification, into a single unified model.

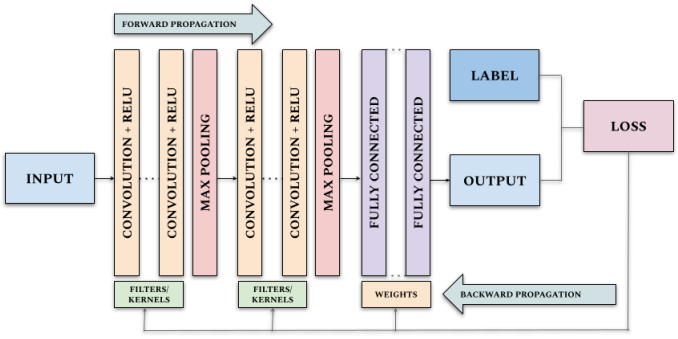


Figure 6: CNN architecture and the training process.

For the two-stage method, in [22], the author introduced the proposal of the regions with CNN features (R-CNN), which took a giant step towards object detection. It extracts a set of object candidate boxes or proposals by using a selective search method, which leads to major improvement in detection accuracy. However, this technique experiences many drawbacks. The enormous quantities of overlapped repetition boxes greatly delay the overall process and some feature losses can also be due to proposals in fixed candidate region. To solve the problems, in [23], the author proposed Spatial Pyramid Pooling Networks (SPPNet). The Spatial Pyramid Pooling (SPP) layer in SPPNet allows a CNN to construct a representation without considering the input image size. Although SPPNet utilizes less processing time than R-CNN since it requires only the single computing for the convolutional layer, the downside of this approach is the extended training period and the high use of disk space. The several advantages from R-CNN and SPPNet were investigated and integrated in the following work, Fast RCNN, but the detection speed is still bounded by the selective search [24]. Therefore, in [25], the author proposed Faster RCNN, which was known as the first ever end-to-end object detector that could almost yield real-time performance. It can break through the speed bottleneck of past efforts by demonstrating the use of the region

proposal networks (RPN) instead of the traditional selective search method. In addition, it uses a separate network to predict the region proposals only. The region proposal network is trained together with the model which results in predicting more accurate proposed regions. Later, there were some further improvement efforts on Faster RCNN, such as Feature Pyramid Networks (FPN) introduced in [26]. With the small object constraints, a basic image pyramid can be used to scale input images into multiple sizes before sending to the network. The FPN has enhanced the performance of multi-scale object detection and has become a model for numerous novel approaches.

On the one hand, the two-stage detector comes up with high localization and classification accuracy. On the other hand, its inference speed is impractical for the real-time applications, especially for the low-resource computational platforms like embedded systems, and the complex procedures of its core methodology could possibly diminish the opportunities of further optimization and improvement over components. The one-stage models have been built and researched in order to overcome these restrictions. It recognized objects based on two distinct neural networks, which lead to high computational time as indicated before in Faster RCNN section. Accordingly, the one-stage method combines feature extraction, region proposal and object classification altogether into a single network. The object classification is also performed regarding the predefined size and number of anchor boxes specified in the model configuration, then the regression concept is applied for the object localization process. Consequently, this novel breakthrough could implement the object detection as a single regression problem by predicting bounding box coordination and class probability simultaneously. The one-stage object detection pipeline used in our work is shown in Figure 7.

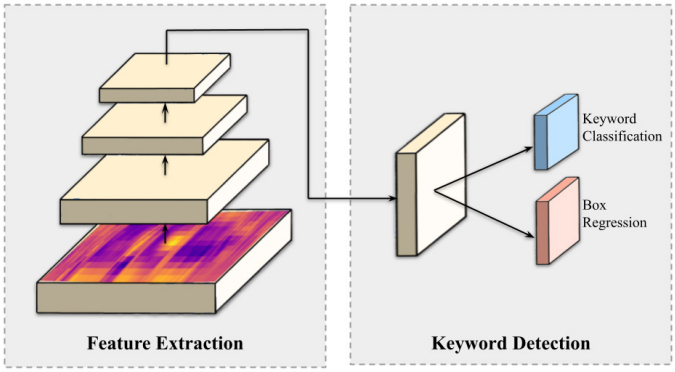


Figure 7: The basic architecture of one-stage object detector using MFCC voice image as the model input.

The one-stage object detectors were demonstrated to attract a great deal of attention amongst researchers not only because they can overcome the problem of a real-time bottleneck in the two-stage approaches, but because it could generate similar detection performance, it the detection speed and accuracy trade-off is well-optimized. Since our application is based on the Automatic Speech Recognition (ASR), both processing time and detection performance must be concerned. Therefore, we have chosen the state-of-the-art one-stage approach called YOLO, which is You Only Look Once. Based on real-time criteria, the computational time must be considered more than 10 frames per second (FPS) or less than 100 milliseconds. The performance evaluation results published in the YOLO article meet our requirements and stated that the third version of YOLO, or YOLOv3, outperforms other popular methods such as Single Shot Multibox Detector (SSD) [27], which had the original implementation on Caffe [28].

You Only Look Once (YOLO) is the prominent one-stage CNN-based object detector which we have chosen as a Thai keyword localizer and classifier in our research due to the high calculation speed and acceptable detection accuracy on the real-time task. YOLO frames object detection as a regression problem, thus it can simultaneously construct the object bounding boxes and predict class probabilities, while it needs only one forward propagation from the input image. The model name, You Only Look Once, might therefore be designated from such properties.

YOLO divides the input image into $S \times S$ grid. Then, each grid cell takes responsibility for detecting the object whenever its center approaches that particular grid cell boundary. Each grid cell will predict B bounding boxes and their predicted confidence scores. The bounding boxes can be represented by five parameters including x, y, w, h and the confidence scores. The parameters (x, y) indicate the center coordination of the box relative to the grid cell area. The next two parameters (w, h) describe the width and height of the box relative to overall image dimensions. The last parameter, the confidence scores can be derived from (12).

$$Pr(object) * IOU_{Pred}^{Truth} \tag{12}$$

where,

$Pr(object)$ tells how likely the object existence is.

IOU_{Pred}^{Truth} tells how accurate the predicted box is.

At that moment, each grid cell also predicts C conditional class probabilities, which is the conditional probabilities of the grid cell detecting an object, regardless of the number of boxes. The prediction is finally encoded as a $S \times S \times (\mathrm{B} \times 5 + \mathrm{C})$ tensor and the final confident scores can be described by Equation (13).

$$Pr(object) * IOU_{Pred}^{Truth} * Pr\left(class_i \middle| object\right) = Pr(class_i) * IOU_{Pred}^{Truth} \tag{13}$$

where,

$Pr\left(class_i \middle| object\right)$ is the conditional class probability.

During the model training step, the loss function is optimized and could be described in (14). From the following equation, we can notice that the loss function only penalizes the classification errors when an object exists in that corresponding grid cell, and likewise, the errors of the bounding box coordinate are penalized when a prediction is literally responsible for the ground truth.

$$Loss_{total} = Loss_{bonding\,box} + Loss_{confidence} + Loss_{classification} \tag{14}$$

where,

$$Loss_{bonding\ box} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} \left\{ \left(x_i - \hat{x}_i\right)^2 + \left(y_i - \hat{y}_i\right)^2 \right\} + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} \left\{ \left(\sqrt{w_i} - \sqrt{\hat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2 \right\}$$

$$Loss_{confidence} = \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} \left(C_i - \hat{C}_i\right)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} \left(C_i - \hat{C}_i\right)^2$$

$$Loss_{classification} = \sum_{i=0}^{S^2} I_{ij}^{obj} \sum_{c \in classes} \left(p_i(c) - \hat{p}_i(c)\right)^2$$

In the first version of YOLO [28], the network architecture was designed based on the GoogLeNet[29] architecture for image classification. YOLOv1 network consists of 24 convolutional layers and 2 fully connected layers. It replaced GoogLeNet's inception modules with 1 x 1 reduction layers and 3 x 3 convolutional layers respectively. The first 20 convolutional layers of the network were pretrained on the classic ImageNet 1000-class competition dataset with the input size of 244 x 244 to have 88% top-5 accuracy for a week. The PASCAL Visual Object Classes (VOC) in 2007 and 2012 were then used for training and validation. On this dataset, 98 bounding boxes per images are predicted. The weak predictions are then filtered out by the Non-maximum Suppression (NMS) technique. The overall training pipeline was originally implemented on the Darknet framework. YOLOv1 could achieve 63.4 mAP (mean average precision) and 45 FPS (frames per second). This achievement implicated that YOLO could perform real-time performance while the detection accuracy was as comparable as Faster R-CNN. To improve YOLO's detection speed, Fast YOLO was introduced [30]. It optimized the original YOLO architecture by decreasing the convolutional layers from 24 to 9 and using fewer filters in those layers. Therefore, it could reach up to 155 FPS, but the accuracy dropped to 52.7% mAP.

For considering the generalizability, YOLOv1 performed well on PASCAL VOC 2007 and it seemed to be the best method when evaluated on the artwork dataset which contained the challenging difficulty on a pixel level. Although YOLOv1 could achieve significant achievement over other methods, it also struggled with some limitations. Since the classification and localization network of this version of YOLO could detect only one object, any grid cell could detect one object too. This constraint resulted in the limit maximum number of objects, which was a total of 49 objects per one detection for 7x7 grid cells. As a result, it caused relatively high localization error especially the small objects that close to each other, such as groups of pedestrians. Moreover, this network architecture found difficulty in object generalization when evaluating on the other input dimensions instead of 244 x 244. To overcome these YOLOv1 constraints and achieve better performance, the second version of YOLO (YOLOv2) was then developed [31]. YOLOv2 concerned mainly about maintaining the prior classification performance while it also worked on improving recall and localization. Thus, it applied a variety of significant modifications to increase mean average precision (mAP). By adding batch normalization on the convolution layers, the model showed drastic impact on the convergence. In detail, the activations in the hidden layer were shifting and scaling which leaded to improving model stability and also reducing the overfitting issue. This technique could raise YOLO's mAP up to 2% and has been widely used for the model regularization propose. Consequently, YOLOv2 also proposed the high-resolution classifier from 224x244 in the first version to 448x448. Initially, the model was trained on the 224x224 input images, then the classification network was fine-tuned with 448x448 resolution on the ImageNet dataset for 10 epochs before detection training. This process gave a room for kernel adjustment on higher resolution images, which therefore resulted in better detection performance and reached slightly below 4% mAP improvement. Moreover, YOLOv2's researcher has introduced the remarkable progress, motivated by Faster R-CNN, which was the usage of anchor box. Instead of using fully-connected layers for bounding box prediction on the feature map, the convolutional layers were modified to predict locations of anchor boxes which leaded to recall increasing, but slightly degraded the overall mAP. The bounding box prediction could then relative to these anchor boxes, which was a small area consisted of width and height, unlike in YOLOv1 that relative to full image dimension. As mentioned in Faster R-CNN, the initial sizes of anchor boxes were chosen by hand-picking method. In contrast, YOLOv2 developer designed the suitable anchor box dimensions using k-mean clustering. The distance metric was calculated based on IoU scores, which is shown in (15).

$$\left\| x, c_i \right\| = 1 - IoU(x, c_i) \tag{15}$$

where,

x is a ground truth candidate box.

c_i is one of the centroids.

The YOLOv2 bounding box prediction spatially related to the offsets of the anchors. It predicted 5 parameters including tx, ty, tw, th and to, then applied the sigma function to limit the offset range of an output, which was calculated by (16).

$$\begin{aligned} b_x &= \sigma(t_x) + c_x \\ b_y &= \sigma(t_y) + c_y \\ b_w &= p_x e^{t_w} \\ b_h &= p_h e^{t_h} \end{aligned}$$

$$Pr(object) * IoU(b, object) = \sigma(t_0) \tag{16}$$

where,

t_x, t_y, t_w, t_h are the predicted offset and scale.

c_x, c_y are the top left corner of the anchor grid cell.

p_w, p_h are the size of the anchor box.

b_x, b_y, b_w, b_h are the predicted box center and size.

$\sigma(t_0)$ is the box confidence score.

These anchor boxes generated by clustering technique could provide higher average IoU. YOLOv2 divided the input into 13x13 grid cells by removing a pooling layer that was later more helpful with larger or smaller items. The goal was also to

overcome a small object limitation in YOLOv1. Furthermore, this version of YOLO model also integrated the multi-scale training strategy that could strengthen the model robustness to various input image sizes. Every 10 batches during training process, the network was resized according to the new random image dimension which was a multiple of 32 dues to convolutional layer down sampling property. Lastly, for reducing the computation time, YOLOv2 used a more light-weighted base model architecture, so called Darknet-19, with 19 convolutional layers and 5 layers of max-pooling. In addition, YOLOv2 also had the extended version which could detect more than 9000 classes, also known as YOLO9000 [31]. YOLO9000 was a slightly modified version of YOLOv2, which considered in merging small object detection dataset with large ImageNet dataset. Since the overlapping class labels of these two datasets could not combine directly, the WordTree hierarchical model was demonstrated for this difficulty. The general labels were placed in the top closer to the root and the fine-grained labels were branched similar to leaves. The probability of each class could calculate by following the path from that node to the root.

YOLOv3 was another improved version of YOLO family which outperformed its predecessor by using the state-of-the-art algorithms and also was practical with real-time scenarios [32]. Firstly, in order to predict the objectness confidence score, YOLOv3 proposed the logistic regression for predicting the confidence score for each bounding box, while the prior YOLOs used the sum of the square errors. Next, in some dataset like ImageNet, not all the labels were mutually exclusive. In such a case, choosing the Softmax function as the activation function of the output layer might not be the best choice, since it just converted predicted scores to distributed probabilities which summed up to one. For instead, YOLOv3 used multiple independent logistic classifier. The input image could then have multiple labels and contained both mutually and non-mutually exclusive objects. The complexity of the YOLOv3 model was also decreased by getting rid of the Softmax function. Consequently, with the new activation functions, the classification loss function must be modified into the binary cross-entropy instead of the Mean Square Error (MSE). Inspired by the idea of the Feature Pyramid Network (FPN), YOLOv3 could then predict the object up to three different box scales for every location in the input image and then perform feature extraction. This technique improved the detection performance on various object scales and was a significant solution especially for detecting small things. The Feature Pyramid Network design was illustrated in Figure 8. Furthermore, YOLOv3 also adopted cross-layer connections between two prediction layers and the fine-grained feature maps. Initially, the coarse-grained feature map was up-sampled, then the concatenation was performed to combine it with the previous one. Moreover, the YOLO's feature extractor was modified from Darknet-19 to Darknet-53, which based on the 53 convolutional layers. Darknet-53 was much deeper and composed of mostly 3x3 and 1x1 kernels including skip connections just like residual network in ResNet. In comparison, it has less BFLOP (billion floating point operations) than ResNet-152, but could still maintain the same detection accuracy at two-times faster. The full YOLOv3 architecture was illustrated in Figure 9.

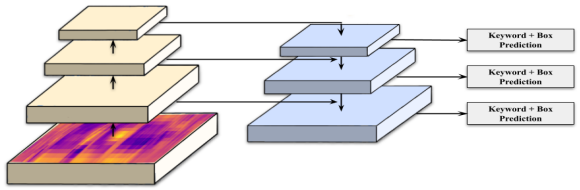


Figure 8: Feature Pyramid Network.

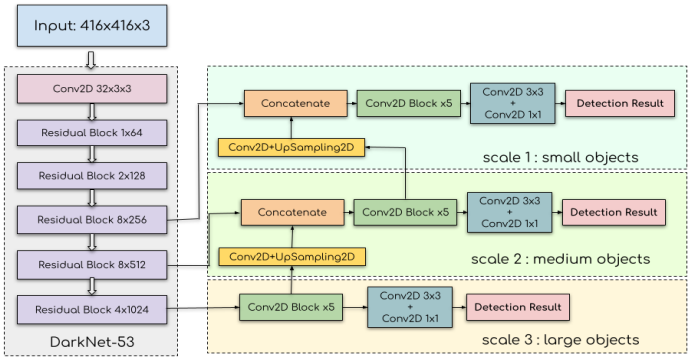


Figure 9: The YOLOv3 architecture.

YOLOv3 may therefore conduct high-detection capabilities for object detecion. In YOLOv3 paper, the researcher argued that YOLOv3 might exceed three times the speed of the Single Shot Multibox Detector (SSD). While RetinaNet was more precise, it also battled with the computational time. In small object criteria, YOLOv3 has also shown substantial progress.

YOLOv3 also had the lightweight version which was known as the tiny-YOLOv3 model. It used the same training strategy as the full version, however, the minor modifications were applied with the full YOLOv3 model architecture to gain a lot more computation speed and consume fewer computational resources for inference. This model was ideally optimized for the use of the low resource platform such as embedded systems and mobile devices. Firstly, the tiny-YOLOv3 simplified the number of convolutional layers. It used only 7 convolutional layers. This tiny version just employed a few numbers of 1x1 and 3x3 convolutional layers for extraction of features. It replaced a convolution layer with the pooling layer with a step size of 2 by focusing on summarized the complex dimension of the standard model. Although it used a slightly different architecture with the full YOLOv3, the model base structure was still the same which composed of 2D-convolutional, batch normalization and Leaky Rectified Linear Unit (ReLU).

In our research, the speech signal can now be handled as an ordinary image by transforming the recorded voice signal into the MFCC feature image. Then, Thai spoken keywords that contain with in the voice image could be compared as objects. Therefore, the object detection technique could be applied to an automatic speech recognition task. In order to understand the Thai speaker purpose, we applied this core methodology to find the corresponding Thai keywords in the voice image by feeding the whole sentence MFCC and matched them with our predefined sentence categories which each group of Thai keywords related to. Moreover, we propose that the localization ability of the object detector could be deployed especially with some input spoken sentence, which relies on the restricted keyword ordering such as

the unique number sequence in the flight number. As mentioned previously, YOLOv3 has been proven to be fast and accurate. It was the most stable and reliable version of YOLO since we started to research this work. We also compared the detection performance between the full-YOLOv3 and its tiny version. Since tiny-YOLOv3 was extremely fast, some mean average precision (mAP) was traded off. Our keyword detection system was illustrated in Figure 10.

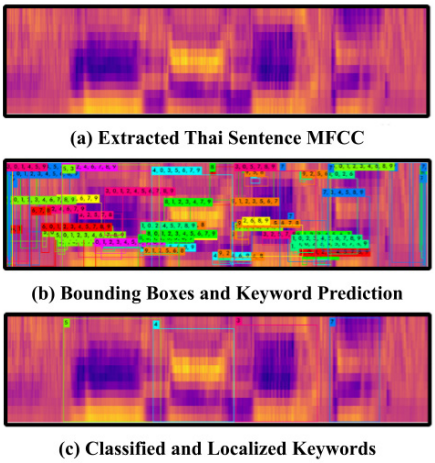
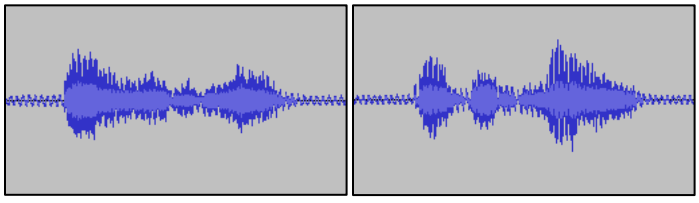


Figure 10: The keyword detection overview with Thai sentence MFCC feature image as an input.

5. Voice Dataset Collecting and Preprocessing

This article draws on speech datasets made by 60 native Thai speakers and recorded in several places proposed to simulate the variety of the background noises. Our selected participants were varied in ages and the genders were equally distributed. In terms of age, the highest number of overall participated subjects was in teenage (20s), which was 18 people (30%), and those aged between 40 and 49 were the smallest, which was 7 people (11.6%). The total datasets contained 2400 WAV files with 40 unique keywords. Table 1 summarized the baseline characteristic of our dataset. At the beginning, each recorded voice was processed using the Audacity software before converting into MFCC feature image, which was shown in Figure 11 and 12.



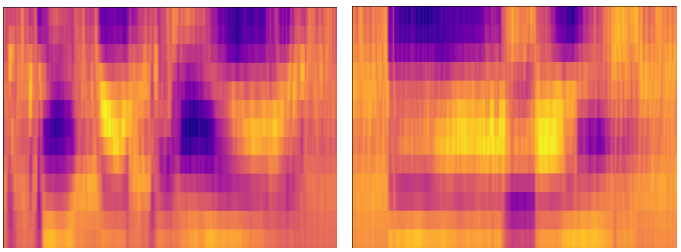
(a) (b)

Figure 11: The examples of recorded voice: (a) · (Air Asia); (b) (Thai Airways)

In order to avoid overfitting and improve the model robustness, some data augmentation methods including the tempo and pitch randomizing were applied to increase the quantity of training data, therefore finally finishing up with a total of 9600 WAV files. The datasets were obtained in accordance with the passenger information machine criteria for the research of the automatic speech recognition (ASR) system in actual settings. The automated system required to give the reliable information, which correlate to the passenger's flight number as a real-time answer to the passenger's frequently asked questions (FAQs). In the aviation industry, a flight number is composed of a two-character airline designator and numbers with a maximum length of four digits. Thus, we proposed three datasets consisted of airline names, numbers and FAQs keywords. The first dataset was selected from the ten popular airlines at the Suvarnabhumi airport in Thailand. Next, the second dataset consisted of ten numbers (0-9). The final dataset contained 20 keywords which could categorize into 7 passenger FAQ topics.

Table 1: The baseline characteristic of our dataset.

Characteristics	Training (80%)	Validation (20%)	Total
Population	48	12	60
Sex			
Male	24	6	30
Female	24	6	30
Age Group (Y)			
10-19	9	2	11
20-29	15	3	18
30-39	6	2	8
40-49	6	1	7
50-59	6	2	8
60-69	6	2	8
.WAV files			
Raw Dataset	1920	480	2400
Raw Dataset + Augmentation	7680	1920	9600



(a) (b)

Figure 12: The examples of extracted MFCC image: (a) · (Air Asia); (b) (Thai Airways).

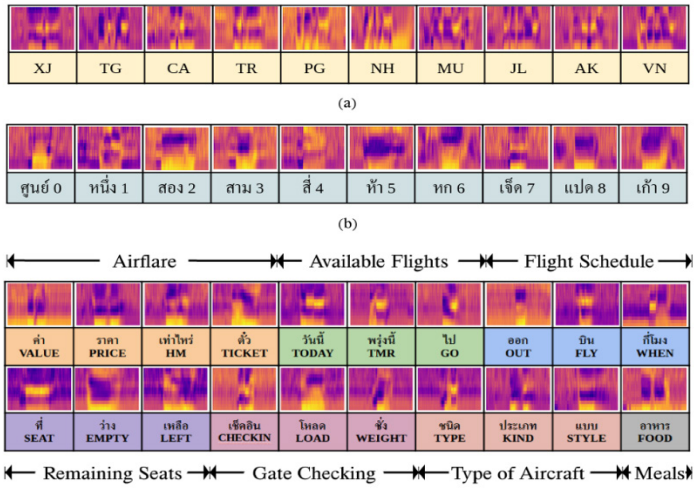
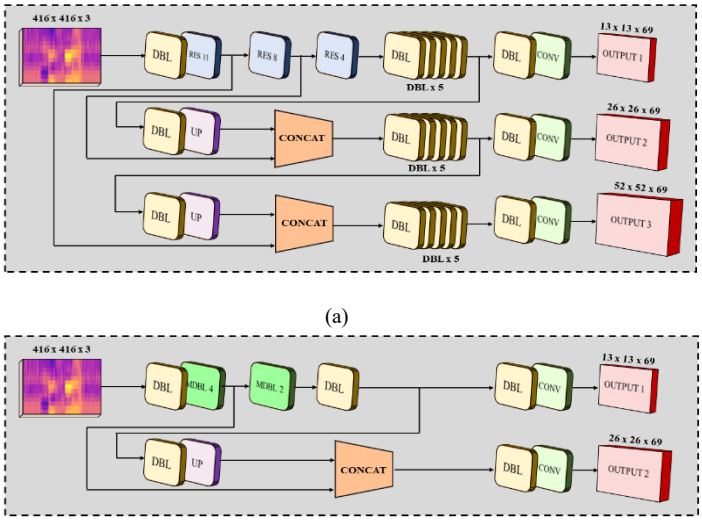


Figure 13: The overview of three datasets: (a) airline designators; (b) numbers 0-9; (c) FAQs keywords

Figure 13 showed the overview of three datasets. In every dataset, the images were randomly split with a ratio of 80:10:10, which included training, validation, and also test sets for further evaluation.

6. Highlights of the Research Work

In Thai linguistics, the absence of large speech corpora and the language morphology could inhibit development and researches in the speech recognition field as mentioned on the introduction section. Although several organizations like NEXTEC concentrated on eliminating this limitation, the complicity of the spontaneous voice input, especially in a noisy environment, seemed to be a problem for many researchers. Through the keyword spotting technique, we were interested in overcoming this problem. This technique could not only filter the sentence redundancies but also be easily maintained in data-scarce domains since we simply need only significantly keywords to train the deep learning model. In order to expand the keywords for more use cases, without retrained the large model like traditional approach, it could be done easily by create new lightweight model with small dataset and included into the latest model which would be clearly described in the experiment section.



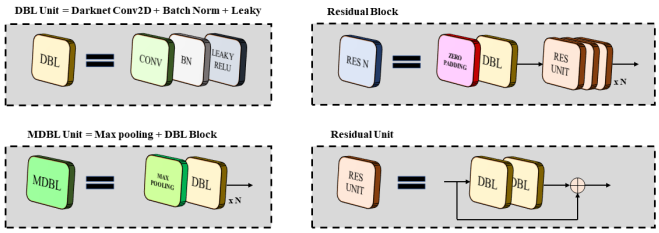


Figure 14: Model architectures: a) Full YOLOv3 architecture; b) Tiny YOLOv3 architecture; c) Legend

In more details, our manuscript has described how to develop the basic and lightweight ASR as the alternative approach of the classical ASR that might be effective in certain complex spoken sentences or languages like Thai, and could be employed in many scenarios where the real-time requirement is crucial. Our data collection is juts utilized to show the specific occurrences that can be applied to our proposed strategy. By applying computer vision technique to the linguistics field, the state-of-the-art one-stage object detector, You Only Look Once, is chosen and responsible for keyword classification and localization due to its detection performances. YOLOv3 also released the lightweight and optimized version of its full architecture as Tiny YOLOv3. Since this version of YOLOv3 is appropriate for performing the real-time computing on the resource-limited platform like edge devices, we also decided to compare this version of YOLOv3 on the same task with the regular YOLOv3. The differences between the full YOLOv3 and tiny YOLOv3 structures are described on Figure 14. The training strategy of our models are clearly explained on the following section.

7. Network Training Method

Since object detection was the supervised learning, we needed to provide the ground truth to the network. Therefore, before the training process, each target keyword on the MFCC feature images of the training set needed to be manually labeled. In this phase, the LabelImg tool has been used because the YOLO annotation format has already been provided. Consequently, the XML file according to each MFCC image would be generated to indicate the necessary labeling information for the network training step. In detail, the actual labeling steps were holding the mouse cursor, framing the target keyword region, and then double-clicking to identify the corresponding keyword class. The YOLO annotation format in XML file was composed of class index, the quotient of x-coordinate of the bounding box's center and the image width, the quotient of y-coordinate of the bounding box's center and the image height, the quotient of the bounding box width and the image width, and the quotient of the bounding box height and the image height. Due to the fact that the keyword in the MFCC feature image had the unique pattern, then it is possible to readily separate the specific keyword region from the uniform background noise region. The example of keyword labeling was described in Figure 15.

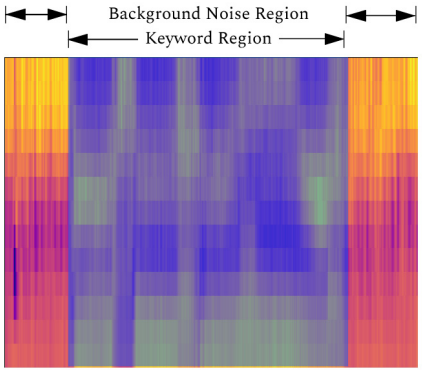


Figure 15: The example of keyword labeling.

The training platform in this article were carried out on the desktop PC (OS:Ubuntu 16.04LTS system, RAM: 16 GB, GPU: Nvidia GeForce GTX 1070Ti, CPU: Intel Xeon 8 cores). Both regular and tiny version of YOLOv3 were trained under the Darknet framework then the training results were analyzed to prove that Tiny YOLOv3 performance was acceptable for the proposed method. The stochastic gradient descent algorithm was used to perform the network training. In order to start training the model, we chose the pre-trained models of the ImageNet dataset as the initial convolutional weights. By considering the training durations, the maximum number of iterations was extended and the training process would be stopped whenever the loss was low enough. We set the Batch parameter to 64 and the Subdivision parameter to 16. As a result, the batch was split into 8 mini-batches which could not only reduce the memory usage, but also accelerate the training process and also generalized the model. Next, the learning rate was chosen to be 0.001. During training, the average loss and mean average precision (mAP) could be visualized in real-time using the graph as shown in Figure 16a. The final results were shown in Table 2.

Table 2: The models' performance of all datasets

Weights	Iterations	Loss	mean Average Precision (mAP)
Airline names			
Tiny YOLOv3	49000	0.12	0.99
Full YOLOv3	21400	0.005	0.99
Numbers (0-9)			
Tiny YOLOv3	50000	0.16	0.98
Full YOLOv3	34800	0.012	0.98
Frequently Asked Questions			
Tiny YOLOv3	287000	0.08	0.99
Full YOLOv3	41100	0.029	0.98

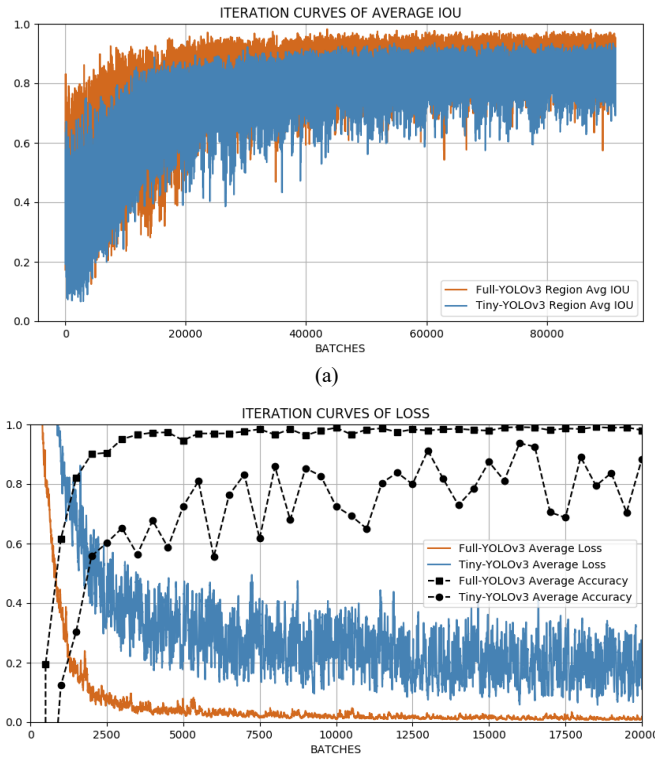


Figure 16: Iteration curves: (a) Iteration curves of loss; (b) Iteration curves of average IoU

As shown in Figure 16, although the fitting degree and the average IoU value of the full version of YOLOv3 was superior compared with the tiny version of YOLOv3, the light-weight version of YOLOv3 appeared to have well-enough training result and could be employed for further experimental evaluations.

8. Experiment and Analysis of Results

In order to investigate the true performance of the finally trained YOLOv3 weights, we considered using the unseen MFCC feature images extracted from the extra recorded voices of 20 individuals (10 males, 10 females) who didn't present in the prior training lists. This experiment was conducted to represent the airport use case of our proposed approach in real life. Whether planning to fly locally or internationally, an airport is always the first priority to asking for the flight information and useful navigation suggestion. From that point of view, we then proposed the autonomous information providing system that could respond to the frequently asked questions (FAQs) of the flight information from the passengers based on the flight number they told in real time. This service might provide passengers all the information they need on their flight only with their voice. This service is therefore straightforward and convenient to use, and additional services such as self-check-in can be integrated and the time and costs of congested and crowded airports can be reduced. For the first task, the automatic speech recognition system was used for detecting the airline names. Then, the 1-to-4-digit numbers would be recognized by the second task. The flight number was then confirmed by the combination of airline name and airline number. In Figure 17, we illustrated some examples of this task. Firstly, our proposed system expected the airline name which was "¢"ƒ‡¹‰" (kan – bin - thai) or Thai Airways (TG) on the upper

left of figure. Next, it would ask for airline number which was "Õç ÃÃ¬ ¡Ã" (haa - see - saam - jet) or 5-4-3-7 on the upper right of figure. The flight number, TG5437, was then generated by combining these two inputs which showed on the ticket.

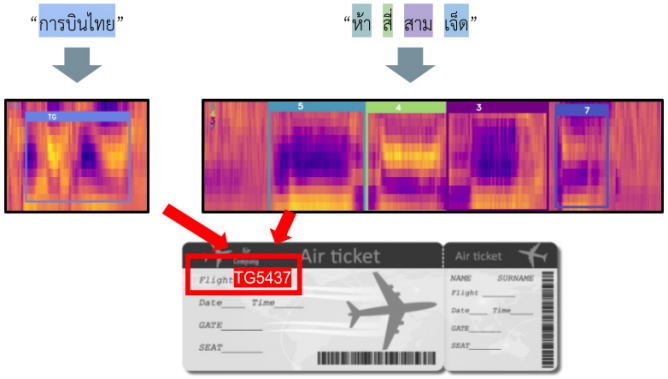


Figure 17: The example of the flight number recognition

Next, the third task was responsible for answering the frequently asked questions (FAQs) according to the provided passenger flight number. The third task example was described in Figure 18 respectively. In Figure 18, we presented some tasks from all seven the frequently asked questions (FAQs) categories of the flight information from the passengers. In the left side of figure, the input interrogative sentence "ราคาเท่าไหร่?" (la-ka-tao-lai) meant "how much is a ticket?". In this sentence, it contained two keywords that related to airflare category (described in Figure 13) which were ราคา (PRICE) and เท่าไหร่ (HM), then the system responded with all available ticket prices. Next, in the right side of figure, the input interrogative sentence "เช็คอินที่ไหน?" (check-in-tee-nai) meant "where is my check-in gate?". This sentence contained one keyword which was เช็คอิน (CHECKIN), thus the system knew that it related to gate checking question and performed the location guidance service.

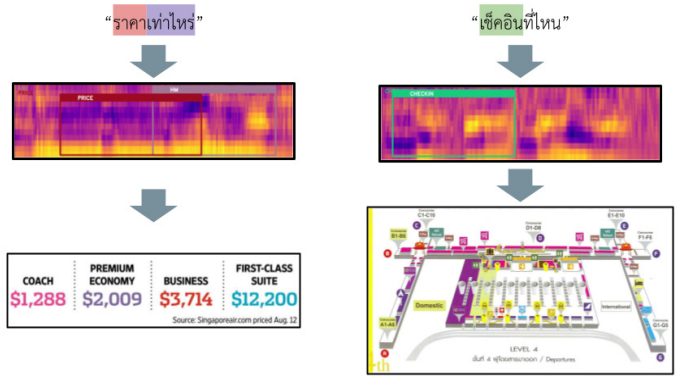


Figure 18: The example of the FAQs recognition

In this experiment, Precision, Recall and F1 score were used as the evaluation metrics. Precision is the proportion of the amount of correctly detected keywords over the entire number of detected keywords. Recall is the proportion of the amount of correctly detected keywords over the entire number of keywords in the experiment set. The F1 score is the harmonic mean of Precision and Recall at particular thresholds. The calculation formulas were shown in Equations (17)-(19).

$$Precision = \frac{TP}{TP + TF} \tag{17}$$

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{19}$$

Specifically, TP, or True Positive, indicates the amount of correctly detected keywords. TN, or True Negative, indicates the amount of correctly detected backgrounds. FP, or False Positive, indicates the number of incorrect detections. FN, or False Negative, indicates the number of missed detections.

The full and tiny version of YOLOv3 were trained and tested separately on three different tasks. Airline name detection was the simplest task and the FAQs recognition was the most challenging task. In detail, Airline name detection required just the one utterance, which might contain just one name, and so resulted in high precision and recall in both tiny and full YOLOv3 model. The next challenge was that of the airline number which required the sequence of 1-to-4-digit numbers, thus it could be seen that the accuracies of both models were slightly dropped. The last assignment was the most difficult, which the input sentence could be varied depending on the speaker experiences. To tackle this challenge, we used the keyword spotting idea to search for only the related terms we defined and categorized. As a result, it seemed that the detection performances of both networks have declined dramatically yet remained over 0.6. The detection threshold and IoU in this experiment were set at 0.20 and 0.50 respectively. The true positive, false positive, false negative, precision, recall, and F1-score results for all test samples were summarized in Table 3. The comparison of the F1-score of both full and tiny YOLOv3 on three different tasks could then be visualized in Figure 19. The test results showed that the first task was highly sensitive for both models, airline name detection, which was the easiest task and then experienced poor detection result on the last task, the FAQs recognition task, the hardest task as we mentioned previously. By compared with the full version of YOLOv3, the light-weight and simplified vision of YOLOv3, so-called tiny YOLOv3, could receive the comparable performance and could unusually outperform in terms of Recall by almost 0.1 on the second task. To clarify the poor performance of both models on the third task, the FAQs recognition task, two potential reasons were given as following. First, the number of classes was doubled in comparison with the first and second tasks and the non-keyword parts might be included in a spoken sentence. Second, the input MFCC feature dimension may vary by the length of the sentence and the words can be compressed when the input images have been scaled, leading to feature loss and false negative.

Table 3: Performances of both models on testing dataset

Weights	IoU = 0.50, Threshold = 0.20					
	TP	FP	FN	Precision	Recall	F1
Airline names						
Full YOLOv3	91	2	1	0.978	0.989	0.984
Tiny YOLOv3	90	6	0	0.938	1	0.968
Numbers (0-9)						
Full YOLOv3	192	39	70	0.831	0.733	0.779
Tiny YOLOv3	221	58	48	0.792	0.822	0.807
Frequently Asked Questions						
Full YOLOv3	121	39	83	0.756	0.593	0.664
Tiny YOLOv3	108	61	74	0.639	0.593	0.615

Based on the detection performances shown in Table 3, the average F1-score of the full and tiny YOLOv3 from three different tasks were 77.54% and 77.23% respectively. In terms of localization performance using the same box labels, the tiny YOLOv3 boxes seemed to have slightly shifted due to the simpler model architecture, which was shown in Figure 20, however, it was not affected the keyword detection proficiency. In conclusion, the experiment results showed that the tiny YOLOv3 could perform well enough and have competitive results with the full YOLOv3.

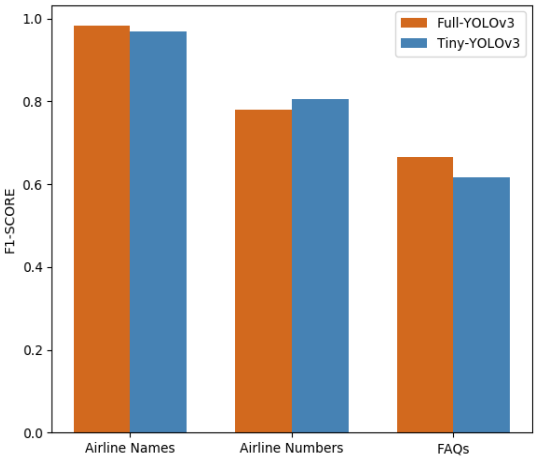
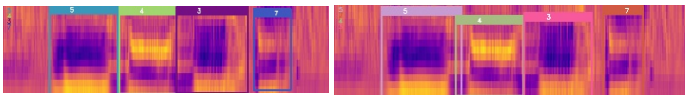


Figure 19: The comparison of three task detection results



(a) (b)

Figure 20: The comparison of the localization accuracy: a) Full YOLOv3; b) Tiny YOLOv3.

Furthermore, in order to test the model robustness, we injected the white noise with the relative amplitude α from 0.1 to 0.4 to each of the audio files using the Audacity software. In accordance with the white light which contains all wavelengths with equal intensity of the visible spectrum, white noise also contains the equally distributed energy in all audible frequencies resulted in a steady humming sound. The F1 score and accuracy measured after increasing the white noise intensity, which belonged to the regular and tiny YOLOv3 were shown in Figure 21, in which the y-axis was the measure and the x-axis was the intensity (α) of the white noise added to the audio files.

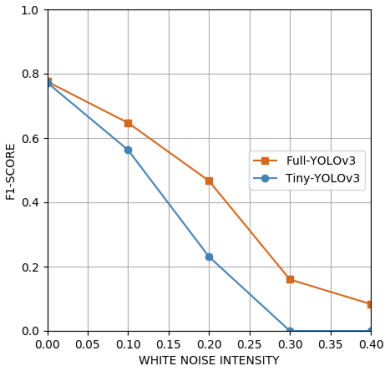
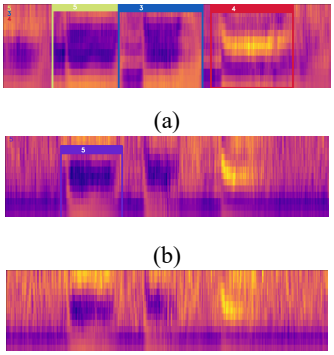


Figure 21: The performances of the full and tiny YOLOv3 when injecting the background noise.

It could be seen from Figure 22 that the full YOLOv3 model had the better adaptability than the tiny YOLOv3 to the noisy environment. Although the higher degrees of noise intensity could significantly degrade the model performances, the F1-score above 0.65 could still be obtained by both models when the variances of the white noise were below 0.05.



(a)

(b)

(c)

Figure 22: The comparison of the detection results:

a) α=0.1; b) α=0.2; c) α=0.3.

9. Conclusions and future work

This study provided the alternative approach for the automatic speech recognition (ASR) was proposed to detect Thai keywords in the spoken sentence. We next proposed the three separate real-world tasks to examine our model detection capabilities. The tasks were different in terms of difficulty, ranging from the standard test that just recognized the Thai Airway names, to the most difficult assignment that detects the specific keywords in the frequently asked questions (FAQs) using the keyword spotting technique. The core methodology is based on the Mel-frequency cepstral coefficients (MFCC), which is chosen as the feature extraction for

the speech signal, because this technique artificially implements the behavior of the human hearing mechanism based on the usage of the non-linear frequency scale of the real human, so it results in the parametrically resemblances between the extracted vectors and human sense of hearing. Then the state-of-the-art convolutional neural network object detector, You Only Look Once (YOLO), was performed as the keyword localizer and classifier. Due to the requirements of real-time speed and high accuracy of the ASR system, the tiny version of YOLOv3, which was the simplified and light-weight version of YOLOv3, was evaluated and compared with the regular version using the precision, recall, and F1-score to verify the feasibility and superiority. From the experiment, the F1-score of the full and tiny YOLOv3 from three different tasks were 77.54% and 77.23% respectively. To conclude, tiny YOLOv3, with the lower computational time and comparable detection accuracy compared to the regular YOLOv3, was proven to meet the ASR requirements and was the most suitable model for using in the low resource platforms.

Thai ASR was still the interesting and challenging topic because of the morphological richness of Thai language and the difficulties in developing Thai ASR model using the traditional technique. The YOLOv3 was applied to the keyword detection. Nevertheless, the experimental results drew back some further applications and additional improvements.

1. Compared with the results on the first and second task, the poor result on the third task told that the dataset should to be increased to produce higher performance.
2. To increase the model robustness to the background noise, injecting noise into dataset and trained together with the original dataset should be useful.
3. This work only proposed the Thai ASR application in the real-world airport scenario. Therefore, the research of the same basis should be increased to extend the application or investigate the new approaches.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] K. Sukvichai, C. Utintu, and W. Muknumporn, "Automatic Speech Recognition for Thai Sentence based on MFCC and CNNs," 2021 Second International Symposium on Instrument, Control, Artificial Intelligance, and Robotics (ICA-SYMP), 2021. doi:10.1109/ica-symp50206.2021.9358451

[2] H. Thaweesak Koanantakool, T. Karoonboonyanan, and C. Wutiwiwatchai, "Computers and the Thai Language," IEEE Annals of the History of Computing, **31**(1), 46-61, 2009. doi:10.1109/mahc.2009.5

[3] T. Pathumthan, ''Thai Speech Recognition Using Syllable Units,'' master's thesis, Chulalongkorn Univ., 1987.

[4] S. Suebvisai, P. Charoenpornsawat, A. W. Black, M. Woszczyna, and T. Schultz, ''Thai Automatic Speech Recognition,'' Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP), 2005. doi:10.1109/icassp.2005.1415249

[5] I. Thienlikit, C. Wutiwiwatchai, and S. Furui, ''Language Model Construction for Thai LVCSR,'' Reports of the Meeting of the Acoustic Soc. of Japan, ASJ, 131-132, 2004.

[6] M. Woszczyna, P. Charoenpornsawat and T. Schultz. "Spontaneous Thai Speech Recognition." The Ninth International Conference on Spoken Language Processing Multimodal Technologies, Inc. Carnegie Mellon University, Australia, 2006.

[7] S. Kasuriya, V. Sornlertlamvanich, P. Cotsomromg, S. Kanokphara, and N. Thatphithakkul, ''Thai Speech Corpus for Speech Recognition,'' Proc. Int'l Conf. Int'l Committee forthe Coordination and Standardization of Speech Databases and Assessments (O-COCOSDA), 105-111, 2003.

[8] S. Tangruamsub, P. Punyabukkana, and A. Suchato, "Thai Speech Keyword Spotting using Heterogeneous Acoustic Modeling," 2007 IEEE International Conference on Research, Innovation and Vision for the Future, 253-260, 2007. doi:10.1109/rivf.2007.369165

[9] P. Khunarsa, J. Mahawan, P. Nakjai, and N. Onkhum, "Nondestructive Determination of Maturity of the Monthong Durian by Mel-frequency Cepstral Coefficients (MFCCs) and Neural Network," Applied Mechanics and Materials, **885**, 75–81, 2016. doi:10.4028/www.scientific.net/amm.855.75

[10] Y. Segal, T. S. Fuchs, and J. Keshet, "SpeechYOLO: Detection and Localization of Speech Objects," Interspeech 2019, 2019. doi:10.21437/interspeech.2019-1749

[11] G. Chen, C. Parada, and G. Heigold, "Small-footprint Keyword Spotting using Deep Neural Networks," 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4087–4091, 2014. doi:10.1109/icassp.2014.6854370

[12] J.R. Rohlicek, W. Russell, S. Roukos, and H. Gish, "Continuous Hidden Markov Modeling for Speaker-independent Word Spotting," International Conference on Acoustics, Speech and Signal Processing (ICASSP), 627–630, 1990. doi:10.1109/icassp.1989.266505

[13] D. Grangier, J. Keshet, and S. Bengio, "Discriminative Keyword Spotting," Automatic Speech and Speaker Recognition, 173–194, 2009. doi:10.1002/9780470742044.ch11

[14] S. Fernández, A. Graves, and J. Schmidhuber, "An Application of Recurrent Neural Networks to Discriminative Keyword Spotting," Artificial Neural Networks – ICANN 2007, 220–229, 2007. doi:10.1007/978-3-540-74695-9_23

[15] T. Sainath and C. Parada, "Convolutional Neural Networks for Small-Footprint Keyword Spotting," 2015.

[16] D. Namrata, "Feature Extraction Methods LPC, PLP and MFCC in Speech Recognition," International Journal for Advance Research in Engineering and Technology (ISSN 2320-6802), 1, 2013

[17] R. Ranjan, A. Thakur, "Analysis of Feature Extraction Techniques for Speech Recognition System," International Journal of Innovative Technology and Exploring Engineering (IJITEE) (ISSN 2278-3075), 8. 2019

[18] Z.Q. Zhao, P. Zheng, S.T. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," IEEE Trans. Neural Networks Learn. Syst., 1-21, 2019.

[19] R.L. Galvez, A.A. Bandala, E.P. Dadios, R.R.P. Vicerra, and J.M.Z. Maningo, "Object Detection Using Convolutional Neural Networks," TENCON, 2018. doi:10.1109/tencon.2018.8650517

[20] H. Jiang and E. Learned-Miller, "Face Detection with the Faster R-CNN," 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition, 2017. doi:10.1109/fg.2017.82

[21] Z.-Q. Zhao, H. Bian, D. Hu, W. Cheng, and H. Glotin, "Pedestrian Detection Based on Fast R-CNN and Batch Normalization," Lecture Notes in Computer Science, 735-746, 2017. doi:10.1007/978-3-319-63309-1_65

[22] R. Girshick, "Fast R-CNN," 2005 IEEE International Conference on Computer Vision (ICCV), 2015. doi:10.1109/iccv.2015.169

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," IEEE Trans. PatternAnal. Mach. Intell., **37**(9), 1904–1916, 2015.

[24] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders, "Selective Search for Object Recognition," International Journal of Computer Vision, **104**(2), 154–171, 2013. doi:10.1007/s11263-013-0620-5

[25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, **39**(6), 1137-1149, 2017. doi:10.1109/tpami.2016.2577031

[26] T.-Y. Lin, P. Doll´ar, R. B. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. doi:10.1109/cvpr.2017.106

[27] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single Shot Multibox Detector," in Lecture Notes in Computer Science, 21-37, 2016. doi:10.1007/978-3-319-46448-0_2

[28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. doi:10.1109/cvpr.2016.91

[29] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov,

D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015. doi:10.1109/cvpr.2015.7298594

[30] M. J. Shafiee, B. Chywl, F. Li, and A. Wong, "Fast YOLO: A Fast You Only Look Once System for Real-time Embedded Object Detection in Video," Journal of Computational Vision and Imaging Systems, **3**(1), 2017. doi:10.15353/vsnl.v3i1.171

[31] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition, 7263-7271, 2017. doi:10.1109/cvpr.2017.690

[32] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 292-297 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Performance of Vertical Axis Wind Turbine Type of Slant Straight Blades

Hashem Abusannuga*, Mehmet Özkaymak

Energy Systems Engineering, Faculty of Technology, Karabuk University, Karabuk, 78050, Turkey

ARTICLE INFO

Article history:
Received: 18 April, 2021
Accepted: 25 July, 2021
Online: 03 August, 2021

Keywords:
VAWT
Multi-Stream Tube
Self-Starting Problem

ABSTRACT

There is no doubt that energy is one of the most important requirements of life, and its importance increases with the passage of time, and this is what make countries to harness the capabilities and scientists in developing energy systems of all kinds, one of the most important energy systems these days is what is known as vertical axis wind turbines. If we compare this type of system with horizontal axis wind turbines, it is characterized by a relatively lower manufacturing cost. But on the other hand, it suffers from less efficiency in addition to the problem of starting the self-movement. The idea of this research revolves around the use of an engineering design for the vertical axis wind rotor that is very rarely used in the field of wind energy. This design takes the geometric shape of two inverted trapezoids. Within the framework of this study, the term "slant straight-blade vertical axis wind turbine" (SS-VAWT) was assigned to the wind rotor. Amendments have been made to the mathematical model of Multi stream tube to make it suitable for application and work on (SS-VAWT), where, it is known that the multi-stream tube model uses primarily and only for the original Darrieus and the H-Darrieus rotors. In order to prove the efficacy of the software used, the results obtained from it were compared with the practical results of previous studies, as it proved its effectiveness in obtaining the satisfactory results that were intended for this analysis. The analyzes and investigations that were conducted on the improved SS design included changing the geometry by changing some of its dimensional parameters represented in rotor height, rotor diameter, number of rotor blades, rotor blade section length, rotor blade section type and rotor blades inclination angle on the horizontal plane. Within the scope of the case studies that were worked on in this research, the results showed that the best efficiency of the SS rotor was achieved in the range of height to radius ratio (0.66 to 1), cord line length to radius ratio about 0.12 The angle of inclination of the blades is between 45- and 65-degrees Degree. In these ranges, the value of Max power factors has reached its turn, and the energetic range of the rotor has increased as a function of the peripheral relative velocity, in addition to a relatively large solution to the problem of starting self-movement, which appears through the highest-power factor values to move away from the limits of negative values in the range Terminal forgetfulness from 1 to 3. In addition, the effect of changing Raynaud's number on the turbine aerodynamic performance has been investigated. The results showed that the higher the Reynolds value, the higher the power factor value, the higher the energy range and the lessening the problem of starting the self-movement.

1. Introduction

This paper is an extension work originally presented in "2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)" [1]**.** Due to the increase in the costs of electric energy consumption with the progression of time, we have called for the need to focus on renewable energies and invest them in order to meet the needs of humanity in terms of energy, because of the advantages of renewable energies such as their presence in nature permanently, inexhaustible, clean, multi-source and free of charge in nature. Where you need systems to harness and convert it into electrical

*Corresponding Author: Hashem Abusannuga, hashem_ebrahim@yahoo.com

https://dx.doi.org/10.25046/aj060432

energy. The most important types of wind energy and solar energy. [2], [3]. When the wind passes according to the "cut in speed Vc" of the vertical axis wind rotor, the air rotor acquires a rotational speed around the rotor shaft to make θ angle from 0^0 to 360^0, and thus the ends of the rotor blades acquire a continuous terminal speed in ωr, which there is a difference in the values and proportionality between the wind speed V in contact with the rotor, this leads to the formation of what is known relative velocity W, which make the angle of attack α with the direction of the ωr, therefore there is a strong relationship between the angle of attack α and the Tip Speed Ratio λ_0 or ($\omega R/V_1$). In the science of aerodynamics, specifically in the field of wings and blades, it is known that the angle of attack α has a direct and strong effect on the forces of lift L and Drag D, meaning that it has an effect on the ratio (L/D). Accordingly, the link between the terminal relative velocity and the ratio (L/D) becomes clear. Linking to the above, in order to achieve optimal efficiency to the rotor, must be sure, that the values of each $\omega R/V_1$, α, θ and L/D are consistent [4]-[7]. With the aim of developing the efficiency of VAWT's, many different engineering designs have emerged that compete with each other in terms of high performance and lower manufacturing cost [8]-[12]. Following some instances from VAWT's engineering designs. Original Darrieus-VAWT, its engineering shape is a parabolic, it gains its rotational movement from the lift force generated on its blades as a result of aerodynamics. On the other hand, this system suffers from a low coefficient of performance Cp, which does not exceed 35%. It also vibrates severely when rotating, which makes its manufacturing cost high due to the increase in structural supports to reduce vibrations [13]-[16]. Darrieus was developed so that his rotary shape became in the form of H, so it was called H-Darrieus, The H-Darrieus is nearly as efficient as the original Darrieus, but has lower manufacturing costs due to its straight blades [17]-[21]. The Helical-turbine is one of the most important developments of vertical axis wind turbines. It is characterized by its aerodynamic equilibrium during its rotation, as well as overcoming structural problems such as bending stress and vibrations. While its efficiency is slightly lower than other types. [22], [23]. All types of vertical axis air fans that were mentioned above and that were not mentioned also, suffer from the inability of the air rotor to start rotating. This problem is relatively addressed by adding a secondary system that relies in its rotation on obstruction to make the air rotor easier to start rotating. But this solution did not completely solve the self-starting problem [24]-[27]. All VAWT's models of vertical axis wind turbines were produced and attempted to be sold globally as HAWT's, but with little traction [28], [29]. The originality of this research appears in its geometric design, which takes the form of two inverted parabolas, and the modification of the famous mathematical method "Multi-Stream Tubes" MST so that it becomes appropriate to apply it to this design. The objective of this improved design is to increase the power factor and reduce the problem of starting movement compared to other types of VAWT's. Figure 1 illustrates the developed design of this research. The work methodology was completed so that twelve geometric shapes of the rotor SS-VAWT were selected, different among them in terms of rotor height 2H, rotor diameter 2R, number of rotor blades N, length of rotor blade cross section C or rotor blade inclination angle β. Subsequently, several analyzes were conducted with the aim of verifying the effect of SS-VAWT rotor geometry variables on its performance. Through the results of the analyses conducted by a software specially designed for this research, the optimum engineering design of the SS-VAWT rotor was reached, which was moved to other stages of this research, including analyzes by Ansys Fluent, manufacturing a prototype and testing it in a wind tunnel. This is a comprehensive research that started with a developed idea that came based on an extensive review of many previous studies and research, this paper shows theoretical aspects that have an applied research extension that is being implemented now through the manufacturing processes of a prototype which will be shown by the experiments that will be conducted on it in subsequent papers in the future.

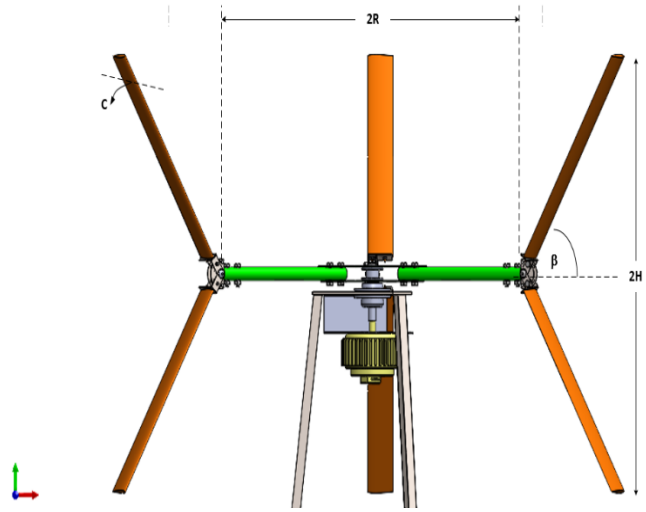


Figure 1: Geometry shape of SS-VAWT model by Solidworks

2. Method

The Multi-Stream Tube is mathematical method used in this work with certain adjustments to be optimal for apply to SS-VAWT and this model is preferred because it has a strong predictability for overall power production of the rotor and is also distinguished by efficiency and ease of use in the study of the influence of geometrical shape parameters on performance of wind rotor. With relation to details provided in [30], a brief overview of this model is offered. The rotor is substituted by the named "imaginary actuator disk." This disk was blocked by a sequence of "stream-tubes." It is an imagined tubs, the top of which consists of a stream-line, and thus the velocity vector is tangent everywhere on its surface. The velocity of the air across the disk differs from one stream-tube to another, much as the velocity of the air during the stream-tube shifts from its "free velocity" value V1 in front of the disk to its air velocity value "V" at the disk stage and even to its velocity "V2" at the wake area behind the disk. This continuous variation in disk velocity happens when part of the kinetic energy in the flow from which it passes is extracted [31]. Previous experimental data obtained in a wind tunnel was used to evaluate the software. The program was created to evaluate three different kinds of vertical axis wind turbines: original Darrieus rotor (Darrieus-VAWT), straight blades Darrieus rotor (H-VAWT), and slant straight blades rotor (SS-VAWT), the latter of which is the subject of this study. Figure 2 depicts the program approach in detail. Table 1 lists the twelve different SS-VAWT designs that have been examined and their efficiency

characteristics evaluated. The aerodynamic efficiency under the control of geometrical variables like blade inclination angle (β), wind rotor height (2H), wind rotor diameter (2R), number of rotor blades (N), and airfoil chord line was expressed using power factor curves derived from the analysis phase of the twelve configurations (C). For the presentation of the results in the shape of the Cp curves, the data for the rotor output are typically shown in non-dimensional formats, which allows such data to be utilized irrespective of the wind rotor scale while preserving geometric continuity across rotors in various dimensions. As a result, power coefficient curves as a characteristic of tip speed ratio are popular. The following diagram depicts the presentation of many important equations in the utilized model.

Table 1: Twelve configurations of SS-VAWT

"H" change configurations			
N	R	H	C
3	1	0.5	0.12
3	1	1	0.12
3	1	1.5	0.12
"R" change configurations			
N	R	H	C
3	0.5	1	0.12
3	1	1	0.12
3	1.5	1	0.12
"N" change configurations			
N	R	H	C
2	1	1	0.12
3	1	1	0.12
4	1	1	0.12
"C" change configurations			
N	R	H	C
3	1	1	0.06
3	1	1	0.12
3	1	1	0.18

2.1. Mathematical Expressions

- following SS-VAWT geometry expression was created for this research:

$$\left(\frac{r}{R}\right) = \left(\frac{\left(\frac{Z}{H}\right)}{\frac{1}{\left(\frac{H}{R}\right)} * \mathrm{Tan}\left(\beta * \frac{\pi}{180}\right)}\right) + 1 \quad (1)$$

- The power P is given by:

$$P = \frac{\rho NC}{2\pi} \int_0^H \int_0^\pi W^2 r \frac{\omega C_t}{\sin\beta} d\theta dz \quad (2)$$

- Tip speed ratio relation is:

$$\lambda_\circ = \frac{\omega R}{V_1} \quad (3)$$

- Mathematical formula of power coefficient, C_p, is:

$$C_P = \frac{P}{1/2 \rho V_1^3 A} \quad (4)$$

and

$$C_p = \frac{NC}{\pi A} \int_0^H \int_0^\pi \left(\frac{W}{V_1}\right)^2 \lambda_0 \frac{r}{R} \frac{C_t}{\sin\beta} d\theta dz \quad (5)$$

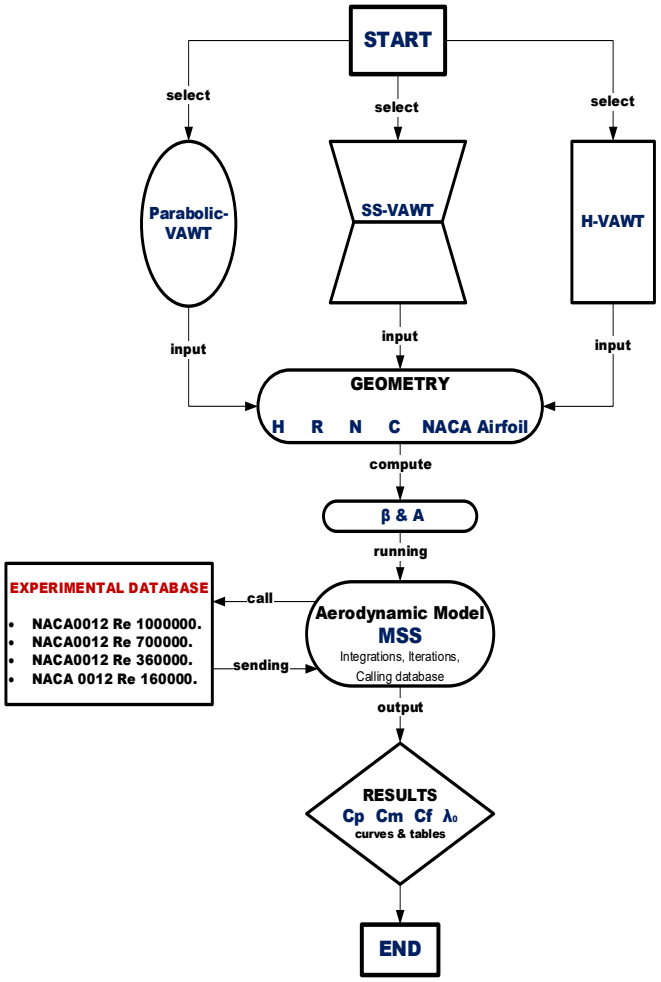


Figure 2: Flowchart of the software utilized

3. Results and Discussion

3.1. (β) Effect at Variable Values of H

The design group specialized in studying the effect of H change with β change is shown in Figure 3**.** The results are divided into four classes of curves, each group representing a different β angle, where we have four different angles as discussed above. In the same way, each category contains three curves that resulted from three different H values. The influence of the b angle on the Cp at three different values of H can be expressed as follows: at an angle of 85 degrees, the curves pattern did not vary much and there was a severe convergence between them, suggesting that there is no major effect of H value difference at this angle. The rotation began with a value of The value of CpM approached 0.3 at a value of λ_0 of around 3.8, and the values of the power factor appeared in a small λ_0 range between 3 and 5, indicating that the turbine is

inefficient in this case due to the low value of CpM, the narrow range of λ_0, and the magnitude of the self-starting issue. When b is equivalent to 65, the turbine's output increases when opposed to the situation where the angle is 85 degrees. In this scenario, the original λ_0 value is between 2 and 2.6, and the CpM at a comparable λ_0 value of 3.2 is calculated to be about 0.36. Furthermore, the value of λ_0 has improved, now varying between 2.2 and 4.8. As a consequence, these circumstances culminated in a relative dominance of 85 over the β. In comparison to the previous two cases of b equal to 85 and 65 degrees, the efficiency excellence of the third party described by β equal to 45 degrees is apparent. As the turbine rotation range increased from λ_0 (approximately 1.2) to 1 (approximately 4 and 5), the turbine rotation range also grew. Furthermore, the CpM values have been improved to approximately 0.36 and 0.42. To be more precise, the wind turbine in this case offered more control, a longer operating range, and a solution to the self-starting dilemma. When the β is equivalent to 25 degrees, the output of the SS-VAWT is investigated, the curves reveal that the turbine works poorly and is unsuitable for electrical energy generation. Where the power factor CpM highest value varied from 0.16 to 0.27.

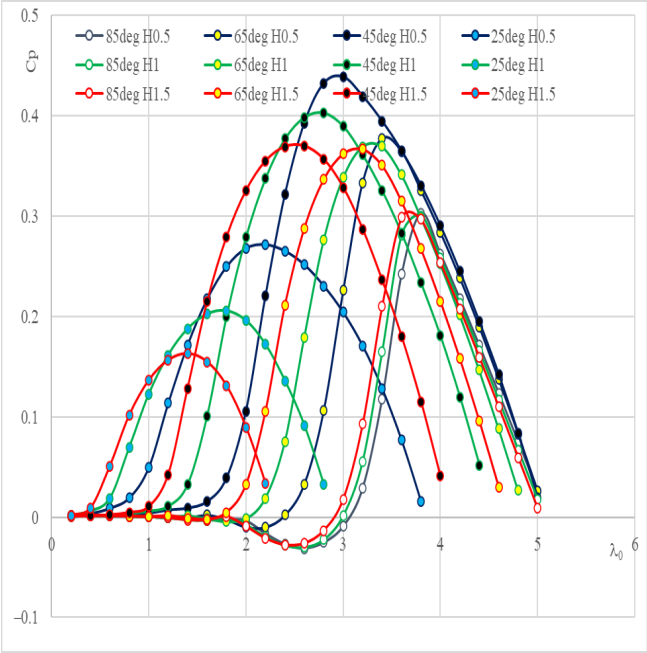


Figure 3: Power factor curves for twelve geometric shapes with change in height and angle of inclination of the blades.

3.2. (β) Effect at variable values of R

As seen in Figure 4 changing the angle and rotor radius values results in different Cp curve action. The engineering configurations that performed best were (β = 85, R = 1.5), (β = 65, R = 1.5, 1), and (β = 45, R = 1.5, 1), with CpM values ranging from 0.36 to 0.45. However, in terms of the energy continuum, (β = 65 degrees, R =1.5) and (β = 45 degrees, R =1.5) is preferred, with the rotor (β= 45 degrees, R =1.5) marked by an early start of rotational speed at the value of = 1.6. The bulk of the instances in Figure 4 had a relatively narrow energy range and a very low Cp, which led to poor performance.

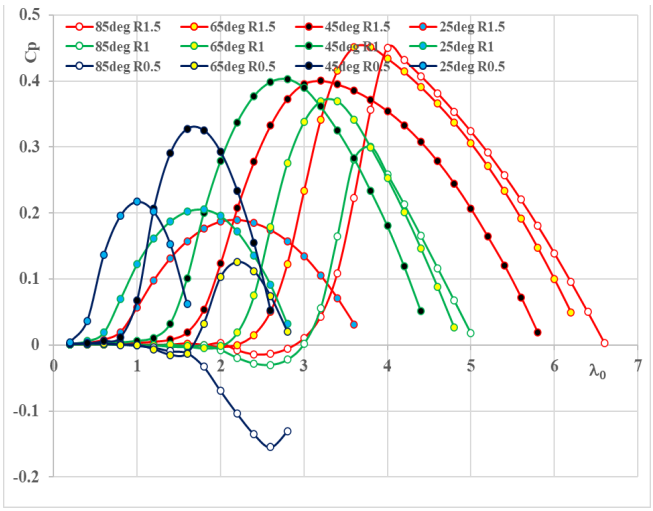


Figure 4: Power factor curves for twelve geometric shapes with change in diameter and angle of inclination of the blades.

3.3. (β) Effect at variable values of N

The three classes of curves in Figure 5 each show a specified N as 2, 3, or 4 and each include four curve angles = 85, 65, 45, and 25 degrees. It's readily evident in each category where N remains unaltered that the value of CpM increases, then falls, with the peak of CpM occurring at angles of 65 and 45 degrees. If we define the curves in another way, dividing them into four classes, each with a different fixed angle and vector N, we can see that when the beta 85 decreases dramatically, the value of the CpM decreases sharply as well, and the value of the energy spectrum decreases as well. It is obvious that modifying the conduct of curves at = 65 degrees is equivalent to changing attitudes at = 85 degrees in the preceding example. In the case of = 45 degrees, changing the value of N has no effect on the value of CpM, where It ranged between 0.37 and 0.40 in both situations, and it is balanced by values varying between 2.4 and 3, and the energy spectrum seems broad as compared to other sample cases. Finally, in the case of 25, the value of CpM rises with increasing N, but it does not surpass 0.24, considering the fact that the energy spectrum stays very small despite the shift in N.

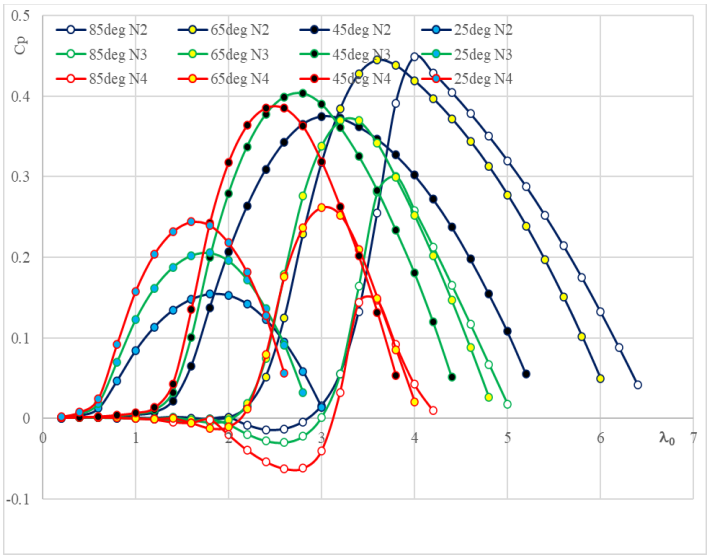


Figure 5: Power factor curves for twelve geometric shapes with change in number of blades and angle of inclination of the blades.

3.4. (β) Effect at variable values of C

Figure 6 expresses the effect of changing the blade section length on the SS-VAWT rotor in terms of power factor curves patterns. It is clear that there is a great similarity in the patterns of the Figure 6 curves with the patterns of the Figure 5 curves in paragraph 3.3. Thus, the ratio N/C can be used instead of N separately and C separately to check the efficiency of the SS-VAWT as a shortening of the analysis time.

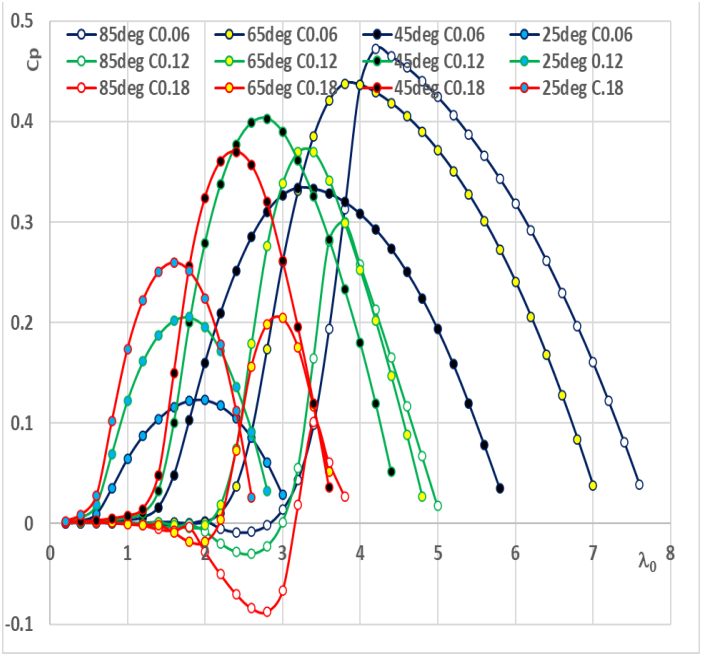


Figure 6: Power factor curves for twelve geometric shapes with change in cord line length and angle of inclination of the blades

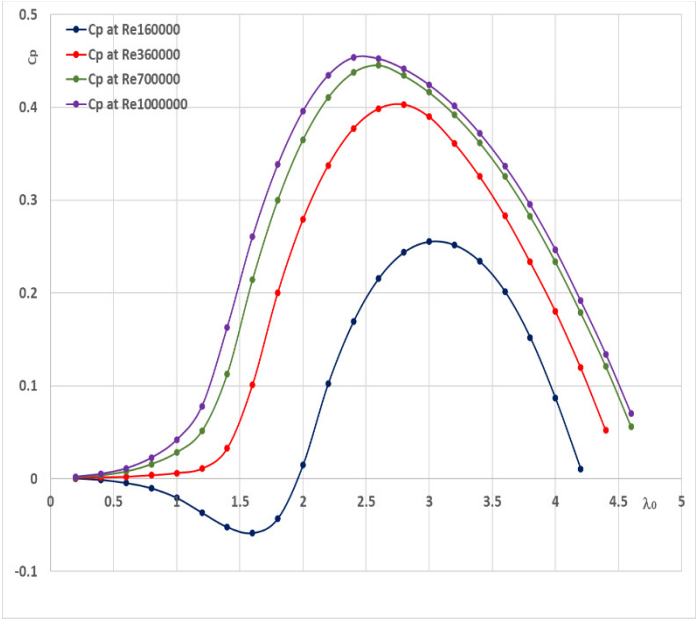


Figure 7: Influence of Reynolds Number on performance of SS-VAWT

3.5. Effect Reynolds Number on power factor of SS-VAWT

Figure 7. shows the change in the power factor values by changing the Reynolds number values. The values are 160000, 360000, 700000 and 1000000 [32]. It can be seen that in the case Reynolds number is equal to 160,000 the value of CpM is 0.25 and corresponding to λ_0 equal to 3. When Reynolds number is equal to 360,000, the CpM value is 0.4 and corresponding to λ_0 is 2.8. As for the Reynolds number equal to 700000, the CpM value is 0.44 and occurred at λ_0 equals 2.6. Finally, the Reynold Number 1000000 yielded CpM is 0.46 corresponding to a λ_0 of 2.4 value. Power factor curves also show that the power range ranges from 2 to 4.2 when the Re is 160,000, and 1.7 – 4.4 for Re equal to 360,000, and when Re = 700,000 shows the power range from 1.4 to 4.6, while at Re =1000000 the power range is 1.2 -4.7.

4. Conclusion

Within the scope of this research, the computer program achieved its objectives, as it predicted with high accuracy the power factor of the SS-VAWT rotor in many different study cases. By comparing with a previous study case that was harnessed for comparison only, the program is in great agreement with the practical results. The most important results extracted from this research are that the highest CpM values and the widest energetic range are achieved at angles of inclination of the blades between 45° and 65°. Regarding the problem of initiation of movement of the rotor, the satisfactory results were in the research cases in which the angles of inclination of the blades are between 25° and 45°. Regarding the effect of the Reynolds number on the performance of the SS-VAWT rotor, when Re value increase, the CP values increase, and consequently the CpM values increase. Moreover, as the Re values rise, the power range rises over the λ_0 values. Also, the problem of the self-starting appeared to diminish as the Re values accented.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] H. Abusannuga and M. Özkaymak, "Towards Evaluating the Performance of Vertical Axis Wind Turbine Consist of Slant Straight Blades," in 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Madrid, 2020.

[2] A. Kulshrestha, O. Mahela, M. Gupta, N. Gupta, N. Patel, T. Senjyu, M. Danish and M. Khosravy, "A Hybrid Fault Recognition Algorithm Using Stockwell Transform and Wigner Distribution Function for Power System Network with Solar Energy Penetration," Energies, **13**(14), 3519, 2020. doi: 10.3390/en13143519.

[3] A. Faramarz, T. Shiro, N. Tsutomu, N. Tomokazu, M.R. Asharif, " Analysis of Non-linear Adaptive Friction and Pitch Angle Control of Small-Scaled Wind Turbine System," (eds) Control and Automation, and Energy System Engineering. Communications in Computer and Information Science, 26-35, 2011. https://doi.org/10.1007/978-3-642-26010-0_4

[4] B. F. Blackwell, "The Vertical-Axis Wind Turbine "How It Works"," Sandia Laboratories, Albuquerque, New Mexico, 1974.

[5] R. Templin, "Aerodynamic Performance Theory for the NRC Vertical-Axis Wind Turbine," National Aeronautical Establishment, LTR-LA-160, Canada, 1974.

[6] M.J. Ralph, S.V. Maria and D. J. Ray, "Theoretical Performance Of Cross-Wind Axis Turbines With Results For A Catenary Vertical Axis Configuration," NASA Langley Research Center, Virginia, 1975.

[7] P. N. Shankar, "On The Aerodynamic Performance of a Class of Vertical Shaft Windmills," Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, **49**(16), 35-51, 1976.

[8] E. Sheldahl, B.F. Blackwell, "Free-Air Performance Tests of a 5-Metre-Diameter Darrieus Turbine," Sandia Laboratories, New Mexico, 1977.

[9] I.Paraschivoiu, "Double-Multiple Streamtube Model For Darrieus Wind Turbines," in Institut de Recherche de l'Hydro-Quebec, Quebec, 1981.

[10] P.M. Kumar, K. Sivalingam, S. Narasimalu, T.C. Lim, S. Ramakrishna and H. Wei "A Review on the Evolution of Darrieus Vertical Axis Wind Turbine: Small Wind Turbines," Journal of Power and Energy Engineering, **7**, 27-44, 2019.

[11] B. Muhammad, H. Nasir, F. Ahmed, A. Zain, J. Sh. Rehan and H. Zahid, "Vertical Axis Wind Turbine – A Review of Various Configurations And Design Techniques," Renewable and Sustainable Energy Reviews, **16**(4), 1926-1939, 2012.

[12] T. Dang, "Introduction, History, And Theory of Wind Power," in 41st North American Power Symposium, Starkville, MS, 2009.

[13] W. Tjiu, T. Marnoto, M. Sohif, H. Ruslan and K. Sopian, "Darrieus Vertical Axis Wind Turbine For Power Generation I: Assessment Of Darrieus VAWT Configurations," Renewable Energy, **75**, 50–67, 2015.

[14] A.D. Thomas, L.M. Timothy, "Developments in Blade Shape Design for a Darrieus Vertical Axis Wind Turbine," Sandia National Laboratories, Albuquerque, New Mexico, 1986.

[15] K.G. Emil, "Characteristics of Future Vertical Axis Wind Turbines," Sandia National Laboratories, Albuquerque, New Mexico, 1982.

[16] W. Tjiu, T. Marnoto, M. Sohif, H. Ruslan, K. Sopian, "Darrieus vertical axis wind turbine for power generation II: Challenges in HAWT and the opportunity of multi-megawatt Darrieus VAWT development," Renewable Energy, **75**, 560-571, 2015.

[17] W. S. Bannister, "Aerodynamic Studies Of A Straight-Bladed Vertical-Axis Wind Turbine," in 1St. B.W.E.A Wind Energy Workshop, Springfield, VA, 1979.

[18] W. S. Bannister, "A Theoretical Analysis Of Small Vertical Axis Wind Turbines," in International Synposium on "Applications of Fluid Mechanics and Heat Transfer to Energy and Environmental Problems, Greece, 1981.

[19] Y. Hara., N. Horita, S. Yoshida, H. Akimoto, T. Sumi, "Numerical Analysis of Effects of Arms with Different Cross-Sections on Straight-Bladed Vertical Axis Wind Turbine," Energies, **12**(11), 2019, https://doi.org/10.3390/en12112106

[20] J. Fadil, Soedibyo, M. Ashari, "Performance Analysis Of Vertical Axis Wind Turbine With Variable Swept Area,", International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, 2017.

[21] A. Gorlov, "Development of the Helical Reaction Hydraulic Turbine," Northeastern University, Boston, 1998.

[22] D. Han, Y. Heo, N. Choi, S. Nam, K. Choi, K. Kim, "Design, Fabrication, and Performance Test of a 100-W Helical-Blade Vertical-Axis Wind Turbine at Low Tip-Speed Ratio," Energies, **11**(6), 2018.

[23] M. Rahman, T. Salyers, , A. Mahbub, A. ElShahat, V. Soloiu, and M. Emile, "Investigation of Aerodynamic Performance of Helical Shape Vertical-Axis Wind Turbine Models With Various Number of Blades Using Wind Tunnel Testing and Computational Fluid Dynamics,", International Mechanical Engineering Congress and Exposition, Phoenix, Arizona, 2016.

[24] R. Dominy, P. Lunt, A. Bickerdyke, J. Dominy, "Self-Starting Capability Of A Darrieus Turbine," Journal of Power and Energy, **221**(1), 111–120, 2007.

[25] J. Zhu, H. Huang, & H. Shen, "Self-Starting Aerodynamics Analysis Of Vertical Axis Wind Turbine," Advances in Mechanical Engineering, **7**(12), 2015.

[26] M.S. Omar, I.A. Ahmed, E.A. Ahmed, A.A. Amr, A. M. Elbaz, "Numerical Investigation Of Darrieus Wind Turbine With Slotted Airfoil Blades," Energy Conversion and Management: X, 5, 100026, 2020.

[27] J. Krishnaraj, Sivakumar Ellappan, M. Anil Kumar, "Additive Manufacturing of a Gorlov Helical Type Vertical Axis Wind Turbine," International Journal of Engineering and Advanced Technology (IJEAT), **9**(2), 2019.

[28] IRENA, "Future of wind: Deployment, investment, technology, grid integration and socio-economic aspects," International Renewable Energy Agency (IRENA), Abu Dhabi, 2019.

[29] J. Damota, I. Lamas, A. Couce, J. Rodríguez, "Vertical Axis Wind Turbines: Current Technologies and Future Trends," in International Conference on Renewable Energies and Power Quality (ICREPQ'15), Coruna, 2015.

[30] H.J. Strickland, "The Darrieus Turbine: A Performance Prediction Model Using Multiple Streamtubes," Sandia Laboratories, Albuquerque, New Mexico, 1975.

[31] N. Batista, R. Melício, V. Mendes, J. Figueiredo, A. Reis, "Darrieus Wind Turbine Performance Prediction: Computational Modeling,", 4th Doctoral Conference on Computing, Electrical and Industrial, Costa de Caparica, 2013.

[32] R. E. Sheldahl, Klimas, P C, "Aerodynamic Characteristics of Seven Symmetrical Airfoil Sections Through 180-Degree Angle of Attack for Use in Aerodynamic Analysis of Vertical Axis Wind Turbines," Sandia National Laboratories, United States, 1981.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 298-305 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Evaluation of Information Competencies in the School Setting in Santiago de Chile

Jorge Joo-Nagata[*,1], Fernando Martínez-Abad[2]

[1]*Universidad Metropolitana de Ciencias de la Educación, Departamento de Historia y Geografía, Ñuñoa, 7760197, Chile*

[2]*Universidad de Salamanca, Instituto Universitario de Ciencias de la Educación, Salamanca, 37008, Spain*

ARTICLE INFO

Article history:
Received: 07 June, 2021
Accepted: 19 July, 2021
Online: 16 August, 2021

Keywords:
Competencies for life
Blended learning
Primary education

ABSTRACT

This study evaluated the competencies related to digital information use through technological tools aiming to acquire applicable knowledge by searching and retrieving information. Methodologically, a quasi-experimental design without a control group was applied to a sample of primary education students from Chile (n=266). First, a diagnosis of the digital-informational skills is performed, and, later, the results of a course in a blended learning context -b-learning- (treatment) are shown. The results show significant differences between the participant groups, confirming the learning in information competencies and distinguishing an initial level from a posterior intermediate level.

1. Introduction

In the present context of Information and Communications Technology (ICT), the development of basic skills, and contexts of data overstocking, they have gradually become the three greatest fields for process of teaching-learning in different schooling levels. Additionally, expectations and objectives for information use in a widely intertwined world are set. This way, key skills in the ICT context are search, use, evaluation, and information processing [1,2].

Thus, it is understood that the integration of new data and communication systems, mainly from the set of technologies such as the Internet, corresponds to a social reality that is challenging the global educational system [3,4] and, particularly, the educational system in Chile. The concern of the educational administration is well reflected in the successive efforts to inquire about and propose improvements in material coverage (facilities, hardware and software material coverage, connectivity at a national level), in training and innovation of teachers or the organization of teachers at a management or curriculum development level. However, the scientific knowledge on acquisition and development of digital and information competencies requires the application of systemic methodologies for collection, analysis, and validation of the information that allows drawing applicable conclusions, with a capacity of generalization to advance in this new curricular content.

The quality of education in Chile, in its different levels, is the reason why it is so stressed on considering including key competencies in the foundations of the mandatory educational curriculum and, a variable that considerably impacts in this improvement is the training level performed at a teacher level and the school education stage [5,6]. The evaluation and training are included in the implementation of ICT innovative actions at educational centers.

The objective of this article is to evaluate digital and information competencies along with the implementation and evaluation of an ICT innovative project leading to the training in searching and processing of information in students at the school educational system. Specifically, a diagnosis of the digital-informational competencies level in school students was performed to later propose an experimental design able to verify the effectiveness of a training program of blended learning for the development of digital-informational education, competencies in the permanent training of school students. Finally, a variation level in the evaluated dimensions is set - search, evaluation, processing, and information communication - at the group of analyzed students.

The rest of the article is organized as follows: firstly, this document provides information on digital-informational skills in a formal educational setting and the current situation of ICT. Then, the applied methodology, and the sample characteristics are detailed. Hereafter, the pre-test and post-test results are summarized. Lastly, discussion and conclusions are shown..

[*]Corresponding Author: Jorge Joo-Nagata, jorge.joo@umce.cl

2. Theoretical framework

2.1. Digital-informational skills and the school setting

The fact of using ICT as learning objects, vehicles of information literacy and tools of permanent training of students from the school system is going to provide an added value to the research field, showing themselves not only as methodological elements but also as findings and achievements. The ICT use in the development of school education is often presented as a positive impact on the development of digital skills that belong to the 21st century and skills in general [7,8]. Thus, digital literacy can be understood as the development of a wide range of skills derived from the use of applied technology, allowing students to research, create contents and communicate digitally, and therefore, to constantly participate in the development of a society with a strong digital component [7–9].

Despite the classic denomination of "digital natives" [10], where it is established that these digital skills are in current students by default, several studies have determined that their knowledge is in medium and low levels [11–15]. Thus, considering this information, it is necessary to develop these skills aiming to prevent differences in the use of digital technologies. In the particular case of Chile, several studies have shown that students are capable of solving tasks related to the use of the information as consumers, as well as organizing and managing digital information. On the other hand, very few students can have success in skills related to the use of information as producers and creators of contents [9,16,17].

Skills related to information management through digital tools –particularly free access resources found on the Internet– with the objective of acquiring valid and verified knowledge through the search and collection of information, along with the interpretation of text information from the reading of new data, reflection, and its evaluation, will be understood as information literacy, also known as ALFIN, by its Spanish initials [18–20].

Beyond the existing controversy on the inportance of terminology used in the very definition of concepts such as "Processing of Information" and "Digital Skill" [21–23], the fact is that "Informational Skill" can be placed within the scope of action of this key competency. At the center of the generic skills the information literacy is found, which has become a new paradigm at the ICT scene, which is understood as the cognitive–affective fabric that allows to people not only recognizing their informational needs but also acting, understanding, finding, evaluating, and using information of the most diverse nature and sources.

2.2. Digital learning in a blending learning setting

For the development of digital and information literacy, the adoption of online modality is increasingly massive, in particular when the courses are integrated into academic study plans, or in contexts where this type of education is required [24]. Besides, the experience of adopting online or b-learning modalities has been strengthened in extreme situations like confinement derived from COVID-19 disease, where teachers understand the context of online learning. However, during the implementation, a variety of problems have arised, like facilities availability, Internet access, the need for new forms of planning, implementation, and evaluation of learning, and collaboration with the parents [25–27]. Notwithstanding the above, digital technologies represent an educational resource for a better adaptation to different types of students and their different personal situations. Elements such as focusing on diversity and encouraging the development of key skills are taken into consideration, while they offer multiple possibilities for collaboration between teachers and partners in new communicative scenes of online character.

Thus, the challenge of providing an interactive learning experience for students in big classmate groups and the concern on the quality of teaching in this type of milieu have been key catalysts for the reconsideration of educational approaches in the different educational levels [28].

Blended learning (b-learning), also known as hybrid learning, constitutes an evolving field of research within the wider dimension of electronic learning (e-learning), and it related to instruction practices combining traditional presential approaches with online learning or mediated by technology with the Internet as the main tool used [24,29]. From the beginning of the 21st century, a significant number of studies in this field, with their diverse implications in educational fields have been developed. However, there is still limited evidence that b-learning improves the results of learning in students [24,30].

3. General context

The interest of organizations such as the European Union, the OECD, and UNESCO for this type of research is relevant [31–33], since the circle of evaluation-formation and educational innovation constitutes one of the strategic axes in the education field for the development, not only of key areas in all the productive fabric but also in critical thought in a globalized world. This way, the formation of citizens regarding skills focused on demands of work environment in the 21st century should constitute a fundamental concern for current governments.

Currently, a great impact in the study of information competencies in research has been reached [34–38]. However, efforts on the progress of evaluation tools of the proficiency level in digital and information skills are not wider enough. Most of the research carried out have established and implemented measure scales of self-perception of their own digital and information skill, of their elaboration, and in a setting of the general evaluation of digital skills [39–47]. It is important to point out that, while most of the research previously analyzed and cited keep a precise and concrete application, only a few of them present an integration of the evaluation and the use of digital and information competencies within the educational curriculum [47–49]. Although there is a specific increase of these digital skills in teachers from diverse areas, particularly teachers who teach one course, the development focused on the digital and information competencies in Chile has not been produced, and rather, they have been settled from the general implementation of digital skills [9,50–54].

Additionally, a significant part of the research carried out in the formal education field, from national as well as international contexts, and especially in primary school and university educational levels, establish training programs that develop themselves from specific curricular or disciplinary aspects

[39,42,55–60]. Nevertheless, due to the multidimensional structure that digital and information skills raise, the experiences within these training programs tend to aim at the specific dimension of technology, collection, and processing of data, rather than a complex, global and holistic view of them.

4. Objectives and hypothesis

Analyze the level of ICT competencies of the school education body to, secondly, present an experimental design able to check the efficacy of a teaching program for the increase of information competencies in the permanent formation of students from school education. Finally, the variation level of the measured dimensions –search, evaluation, processing, and communication of the information– is established in the group of students that were analyzed.

5. Materials and methods

5.1. Research design

This is an information analysis study based on investigation data in students of the last stage of primary school [61], where a quasi-experimental test type design without a control group was used [62,63]. In this study, an educational intervention with ICT and information competencies was performed to support innovation and integration models in the teaching and learning process..

5.2. Participants

The study sample is students of both sexes who belong to the last year of primary education in Chile, in the Province of Maipo area. The sample were stratified according to the school levels they belonged to, along with their source schools.

The sample was composed of 266 students and, according to the socio-demographic features, distribution can be recapitulated as:

- By variable sex of the sample was 50.4% of female students and 49,6% of male students.
- By age of students was established between 12 to 17 years, with an average of 13.89 years.
- The distribution based on education level was made according to sex: 63.53% of students belonged to the 8º grade of primary school, where 50.88% were women and 49.11% were men. 36.46% of students belonged to the 7º grade of primary school, where 50.51% were men and 49.48% were women.

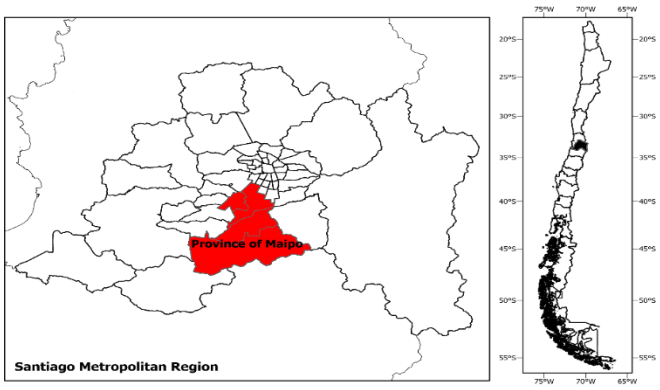


Figure 1: Study area: Province of Maipo

The sample derivates from the population of primary school students, whose ages range between 12 and 17 years old, belonging to the Province of Maipo in the Santiago Metropolitan Region (Figure 1), equivalent to N=106,745 [64], with a level of significance of $\alpha = 0.05$ and a maximum of homogeneity $p = q = 0.5$ for the sample of 266 participants, where the sampling error obtained is equal to 8.2%

5.3. Data collection instrument (test)

The dependent variable has been defined as the level of acquired competency of the students on digital and information skills, measured before (pre-test) and after (post-test) the implementation of the educational intervention (treatment).

The instrument was applied in the pre-test stages as a diagnosis of the competencies, and the post-test was applied for the evaluation of the information competencies reached by the participants students, measuring the level of performance obtained after the treatment. The test is composed of 29 elements of dichotomous nature, from 37 questions of single and multiple selection. 7 questions in the information search dimension, 10 questions in an evaluation dimension, 5 questions in the information processing dimension and other 9 questions on communication and dissemination of the information are presented (table 1).

Table 1: Dimensions and variables used in the instrument (test)

Dimension	Source	Metric	Description
Predictor variables			
Socio-demographic	- Test. - Data provided by the educational center or the Ministry of Education.	- Sex. - Age. - Grade - Municipality	Socio-demographic data of the participants of the study (sample).
Criterion variables			
Digital and information skills focused on the informational part.	Test	-Search of the information.	Grade obtained in the survey (instrument) about the item. Questions 1 to 17
		- Evaluation of the information	Grade obtained in the survey (instrument) about the item. 9 questions going from items 8 to 9.
		Analysis and selection of information (processing)	Data obtained through the survey (instrument) about item. 9 questions going from items 10 to 14.

		Communication of the information.	Grade obtained in the survey (instrument) about item. 9 questions going from items 15 to 18.

5.4. Treatment

The independent or treatment variable involved the teaching intervention applied to the participant during 30 training hours in a blended context [65–67], and adapted from the proposal described in [68]. This way, the teaching intervention applied in the training of sample corresponds to an adaptation of the instrument for the development of information competencies of secondary education [65,68] which has been tested and validated in that area.

5.5. Data analysis

Concerning the data analysis, after the exploratory initial analysis of the distributions of the variables and the evenness of the variances and covariances structures, parametric techniques, ANOVA with repeated measures are applied. Intra-subjects factors (pre-test/post-test) and inter-subjects factors (type of school) are incorporated. After the study with repeated measures, other techniques complementing the results are applied, such as the t-test. Additionally, all scores have been calculated so that each of the items is valued with a maximum score of 1 and a minimum score of 0. Besides, the scores of dimensions such as the average score of the set of items of the n dimension multiplied for 10 have been calculated. Thus, all dimensions are ranged from 0 to 10 points, which facilitates interpretation and contrast. Finally, the total score is calculated as the sum of the scores in the 4 dimensions. To ensure the internal consistency of the test results, the Cronbach's alpha test was performed, whose value was equal to 0.731, considered as adequate or acceptable [69,70]. All these statistical processes have been carried out in the SPSS 25 and R 4.0 software.

5.6. Procedure

The data in tests, in all cases, was collected mainly through online questionnaires. Responsible teachers were contacted, and the students participated by performing both pre-test and post-test online.

6. Results

6.1. General and descriptive results

Firstly, it is possible to determine values that approximate an average of 5 points (table 2) for intra-subjects characteristics (sex, grade and school). In parallel, there may be an important dispersion of values, which increases in post-test results.

It is observed that at the sample's level, the average reaches higher scores in Search, Processing, and Communication dimensions, while lower scores are registered in the Evaluation dimension (table 3). At a general level, the average score is greater in the student samples in the post-test. The dispersion of values increases in the post-test, so the teaching e-learning process has created higher levels of inequity between the measured students.

Table 2: total descriptive scores (n = 266)

		Pre-test		Post-test	
		Mean	**S.D**	**Mean**	**S.D.**
Sex	**Male**	4.809	1.263	5.207	1.413
	Female	4.694	1.213	5.331	1.449
Grade	**7º**	4.778	1.089	4.998	1.446
	8º	4.736	1.317	5.425	1.401
School	**School 1**	4.682	1.236	5.171	1.399
	School 2	5.051	1.207	5.697	1.494

Table 3: Descriptive statistics per dimension (n = 266)

		Mean	**S_x**	**P_{25}**	**P_{50}**	**P_{75}**
PRE-TEST	**SEARCH**	4.540	1.532	3.654	4.615	5.385
	EVALUATION	5.079	1.939	3.333	5.556	6.667
	PROCESSING	4.511	1.692	3.636	4.545	5.455
	COMUNICATION	5.000	2.044	4.000	5.000	6.000
	TOTAL	**4.752**	**1.237**	**3.953**	**4.651**	**5.581**
POST-TEST	**SEARCH**	5.552	1.824	3.846	5.385	6.923
	EVALUATION	5.046	1.985	3.333	5.000	6.667
	PROCESSING	4.863	1.876	3.636	4.545	6.364
	COMUNICATION	5.550	2.128	4.000	6.000	7.000
	TOTAL	**5.270**	**1.430**	**4.419**	**5.233**	**6.337**

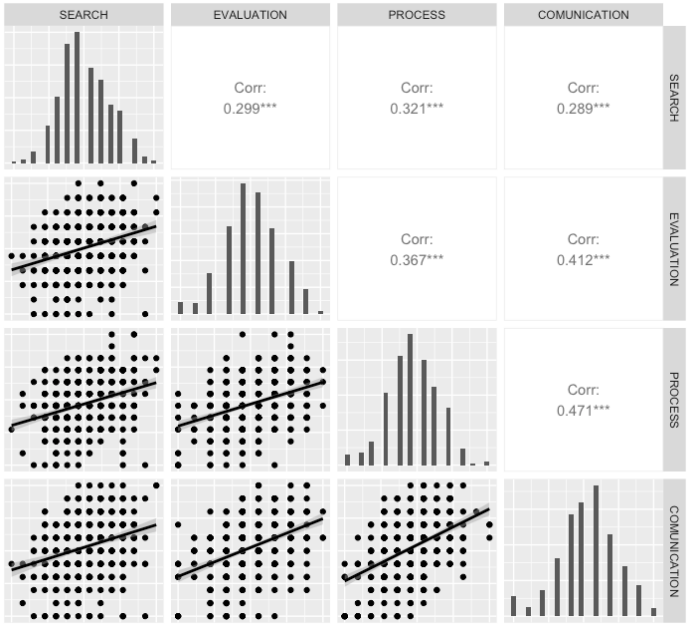


Figure 2: relation and distribution between dimensions

Moreover, it is observed how the distribution of pretest and post-test is similar in both groups, and their progress is similarly reached, while in the Evaluation dimension there is a decline or no

progress at all for both cases. It does not seem to be interaction according to the school; therefore, it is not considered a covariable. By obtaining correlations between dimensions, it is possible to establish that all values are significant and positive (figure 2).

By comparing between groups and the results obtained in the post-test, it is possible to perceive differences between the course and the school groups. However, based on the sex variable, values in the scores in different dimensions do not show great differences (figure 3).

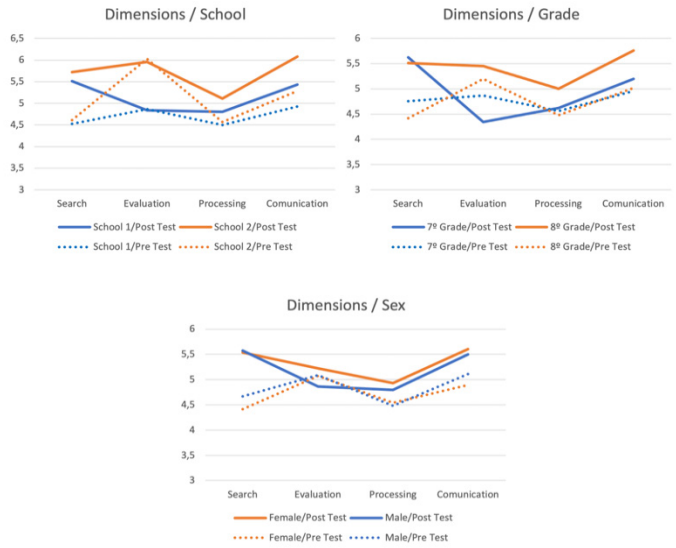


Figure 3: Post-test results based on dimensions and groups (n=266)

It is also possible to establish that in total scores, from all the variables analyzed in the contrasts, the sex variable is not significant in an inter-subject level nor an intra-subject interaction (table 4).

Table 4: Comparison per variables (n = 266)

		Pre-test		Post-test	
		Mean Dif.	t (p.)	Mean Dif.	t (p.)
Sex		-0.115	-0.758 (0.449)	0.123	-0.704 (0.482)
Grade		0.041	- 0.264 (0.792)	-0.427	-2.365 (0.019)
School		-0.368	-1.909 (0.057)	-0.526	-2.366 (0.019)
			Post-test – Pre-test		
			Mean Dif.		t (p.)
Sex	**Male**		-0.398		-3.997 (<.001)
	Female		-0.636		-5.614 (<.001)
Grade	**7º**		-0.220		-1.667 (0.099)
	8º		-0.689		-7.688 (<0.001)
School	**School 1**		-0.488		-5.631 (<0.001)
	School 2		-0.646		-4.359 (<0.001)
Complete sample			-0.518		-6.841 (<0.001)

6.2. ANOVA results of repeated measures

A hypothesis in the use of the ANOVA test with repeated measures is the matrix homogeneity of the covariances of the dependent variables [71,72]. This hypothesis is determined through the Box test, with the following results (table 5):

Table 5: Box test of the equality of covariances matrix

Box M.	**11.194**
F	**1.842**
df1	**6**
df2	**268376.644**
Sig.	**0.087**

Thus, the significance level showed that in this test a value of 0.087 was obtained, which exceeds 0.05 and, therefore, with a probability grade of 95% that the hypothesis on covariances matrix observed of the dependent variables are equal between groups.

Descriptive statistics and the results of the ANOVA test of repeated measures applied to data (table 6) showed that there is a significant interaction depending on the Grade, so it was decided to keep this variable, regardless that it is not significant at an individual level.

Table 6: Intra-subjects effects

INTRA-SUBJECT EFFECTS	F	p.
PRE-POST	**30.92**	**<.001**
PRE-POST * Grade	**8.518**	**0.004**
PRE-POST * School	**0.088**	**0.767**

The training action was significantly more effective in the eighth-grade group course than in the seventh-grade course. While the levels are similar in the pre-test scores - slightly higher in the eighth-grade group -, the training action had better results in the eighth-grade students. Nevertheless, the School variable does not have interaction in pre-post test scores (Figure 4).

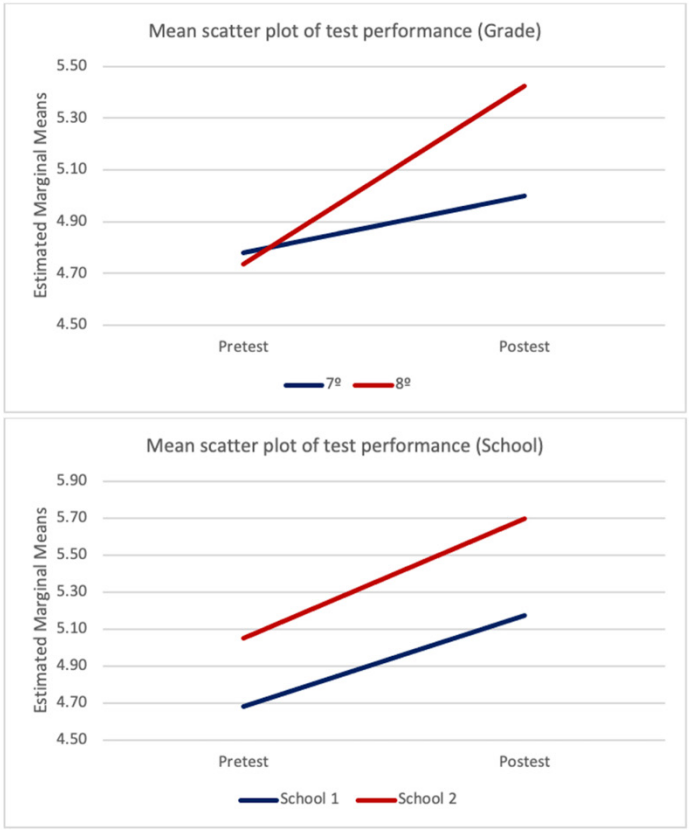


Figure 4: interactions between pre-test and post-test scores in grade and school variables

In table 7 it is possible to see the effects caused by the characteristics of the subjects as a transversal measure. Thus, it is possible to establish that the School variable is significant, but that is not the case for the Grade variable. This implies that at a general level, and calculating the average score of each subject in the pre-test and post-test, and also considering the criteria of the average score variable, there are significant differences for the School variable, but not for the Grade variable.

Table 7: Inter-subjects effect

INTER-SUBJECT EFFECTS	F	p.
Grade	0.186	.666
School	4.408	.037
Grade * School	-	-

Although all students in different schools and courses reached substantial progress in their digital competencies, there are also constant differences between the schools participating in the research.

7. Discussion

The students participating in the study, despite belonging to two different schools and two different teaching levels, in which there are different methodologies and contents within the elementary school, behave as a heterogeneous group in the information competencies area, with statistically significant differences between groups compared in the totals as much as in the analyzed dimensions (tables 1-7 and figures 1-2), which is complemented with the average age of the test responding students along with the sex variable, factors that do not have great incidence in the results due to the developed b-learning educational intervention.

Regarding the diverse analyzed variables, students behave in a heterogeneous manner, despite all the participants increasing their scores in the post-test measures. Within the established dimensions, the differences existing in the evaluation and the processing of the digital information areas are highlighted, a situation that has been stated in other investigations [16]. This way, particular uses in the application of technologies skills are presented. Regarding values on indicators of digital competencies and forms in which they show, participant students keep medium to low digital competencies indexes, which coincide with what has been carried out in others similar research [4,49,73].

8. Conclusions

Within the complexity of the Information Society, which is constantly mediated by the impact of ICT, it demands to the educational processes the incorporation of key skills, where the digital and information skills are prominent, related to the treatment of information in a virtual setting and the competency of digital processing. From this aspect, the training of the students from the first cycles of teaching gains importance. Thus, this study evaluated the efficacy of an educational intervention in a b-learning setting for the training in information competencies at an elementary educational level (seventh and eighth grades), obtaining progress considered significant but not enough for the ICT context where we live.

Empirical values obtained in the different dimensions confirm an appropriation in the digital competencies learning, where it is possible to differentiate an initial level from a posterior level after the applying of the educational intervention, and also between other variables as grade from school variables, establishing different realities in each educational context.

In the conclusions, the importance of the evaluation and training in digital and information skills is considered, addressing the fundamental dimensions that compose them, and establishing the efficacy of the implemented educational intervention. On the other hand, the research contributes to the study area according to the scale of reference - Santiago de Chile - and the training context - elementary students. Even though in Chile there are efforts to know the use of technologies in secondary teaching students [9], in the case of elementary students there are no studies that establish the characteristics and the levels concerning the information competencies in a digital setting.

Finally, after an analysis of the research contributions, weaknesses are established, which are focused on the design and development of the evaluation instrument and the experimental level of the applied design.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research was supported by the Research Direction of the Universidad Metropolitana de Ciencias de la Educación, Chile [Proyecto FIPEA 02-18].

References

[1] M. Bielva Calvo, F. Martinez Abad, M.E. Herrera García, M.J. Rodríguez Conde, "Diseño de un instrumento de evaluación de competencias informacionales en Educación Secundaria Obligatoria a través de la selección de indicadores clave," Education in the Knowledge Society (EKS), **16**(3), 124–143, 2015, doi:10.14201/eks2015163124143.

[2] UNESCO, A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2, 2018.

[3] A. Aslan, C. Zhu, "Pre-service teachers' perceptions of ICT integration in teacher education in Turkey," Turkish Online Journal of Educational Technology, **14**, 97–110, 2015.

[4] V. Dagiene, Development of ICT competency in pre-service teacher education, IGI Global, Lazio: 65–75, 2013.

[5] E. Palma Gajardo, "Percepción y Valoración de la Calidad Educativa de Alumnos y Padres en 14 Centros Escolares de la Región Metropolitana de Santiago de Chile," REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, **6**(1), 85–13, 2008.

[6] M. Perticará, M. Román, Los desafíos de mejorar la calidad y la equidad de la educación básica y media en Chile, Konrad-Adenauer-Stiftung, Santiago, Chile: 95–122, 2014.

[7] J. Fraillon, J. Ainley, W. Schulz, T. Friedman, E. Gebhardt, Preparing for life in a digital age: The IEA international computer and information literacy study. International report, Springer International Publishing, London: 15–25, 2014, doi:10.1007/978-3-319-14222-7_1.

[8] R. Schmid, D. Petko, "Does the use of educational technology in personalized learning environments correlate with self-reported digital skills and beliefs of secondary-school students?," Computers & Education, **136**, 75–86, 2019, doi:10.1016/j.compedu.2019.03.006.

[9] M. Claro, D.D. Preiss, E. San Martín, I. Jara, J.E. Hinostroza, S. Valenzuela, F. Cortes, M. Nussbaum, "Assessment of 21st century ICT skills in Chile: Test design and results from high school level students," Computers & Education, **59**(3), 1042–1053, 2012, doi:10.1016/j.compedu.2012.04.004.

[10] M. Prensky, "Digital natives, digital inmigrants," On the Horizon, **9**(5), 1–6, 2001.

[11] C.J. Asarta, J.R. Schmidt, "Comparing student performance in blended and traditional courses: Does prior academic achievement matter?," The Internet and Higher Education, **32**, 29–38, 2017, doi:10.1016/j.iheduc.2016.08.002.

[12] F.D. Guillen-Gamez, M.J. Mayorga-Fernández, M.T.D. Moral, "Comparative research in the digital competence of the pre-service education teacher: face-to-face vs blended education and gender," Journal of E-Learning and Knowledge Society, **16**(3), 1–9, 2020, doi:10.20368/1971-8829/1135214.

[13] M.A. Harjoto, "Blended versus face-to-face: Evidence from a graduate corporate finance class," Journal of Education for Business, **92**(3), 129–137, 2017, doi:10.1080/08832323.2017.1299082.

[14] L. Mohebi, Leaders' Perception of ICT Integration in Private Schools: An Exploratory Study from Dubai (UAE), SSRN Scholarly Paper ID 3401811, Social Science Research Network, Rochester, NY, 2019, doi:10.2139/ssrn.3401811.

[15] P. Slechtova, "Attitudes of Undergraduate Students to the Use of ICT in Education," Procedia - Social and Behavioral Sciences, **171**, 1128–1134, 2015, doi:10.1016/j.sbspro.2015.01.218.

[16] T. Ayale-Pérez, J. Joo-Nagata, "The digital culture of students of pedagogy specialising in the humanities in Santiago de Chile," Computers & Education, **133**, 1–12, 2019, doi:10.1016/j.compedu.2019.01.002.

[17] J. Joo Nagata, P. Humanante Ramos, M.Á. Conde González, J.R. García-Bermejo, F. García Peñalvo, "Comparison of the Use of Personal Learning Environments (PLE) Between Students from Chile and Ecuador: An Approach," in Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality, ACM, New York, NY, USA: 75–80, 2014, doi:10.1145/2669711.2669882.

[18] D. Bawden, "Revisión de los conceptos de alfabetización informacional y alfabetización digital," Anales de Documentación, **5**, 361–408, 2002.

[19] I. Szőköl, K. Horváth, "Introducing of New Teaching Methods in Teaching Informatics," in: Auer, M. E. and Tsiatsos, T., eds., in The Challenges of the Digital Transformation in Education, Springer International Publishing, Cham: 542–551, 2020, doi:10.1007/978-3-030-11932-4_52.

[20] S. Virkus, "Information literacy in Europe: a literature review," Information Research, **8**(4), 2003.

[21] J. Castaño, J.M. Duart, T. Sancho, "A second digital divide among university students," Cultura y Educación, **24**(3), 363–377, 2012, doi:10.1174/113564012802845695.

[22] J. Castaño-Muñoz, "Digital inequality among university students in developed countries and its relation to academic performance.," International Journal of Educational Technology in Higher Education, **7**(1), 43–52, 2010.

[23] L. Starkey, "A review of research exploring teacher preparation for the digital age," Cambridge Journal of Education, **50**(1), 37–56, 2020, doi:10.1080/0305764X.2019.1625867.

[24] C. McGuinness, C. Fulton, "Digital Literacy in Higher Education: A Case Study of Student Engagement with E-Tutorials Using Blended Learning," Journal of Information Technology Education: Innovations in Practice, **18**, 001–028, 2019.

[25] S. Dhawan, "Online Learning: A Panacea in the Time of COVID-19 Crisis," Journal of Educational Technology Systems, **49**(1), 5–22, 2020, doi:10.1177/0047239520934018.

[26] I. Fauzi, I.H.S. Khusuma, "Teachers' Elementary School in Online Learning of COVID-19 Pandemic Conditions," Jurnal Iqra' : Kajian Ilmu Pendidikan, **5**(1), 58–70, 2020, doi:10.25217/ji.v5i1.914.

[27] D. Moszkowicz, H. Duboc, C. Dubertret, D. Roux, F. Bretagnol, "Daily medical education for confined students during coronavirus disease 2019 pandemic: A simple videoconference solution," Clinical Anatomy, **33**(6), 927–928, 2020, doi:10.1002/ca.23601.

[28] D.J. Hornsby, R. Osman, "Massification in higher education: large classes and student learning," Higher Education, **67**(6), 711–719, 2014, doi:10.1007/s10734-014-9733-1.

[29] D.R. Garrison, H. Kanuka, "Blended learning: Uncovering its transformative potential in higher education," Internet and Higher Education, **7**(2), 95–105, 2004, doi:10.1016/j.iheduc.2004.02.001.

[30] G. Siemens, D. Gašević, S. Dawson, Preparing for the digital university: A review of the history and current state of distance, blended, and online learning., Athabasca University, Arlington: Link Research Lab.: 234, 2015.

[31] Eurydice, La profesión docente en Europa. Prácticas, percepciones y políticas, Oficina de Publicaciones de la Unión Europea, Luxemburgo, 2015.

[32] OCDE, La definición y selección de las competencias clave. Resumen ejecutivo., DeSeCo, OCDE: 20, 2005.

[33] T. Valencia-Molina, A. Serna-Collazos, S. Ochoa-Angrino, A.M. Caicedo-Tamayo, J.A. Montes-González, J.D. Chávez-Vescance, Competencias y estándares TIC desde la dimensión pedagógica: una perspectiva desde los niveles de apropiación de las TIC en la práctica educativa docente., Cali, 2016.

[34] M. Bouckaert, "Designing a materials development course for EFL student teachers: principles and pitfalls," Innovation in Language Learning and Teaching, **10**(2), 90–105, 2016, doi:10.1080/17501229.2015.1090994.

[35] C. Flores-Lueg, R. Roig-Vila, "Percepción de estudiantes de Pedagogía sobre el desarrollo de su competencia digital a lo largo de su proceso formativo," Estudios pedagógicos, **42**(3), 7, 2016.

[36] J. Garrido M, D. Contreras G, C. Miranda J, "Analysis of the pedagogical available of preservices teacher to use ICT," Estudios Pedagógicos (Valdivia), **39**(ESPECIAL), 59–74, 2013, doi:10.4067/S0718-07052013000300005.

[37] A. Martínez, J.G. Cegarra Navarro, J.A. Rubio Sánchez, "Aprendizaje basado en competencias: Una propuesta para la autoevaluación del docente," Profesorado. Revista de Currículum y Formación de Profesorado, **16**(2), 325–338, 2012.

[38] P. Martínez Clares, B. Echeverría Samanes, "Formación basada en competencias," Revista de Investigación Educativa, **27**(1), 125–147, 2009.

[39] A. Blasco Olivares, G. Durban Roca, "La competencia informacional en la enseñanza obligatoria a partir de la articulación de un modelo específico," Revista española de Documentación Científica, **35**(Monográfico), 100–135, 2012, doi:10.3989/redc.2012.mono.979.

[40] L. González Niño, G.P. Marciales Vivas, H.A. Castañeda Peña, J.W. Barbosa Chacón, J.C. Barbosa Herrera, "Competencia informacional: desarrollo de un instrumento para su observación," Lenguaje, **41**(1), 105–131, 2013.

[41] S.U. Kim, D. Shumaker, "Student, Librarian, and Instructor Perceptions of Information Literacy Instruction and Skills in a First Year Experience Program: A Case Study," The Journal of Academic Librarianship, **4**(41), 449–456, 2015, doi:10.1016/j.acalib.2015.04.005.

[42] E. Kuiper, M. Volman, J. Terwel, "Developing Web literacy in collaborative inquiry activities," Computers & Education, **52**, 668–680, 2009, doi:10.1016/j.compedu.2008.11.010.

[43] F. Martinez-Abad, P. Torrijos-Fincias, M.J. Rodríguez-Conde, "The eAssessment of Key Competences and their Relationship with Academic Performance:," Journal of Information Technology Research, **9**(4), 16–27, 2016, doi:10.4018/JITR.2016100102.

[44] E. Resnis, K. Gibson, A. Hartsell-Gundy, M. Misco, "Information literacy assessment: a case study at Miami University," New Library World, **111**(7/8), 287–301, 2010, doi:10.1108/03074801011059920.

[45] H. Saito, K. Miwa, "Construction of a learning environment supporting learners' reflection: A case of information seeking on the Web," Computers & Education, **49**(2), 214–229, 2007, doi:10.1016/j.compedu.2005.07.001.

[46] S. Santharooban, "Analyzing the level of information literacy skills of medical undergraduate of Eastern University, Sri Lanka," Journal of the University Librarians Association of Sri Lanka, **19**(2), 2016.

[47] S. Santharooban, P.G. Premadasa, "Development of an information literacy model for problem based learning," Annals of Library and Information Studies (ALIS), **62**(3), 138–144, 2015.

[48] S.C. Kong, "A curriculum framework for implementing information technology in school education to foster information literacy," Computers & Education, **51**, 129–141, 2008, doi:10.1016/j.compedu.2007.04.005.

[49] A. Pérez Escoda, M.J. Rodríguez Conde, "Evaluación de las competencias digitales autopercibidas del profesorado de Educación Primaria en Castilla y León (España)," Revista de Investigación Educativa, **34**(2), 399–415, 2016, doi:10.6018/rie.34.2.215121.

[50] M. Claro, A. Salinas, T. Cabello-Hutt, E. San Martín, D.D. Preiss, S. Valenzuela, I. Jara, "Teaching in a Digital Environment (TIDE): Defining and measuring teachers' capacity to develop students' digital information and communication skills," Computers & Education, **121**, 162–174, 2018, doi:10.1016/j.compedu.2018.03.001.

[51] Enlaces, Competencias TIC en la Profesión Docente, Ministerio de Educación, Chile, Santiago de Chile, 2007.

[52] Enlaces, Estándares en Tecnología de la Información y la Comunicación para la Formación Inicial Docente, Ministerio de Educación, Chile, Santiago de Chile, 2007.

[53] Enlaces, Matriz de habilidades TIC para el Aprendizaje, 2013.

[54] I. Jara, M. Claro, J.E. Hinostroza, E. San Martín, P. Rodríguez, T. Cabello, A. Ibieta, C. Labbé, "Understanding factors related to Chilean students' digital skills: A mixed methods analysis," Computers & Education, **88**, 387–398, 2015, doi:10.1016/j.compedu.2015.07.016.

[55] I. Aguaded, I. Marín-Gutiérrez, E. Díaz-Pareja, "La alfabetización mediática entre estudiantes de primaria y secundaria en Andalucía (España)," RIED. Revista Iberoamericana de Educación a Distancia, **18**(2), 275–298, 2015, doi:10.5944/ried.18.2.13407.

[56] M. Fuentes Agustí, C. Monereo Font, "Cómo buscan información en Internet los adolescentes," Investigación en la escuela, (64), 45–58, 2008.
[57] M.J. Grant, A.J. Brettle, "Developing and evaluating an interactive information skills tutorial," Health Information & Libraries Journal, **23**(2), 79–88, 2006, doi:10.1111/j.1471-1842.2006.00655.x.
[58] N. Landry, J. Basque, "L'éducation aux médias : contributions, pratiques et perspectives de recherche en sciences de la communication," Communiquer. Revue de communication sociale et publique, (15), 47–63, 2015, doi:10.4000/communiquer.1664.
[59] G.P. Marciales Vivas, H.A. Castañeda-Peña, J.W. Barbosa-Chacón, I. Barreto, L. Melo, "Fenomenografía de las competencias informacionales: perfiles y transiciones," Revista Latinoamericana de Psicología, 58–68, 2016, doi:10.1016/j.rlp.2015.09.007.
[60] M.L. Tiscareño Arroyo, J. de J. Cortés-Vera, "Competencias informacionales de estudiantes universitarios:: una responsabilidad compartida. Una revisión de la literatura en países latinoamericanos de habla hispana.," Revista Interamericana de Bibliotecología, **37**(2), 117–126, 2014.
[61] J.H. McMillan, S. Schumacher, Research in Education: Evidence-Based Inquiry, 7th Edition, Pearson, New Jersey, 2010.
[62] C. Anwar, A. Saregar, Y. Yuberti, N. Zellia, W. Widayanti, R. Diani, I.S. Wekke, "Effect size test of learning model arias and PBL: Concept mastery of temperature and heat on senior high school students," Eurasia Journal of Mathematics, Science and Technology Education, **15**(3), 2019, doi:10.29333/ejmste/103032.
[63] R. Sagala, R. Umam, A. Thahir, A. Saregar, I. Wardani, "The effectiveness of stem-based on gender differences: The impact of physics concept understanding," European Journal of Educational Research, **8**(3), 753–761, 2019, doi:10.12973/eu-jer.8.3.753.
[64] Asociación Chilena de Municipalidades, Universidad San Sebastián, COMUNAS Y EDUCACIÓN: Una aproximación a la oferta educativa comunal, Dirección de Estudios AMUCH Asociación de Municipalidades de Chile, Santiago (Chile): 34, 2017.
[65] M. Bielva Calvo, F.M. Martínez Abad, M.J. Rodríguez Conde, "Validación psicométrica de un instrumento de evaluación de competencias informacionales en la educación secundaria," Bordón. Revista de Pedagogía, **69**(1), 27–43, 2016, doi:10.13042/Bordon.2016.48593.
[66] J. Cabero Almenara, "Formación del profesorado universitario en TIC. Aplicación del método Delphi para la selección de los contenidos formativos," Educación XX1, **17**(1), 111–132, 2013, doi:10.5944/educxx1.17.1.10707.
[67] J. Cabero Almenara, M. Llorente Cejudo, V. Marín Díaz, "Hacia el diseño de un instrumento de diagnóstico de 'competencias tecnológicas del profesorado' universitario," 2010.
[68] F. Martínez Abad, S. Olmos Migueláñez, M.J. Rodríguez Conde, María J., "Evaluación de un programa de formación en competencias informacionales para el futuro profesorado de E.S.O.," Revista de Educacion, **370**, 45–70, 2015.
[69] D. George, P. Mallery, SPSS for Windows Step by Step: A Simple Guide and Reference, 11.0 Update, Allyn and Bacon, 2003.
[70] J.C. Nunnally, Psychometric theory, 2nd ed., McGraw-Hill, New York, 1978.
[71] M.E. Legget, M. Toh, A. Meintjes, S. Fitzsimons, G. Gamble, R.N. Doughty, "Digital devices for teaching cardiac auscultation - a randomized pilot study," Medical Education Online, **23**(1), 1524688, 2018, doi:10.1080/10872981.2018.1524688.
[72] A. Tahriri, M.D. Tous, S. MovahedFar, "The Impact of Digital Storytelling on EFL Learners' Oracy Skills and Motivation," International Journal of Applied Linguistics and English Literature, **4**(3), 144–153, 2015, doi:10.7575/aiac.ijalel.v.4n.3p.144.
[73] J. García-Martín, J.-N. García-Sánchez, "Pre-service teachers' perceptions of the competence dimensions of digital literacy and of psychological and educational measures," Computers & Education, **107**(Supplement C), 54–67, 2017, doi:10.1016/j.compedu.2016.12.010.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 306-319 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Segmentation of Stocks: Dynamic Dimensioning and Space Allocation, using an Algorithm based on Consumption Policy, Case Study

Anas Laassiri*, Abdelfettah Sedqui

Innovative Technologies Laboratory, Abdelmalek Essaadi University, ENSAT, Tangier, 90000, Morocco

ARTICLE INFO

Article history:
Received: 07 June, 2021
Accepted: 08 August, 2021
Online: 16 August, 2021

Keywords:
Stock management
Warehousing
Dynamic dimensioning
Storage location assignment
Bin packing

ABSTRACT

This paper addresses the stock management aspect. Through this work, we provide a dynamic model of dimensioning and allocation of stocks to storage location for the automotive industry field. This model takes into consideration all constraints of the supply chain (24 constraints) from the suppliers passing by production, storage up to customer delivery and transport. At the end of this paper we will be able to specify the stock replenishment policy, particularly the definition of stock alerts (minimum, nominal, maximum) in quantity and days of stock, in space occupied, and in financial value. These stock thresholds will be integrated in material resource planning, storage allocation procedure and financial budget follow-up. The tool developed is decision-making support for logisticians. The algorithm proposed has solved a real instance and ensure a balanced stock fill rate (99%) in 1200 seconds.

1. Introduction

In every supply chain, the total costs that affect the logistic cost center are very high compared to all production costs. This has a direct impact on companies' margin, and therefore, on the competitiveness of the products in the market knowing that a super high selling price cannot be accepted by all customers. In this context, companies are trying to set a demarche to control its logistics costs. The general framework of this work concerns a key element for the success of this demarche; we are talking about the immobilized cash (stocks) as well as the charges related to the surface occupied by these stocks. Inventories are the basic and main logistical data for making decisions, which can affect the operational, tactical, and sometimes strategic level. Moreover, defining a replenishment policy is fixing a minimum, average, and maximum thresholds, this implies the necessary surface area and the allocation of products to the available storage surface. The main goal is to guarantee cost control, align with budget and at the same time simplify and streamline the internal physical flow. Therefore, inventory management becomes a compulsory.

In the literature, stock management arouses the interest of researchers and manufacturers who have defined procurement policies for all contexts and types of products. In this extended paper of research, originally presented in the 5th international conference on logistics operations management (GOL'20) [1], we will define, by considering the constraints of logistics management, a dynamic stock model which will be used to define stock thresholds and serves as a decision support tool and facilitate the allocation of products in the warehouse. We will define the minimum and maximum stock alarms as well as the corresponding storage area in order to better manage both space and the cash of the company and to follow the perspectives announced in the conference paper presented in the 5th international conference on logistics operations management (GOL'20), regarding the storage location assignment problem which can be modeled as a bin packing problem based on segmentation phase results.

The rest of this article will be organized as follows: we will mention different stock policies in the state of the art. Then, we will detail the constraints of logistic management which formulate our problem of stock policy definition in addition to the assignment of stocks demarche, develop this model on Excel VBA. Finally, a real case study from an industrial background will be studied (Moroccan automotive company).

2. State of art

The stock management in a literature point of view depicts a multitude of replenishment policies. The most addressed are the reorder point methods (Q, R) (figure 1, part 1 taken from [2]), this demarche assumes that within a lead-time of a product, the quantity ordered should be equal to the total quantity consumed. The second method was introduced by [3] and controls stocks through up to level ordering intervals (T, S) (figure 1, part 2 taken from [2]), the Third method defines the economic order quantity

*Corresponding Author: Anas LAASSIRI, laassiri.anas@gmail.com

https://dx.doi.org/10.25046/aj060434

(EOQ) (figure 1, part 3) as part of Wilson model for inventory management [4] considering the ordering and holding cost. Supposing that the ordering cost and delivery time is constant, moreover, the minimum stock alert is considered null, and the maximum stock that can be reached is equal to the EOQ.

There are other policies, based on the ECR: efficient consumer response. This model proposes a collaborative policy in an efficient way between the manufacturers and distributors to boost the gains and achieve high level of productivity from the supplier up to the consumer among of them we can address:

- Continuous replenishment planning (CRP), it is a base that supports the efficient consumer response, which refers to a program that triggers the production and the flow of a product through the supply chain as soon as an identical product have been purchased by the end consumer. As shown by [5] brings varying benefits in terms of inventory cost savings.
- SSM (GPA in French) which means the shared supply management and includes two types: vendor managed inventory (VMI) with or without consignment, the co-managed inventory (CMI).
- The consignment stock (CS) which is an innovative approach to manage inventories in which the vendor removes his inventory and maintains a stock of materials at the buyer's plant. The reference [6] treats the consignment stock policy considering economical and logistics constraints and define the minimum and maximum stock levels of a vendor managed inventory.
- We find also the deterministic and stochastic demand assumption, and stock out assumptions if we lose the customer or there is a stock out cost to pay or assume.

To the best of our knowledge, none of the models addressed in the literature, have treated the constraints of inventory management that we will approach all together.

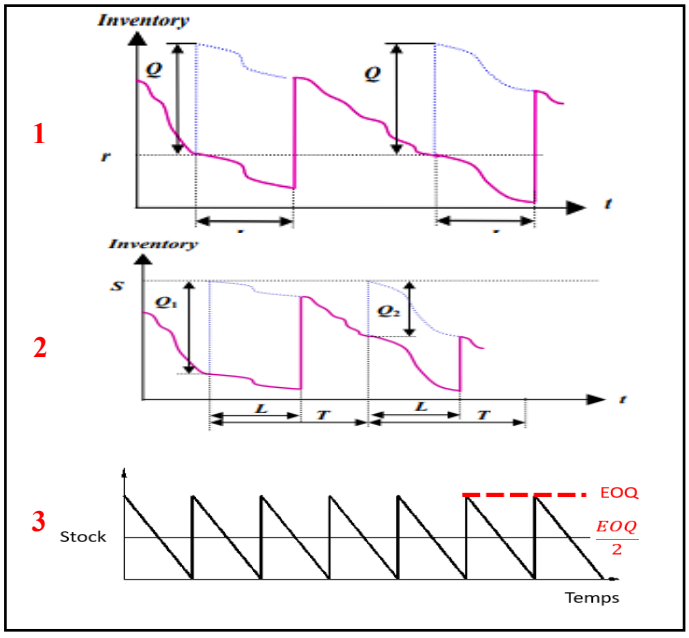


Figure 1: Part 1 (Q, R) policy, part 2 (T, S) policy and part 3 (EOQ) policy

3. Problematic

In an automotive industry environment, between the most uncertain goals to attain we can cite inventory objective level; a supply manager is facing difficulties to determine the policy of procurement in such an uncertain environment. Defining the minimum and maximum replenishment level to ensure production and customer needs, and meet the financial targets set by the top management. This target is calculated in function of the turnover and working days and without integrating the real logistical constraints. To avoid the problem of shortage, the stock managers are obliged to put more and more days in stock, for most cases in a non-efficient way. The model we propose will stands for a decision making support tool that assists stock managers, to determine fixed and variable stock parameters: in order to calculate the minimum and maximum alerts of stock in function of the EOQ. These parameters will be integrated in the material resource planning (MRP) which involves the calculation of requirements, but in an efficient and dynamic way, considering also the main variables impacting the EOQ : type of stocks, the daily average consumption and other vital parameters such as: transit time, minimum ordered quantity. Then assigning the defined stock targets to the storage location areas.

3.1. Objectives

By means of this research, we intent to:

- Define the minimum, nominal, maximum stock alerts, in value and in quantity and in days of stock taking into account the various parameters of stock.
- Define the correct storage area for products in the warehouse in function of the stock alerts defined: dynamic storage.
- Assign products to storage location by maximizing the net surface used while minimizing the distance between the same families of product (multi-purpose).
- Check the correspondence of the defined stock with the available surface.
- Guarantee the Fluidity of goods through the different process of the supply chain.

4. Segmentation of stock: dynamic model

4.1. Constraints of management

In a supply chain, the flow from the supplier up to the customer involves many parameters of management specific to each stage or part of this chain; these parameters directly affect the policy of stock management and make it complicated to define.

If we take the procurement phase as an example, the ordered quantity by the purchaser must abide by the supplier terms defined in the supplier schedule agreement, particularly the minimum batch supply. The ordered quantity should be greater to the minimum batch defined by the supplier to maintain the price defined in the contract; otherwise, a new price will be applied because of the new batch order which will be greater than the defined one. In the other hand, if the gap between the two quantities is really huge compared to the weekly average consumption, the purchaser have to assume the gap in stock even if it will not be consumed immediately, moreover, he has to reserve a space for this gap while waiting for its consumption.

This kind of constraints can have less impact in some cases, for example if the daily average consumption is equal or is a multiple of the supplier batch size, in this case we can order only the required quantity and respect the targeted days of coverage set by the logistic manager. To deal with these constraints and to define the stock management policy, we must start by listing and classifying those constraints all over the supply chain, then taking them into consideration while dimensioning the stocks.

The identified constraints in a supply chain while managing the procurement or the sale of products are defined below and was collected from the industrial sector (several meetings were conducted with logistic managers, engineering pilots, purchasing teams were done in order to regroup all the constraints):

For the raw materials we define:

- Packaging: define the parts in one package, it can be expressed in Kilograms, Meters or Pieces; this data could be found in the logistic contract of each product.
- Batch supply: minimum quantity ordered from a supplier with a normal price.
- Transport Frequency: procurement delivery frequency and represents the number of inbounds in a period (week, two weeks, one month) divided by the period expressed in days.
- Scrap Rate: rejection due to the process (production) and all rejected parts by quality controllers
- DPM: defective parts per million due to the supplier quality of delivered pieces.
- Supplier service rate in full: quantity of parts received compared to the ordered quantity in the firm horizon.
- Supplier service rate on time: quantity of supplies on time divided by total quantity supplied in a specific period.
- Production Reliability rate: the volume declared in production divided by total volume scheduled by logistics.
- Daily average consumption: average consumption of a product due to production and related to the explosion of bill of materials.
- Supplier Firm lead time: number of working days separating the day following the order and the day of the delivery.
- Supplies Frequency: the procurement program update frequency
- Supplier Transit Time: duration in working days between the supplier's plant and the customer point of delivery.

Price: unit price of the product in monetary value.

- Stock for Specific issues: safety stock resulting of a managerial Decision or imposed by the supply policy.
- Packaging details: dimensions authorized stacking level.

For the finished goods we define:

- Packaging: define the parts in one package.
- Minimum production batch size: minimum quantity to launch in production lines.
- Delivery Frequency: customer delivery frequency, it represents the number of trucks loaded to deliver in a period divided by the period expressed in days.
- Rejection Rate and Scrap rate: scrap rate at the end of production and rejected parts following a quality control.
- Production Reliability Rate: the ratio between the volumes produced and the planned volumes.
- Service rate on time in full: the ratio between deliveries on time and the total expected deliveries on a specific period.
- Customer DPM: Defective parts per million delivered to the customer.
- Customer variability: maximum tolerated variability of delivery instructions (firm orders compared to forecast)
- Daily average deliveries. : the mean daily deliveries.
- Customer firm lead time: number of working days separating the day following the release of orders, and the expected day of the delivery.

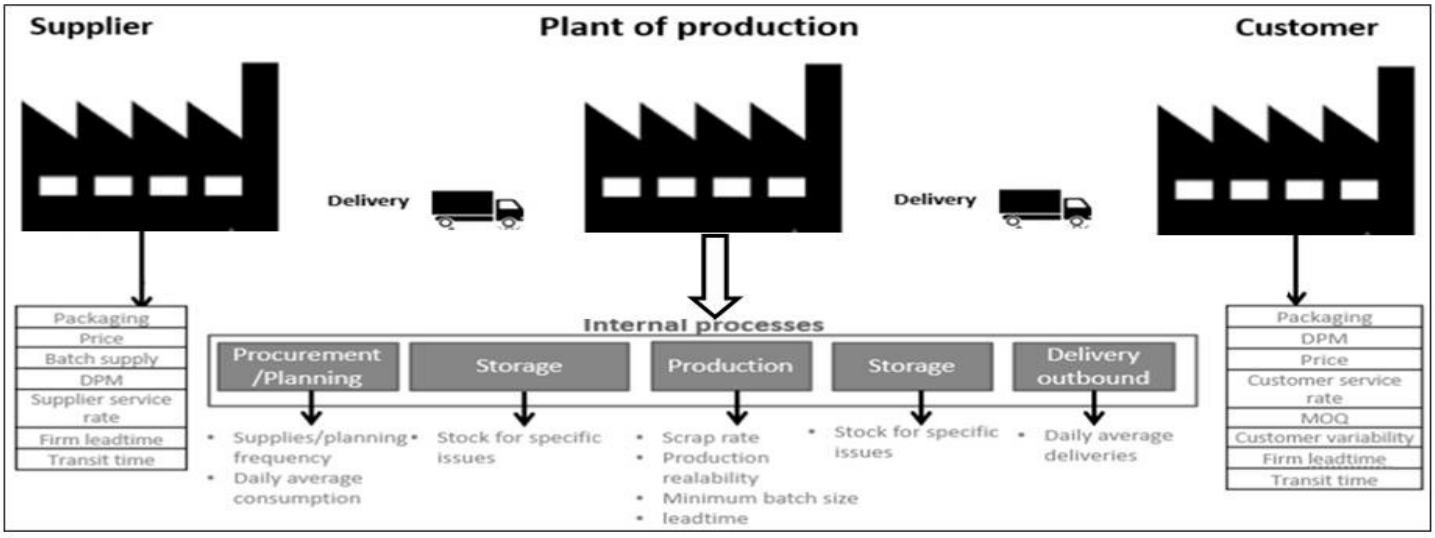


Figure 2: Constraints of management of products in a supply chain

- Planning frequency: the master Production Schedule update frequency.
- Production lead time: time between the Raw Materials out of stock and Finished Goods produced. Must be expressed in working days.
- Transit time: the time between the warehouse and the agreed delivery place.
- Stock for specific issues: safety stock due to management or customer decision.
- Price: unit price in monetary value.

4.2. Proposed model of stock

The operating stocks in the model proposed are defined as follows:

- *Static stock*

For raw materials / finished goods:

- The stock covering the non-respect of the production schedule.
- The stock covering the non-quality, non-compliance, rejects.
- The defined stock for specific issues: decision of top management committee or customer.

- *Dynamic stock*

For raw materials:

- The stock due to deliveries, in function of supply delivery frequency.
- The stock covering the gap between the order batch and the delivery frequency in case they are not synchronized.
- The stock due to variability covering the non-respect of the production schedule.

For finished goods:

- The stock due to deliveries, which depends on the customer delivery frequency.
- The stock covering the gap between the production schedule and the delivery frequency in case they are not synchronized.
- The stock covering the variability of customer orders.

- *Stock alerts*

- Minimum stock: stands for the stock related to total service rate and quality level, in addition to the stock for specific issues. This stock includes internal constraints related to both production and quality processes. It will integrated in the master production schedule MPS and supply program SP as the minimum alert to launch either the replenishment of raw materials, or production orders to maintain the level of stock for finished goods.
- Nominal stock: stock covering all the constraints mentioned before and ensures the availability of material for production, the total stock hovers around this nominal stock.
- Maximum stock: represents the nominal stock added to stock due to the supplier batch size (for raw materials) and production batch (for finished goods) in addition to the stock due to variability. This maximum stock will be integrated also in the MPS/SP program and stands for the upper limit of total stock. This maximum stock could be reached and tolerated by the top management in these two cases :
- While receiving a new inbound batch (new reception).
- Before an outbound delivery (stock assigned to the temporary preparation area/ Bin waiting to be loaded in customer truck).

- *Dynamic surfaces:*

Dynamic surfaces have a direct link with the minimum, nominal and maximum stock, and are generated to follow the variability of these stocks:

- Minimum surface: stands for the minimum space that would accommodate the minimum stock defined expressed in handling units (number of pallets for example), integrating the unit area of one handling unit and the possible stacking level as well as the type of storage (flat or shelving).
- Nominal surface: stands for the nominal space that would accommodate the nominal level of stock defined expressed in handling units, integrating the unit area of one handling unit and the possible stacking level as well as the type of storage (flat or shelving).
- Maximum surface: stands for the maximum area that would accommodate the maximum stock defined expressed in handling units, integrating the unit surface of the handling unit and the number of stacking levels as well as the type of storage space (flat, which means massive or shelving using metallic structures).

- *Storage location allocation:*

The problem of allocating materials to the storage location area can be modeled as a task assignment problem as schematized below:

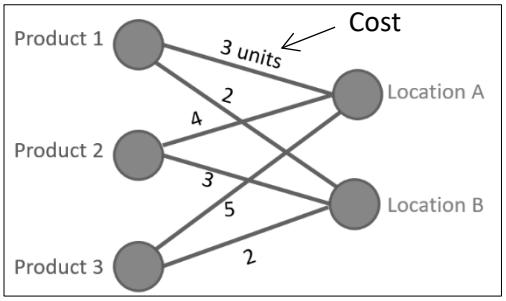


Figure 3: modeling of task assignment problem

As shown in the figure above, we will assign products (represented with handling units) to the storage location, which is divided to sub storage location named bins. Our problem can also be considered as a bin packing problem. It is about finding the most economical storage possible for a set of items into bins.

In our case, it is about finding the closest optimal assignment for a set of products that are assigned to storage location. Respecting lean flow directives (raw materials and finished goods should be stored separately to avoid flow congestion). The storage locations will be represented by boxes (Bins) and the products will be represented by items. To solve this NP hard problem we will use tree dimensional Bin packing algorithm, and to improve results we will pair the bin packing algorithm to an algorithm of list (first fit decreasing problem [7].

The table below summarizes the analogies between our assignment problem and the core model of bin packing problem modified from [8] considering the 3 dimensions and the constraints related to our context :

Table 1: Analogies between bin packing and our problem

Criteria	**Bin packing problem**	**Stock allocation problem**
Data	Article Bin Volume of the article	products Storage location Dimensions of the product
Objective function	Assigning items to bins	Assign stocks to volume
Constraints	The volume capacity of the bin	The volume of the storage location
Objective	Minimize the number of bins used	Minimize the volume used
Other		Compatibility constraint between raw materials and final products, they must be separated to better flow organization, and between bins and products in function of storage type

4.3. Parameters and Equations

- Notation

Table 2: Stock parameters

Raw Materials	
Pck_i	Parts per packaging of the raw material (RM) i
B_i	Batch Supply of the RM i
$Sfreq_i$	Supply Frequency of the RM i
S_i	Scrap Rate of the RM i
DPM_i	DPM of the RM i
Sf_i	Supplier service rate in full of the RM i
St_i	Supplier service rate on time of the RM i
PR_i	Production Reliability rate of the RM i
DAC_i	Daily average consumption of the RM i
F_i	Supplier Firm lead time of the RM i
$Freq_i$	Supplies Frequency of the RM i
TT_i	Supplier Transit Time of the RM i
SSI_i	Stock for Specific issues of the RM i
p_i	Price of the RM i
L_i	Length of the RM i
l_i	Width of the RM i
G_i	Stacking level of the RM i
Finished Goods	
Pck_j	Packaging of the finished good (FG) j
B_j	Minimum production batch size of the FG j
$Dfreq_j$	Delivery Frequency of the FG j
S_j	Rejection Rate and Scrap rate of the FG j
PR_j	Production Reliability Rate of the FG j
Sr_j	Service rate on time in full of the FG j
Var_j	Customer variability of the FG j
DAD_j	Daily average deliveries of the FG j
C_j	Customer firm lead time of the FG j
$Plan_j$	Planning frequency of the FG j
PLT_j	Production lead time of the FG j
DPM_j	DPM of the FG j
SSI_j	Stock for specific issues of the FG j
TT_j	Customer Transit time of the FG j
p_j	Price of the FG j
L_j	Length of the FG j
l_j	Width of the FG j
G_j	Stacking level of the FG j
DAC	Daily average consumption
	Stock alerts in quantity
$Minstock_k$	Minimum stock of the part number k
$Nomstock_k$	Nominal stock of the part number k
$Maxstock_k$	Maximum stock the part number k
	Dynamic surface
l_k	Length of the part number k
w_k	Width of the part number k
$Hustock_k$	Number of packages in part in a handling unit for the part number k
$Stacking_k$	Authorized stacking level for the part number k
$Minar_k$	Minimum surface for the part number k
$Momar_k$	Nominal surface for the part number k
$Maxar_k$	Maximum surface for the part number k
	Stock alerts value
$MinstockV_k$	Minimum stock of the part number k
$NomstockV_k$	Nominal stock of the part number k
$MaxstockV_k$	Maximum stock the part number k
$Cost_k$	Unit cost of the part number k

The graphic bellow synthetizes the static and dynamic stocks taken into consideration in our model:

- ✓ Dynamic stock: represents the level of stock made available to the planner for smoothing.
- ✓ Static stock: represents the incompressible level of stock.

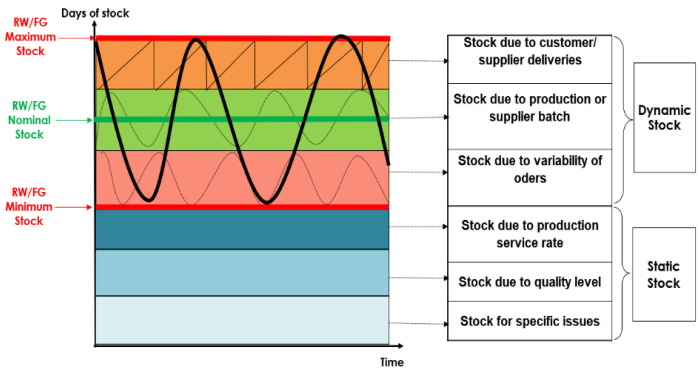


Figure 4: Static and Dynamic Stocks

- *Static stock*

RM

$$(S_i + DPM_i) \times \sqrt{\frac{F_i}{Freq_i}} \times Freq_i \quad (1)$$

$$1 - Sf_i \times St_i \times \sqrt{\frac{F_i}{Freq_i}} \quad (2)$$

$$SSI_i \quad (3)$$

FG

$$(S_j + DPM_j) \times \sqrt{\frac{C_j}{Plan_j}} \times Plan_j \quad (4)$$

$$1 - PR_j \times Sr_j \times \sqrt{\frac{C_j}{Plan_j}} \quad (5)$$

$$SSI_j \quad (6)$$

- *Dynamic stock:*

RM

$$\frac{Freq_i}{2} + TT_i \quad (7)$$

$$(1 - PR_i) \times \sqrt{\frac{F_i}{Freq_i}} \times Freq_i \quad (8)$$

$$\frac{B_i}{DAC_i \times 2} + TT_i - \left(\frac{Freq_i}{2} + TT_i\right) - (S_i + \frac{DPM_i}{1\,000\,000}) \times \sqrt{\frac{F_i}{Freq_i}} \times \sqrt{\frac{F_i}{Freq_i}} \times Freq_i - \left(1 - Sf_i \times St_i \times \sqrt{\frac{F_i}{Freq_i}}\right) - (1 - PR_i) \times \sqrt{\frac{F_i}{Freq_i}} \times Freq_i \quad (9)$$

FG

$$\frac{DFreq_j}{2} + TT_j \quad (10)$$

$$Var_j \times \sqrt{\frac{C_j}{Plan_j}} \times Plan_j \quad (11)$$

$$\frac{B_j}{DAC_j \times 2} + TT_j - \left(\frac{DFreq_j}{2} + TT_j\right) - (S_j + DPM_j) \times \sqrt{\frac{C_j}{Plan_j}} \times Plan_j - (1 - PR_j \times Sr_j \times \sqrt{\frac{C_j}{Plan_j}}) - Var_j \times \sqrt{\frac{C_j}{Plan_j}} \times Plan_j \quad (12)$$

- ✓ (1): Stock due to quality level.
- ✓ (2): Stock due to production reliability.
- ✓ (3): Stock for specific issues.
- ✓ (4): Stock due to quality level.
- ✓ (5): Stock due to production reliability.
- ✓ (6): Stock for specific issues.
- ✓ (7): Stock due to supply frequency.
- ✓ (8): Stock due to production variability.
- ✓ (9): Stock due to supplier batch.
- ✓ (10): Stock due to delivery frequency.
- ✓ (11): Stock due to customer variability.
- ✓ (12): Stock due to production batch.

- *Stock alerts in quantity:*

For RM:

$$Minstock_k = (S_i + DPM_i) \times \sqrt{\frac{F_i}{Freq_i}} \times Freq_i + 1 - Sf_i \times St_i \times \sqrt{\frac{F_i}{Freq_i}} + SSI_i \quad (13)$$

$$Nomstock_k = Minstock_k + \frac{Freq_i}{2} + TT_i + (1 - PR_i) \times \sqrt{\frac{F_i}{Freq_i}} \times Freq_i + \frac{B_i}{DAC_i \times 2} + TT_i - \left(\frac{Freq_i}{2} + TT_i\right) - (S_i + \frac{DPM_i}{1\,000\,000}) \times \sqrt{\frac{F_i}{Freq_i}} \times \sqrt{\frac{F_i}{Freq_i}} \times Freq_i - (1 - Sf_i \times St_i \times \sqrt{\frac{F_i}{Freq_i}}) - (1 - PR_i) \times \sqrt{\frac{F_i}{Freq_i}} \times Freq_i \quad (14)$$

$$Maxstock_k = SSI_i + (S_i + DPM_i) \times \sqrt{\frac{F_i}{Freq_i}} \times Freq_i + 1 - Sf_i \times St_i \times \sqrt{\frac{F_i}{Freq_i}} + (1 - PR_i) \times \sqrt{\frac{F_i}{Freq_i}} \times Freq_i + TT_i + Max\left(Freq_i, \frac{B_i}{DAC_i}\right) \quad (15)$$

- *Stock alerts in value:*

$$MinstockV_k = Minstock_k \times Cost_k \quad (16)$$

$$NomstockV_k = Nomstock_k \times Cost_k \quad (17)$$

$$MaxstockV_k = Maxstock_k \times Cost_k \quad (18)$$

- *Dynamic surface:*

$$Minar_k = Minstock_k \times (Hustock_k \div Stacking_k) \times l_k \times w_k \quad (19)$$

$$\mathrm{Nomar_k} = \mathrm{Nomstock_k} \times (Hustock_k \div \mathrm{Stacking_k}) \times \mathrm{l_k} \times \mathrm{w_k}) \quad (20)$$

$$\mathrm{Maxar_k} = \mathrm{Maxstock_k} \times (Hustock_k \div \mathrm{Stacking_k}) \times \mathrm{l_k} \times \mathrm{w_k}) \quad (21)$$

- *Allocation problem: bin packing model*

The bin packing problem is NP hard in strong sense, this has been demonstrated using the reduction with 2-Partition, to solve medium to large instances we need to adopt heuristics developed particularly to solve this problem. The most popular heuristics in literature are first fit, Next fit and best fit, which have proven its effectiveness. The most adapted heuristic to our context is First fit, since the articles are assigned in a given order (in our case the order is in function of the daily average consumption). Then each article is placed in the first bin that can contain it.
A new box is considered only if this article does not fit in any box (considering the remaining empty volume in each box).The box remains open until the end of the execution of the algorithm, in case we engage the constraints of compatibility between articles, if an article is not compatible with a bin or an article or with another article already assigned in the current box, it will be assigned to the next box that ensure compatibility first then ensure the volume constraint contrary to the next fit heuristic where box is closed permanently if the volume of the current article cannot be contained in the box. The box is permanently closed, and it is impossible to assign a new article even if the remaining empty volume is enough to contain a new article. A new box is considered and becomes the current box. The reference [9] have proven the result of First fit:

$$First\ Fit\ (O) \leq Optimal(O) + 2\ \forall O \quad (22)$$

Research in the article [10] have demonstrated that the sequential coupling of list algorithms and bin packing heuristics guarantee better results, the first fit decreasing is the one that provides the best results, at a factor 2 of the optimal as proved by the reference [2]. Our assignment problem involves the allocation of stocks (total handling units) to the storage location (divided into sub-storage locations) with a flow constraint, that the raw materials and final products should be stored separately to remain an organized physical flow. It serves as a simulation tool to check if the policy of replenishment proposed sticks with the space that we have in the warehouse considering that the layout and the storage location are already defined.

Next, we will adopt the first fit decreasing heuristic to resolve the storage location assigning problem based on the daily average consumption of the handling unit, which refers to the daily consumption of the product marked in the HU label, sorted in the descending order (the HU of high runner products will be assigned first). We will use the following formalization in addition to some modifications related to our context:

- i: The index that represents the handling unit i.
- j: The index that represents the storage location bin j.
- S: The set of storage location bins in the warehouse.
- HU: The set of handling units.
- V_i: The volume occupied by the handling unit i (3D).
- V_j : The volume of the bin j.
- Dac_i: The daily average consumption of the handling unit i.

The decision variables are described below:

$$x_{i,j} = \begin{cases} 1, if\ the\ Hu\ i \in HU\ is\ assigned\ to\ the\ storage\ bin\ j \in S \\ 0, else \end{cases}$$

$$C_{i,j} = \begin{cases} 1, if\ the\ HU\ i \in HU\ is\ compatible\ with\ the\ Bin\ j \in S \\ 0, else \end{cases}$$

The algorithm bellow describes the first fit decreasing algorithm adopted which is the sequential coupling of the list algorithm based on the consumption of productions and the first fit algorithm:

Algorithm 1: The principle of the first fit decreasing algorithm

Result: $A_j \forall j \in S$

Input data:
- HU, $i \in \{1, \dots, \mathrm{card(HU)}\}$, V_i , Dac_i $\forall \mathrm{i} \in \mathrm{HU}$
- S , $j \in \{1, \dots, \mathrm{card(S)}\}$, V_j , $\forall j \in S$
- $C_{i,\mathrm{j}}$, $\forall$ j $\in$ S

Initialization:
- A_j =0, the initial volume of the bin j $\forall$j$\in$ S, all bins is empty at the beginning.
- $Chart\ A_{(3,card(HU))}$, the chart with card (HU) columns and 3 lines that contains respectively: Hu_i, V_i, Dac_i
- $Chart\ B_{(3,Card(HU))}$, the chart with 3 columns that contains: Hu_i, V_i, Dac_i after the step of sorting.

Sort $Dac_i, \forall i \in HU$ in a descending order, then place them with the corresponding HU and volume in $Chart\ B$

for k = 1 ... $card(HU)$ **do**
  j : = 1, Assigned: = false
  while *(assigned= false)* & $j \leq$ card(S)) **do**
    if the HU $Chart\ B_{1,k}$ is compatible with the bin j ($C_{chart\ B_{(1,k)},j} = 1$) **then**
      if $Chart\ B_{(2,k)}$ holds in j, ($A_j + Chart\ B_{(2,k)} < V_j$) **then**
        Assign $Chart\ B_{(1,k)}$ to the bin j:
        $x_{Chart\ B_{(1,k)},j} = 1$
        $A_j = A_j + Tab\ B_{(2,k)}$
        Assigned: = True
        j : = j + 1
      end
    end
  end
end

4.4. Resolution method: Segmentation and assignment of stocks

The demarche of segmentation starts from the integration of input data, until the calculation of the stock alerts: in quantity, in number of packages (HU), in financial value and in square meters.

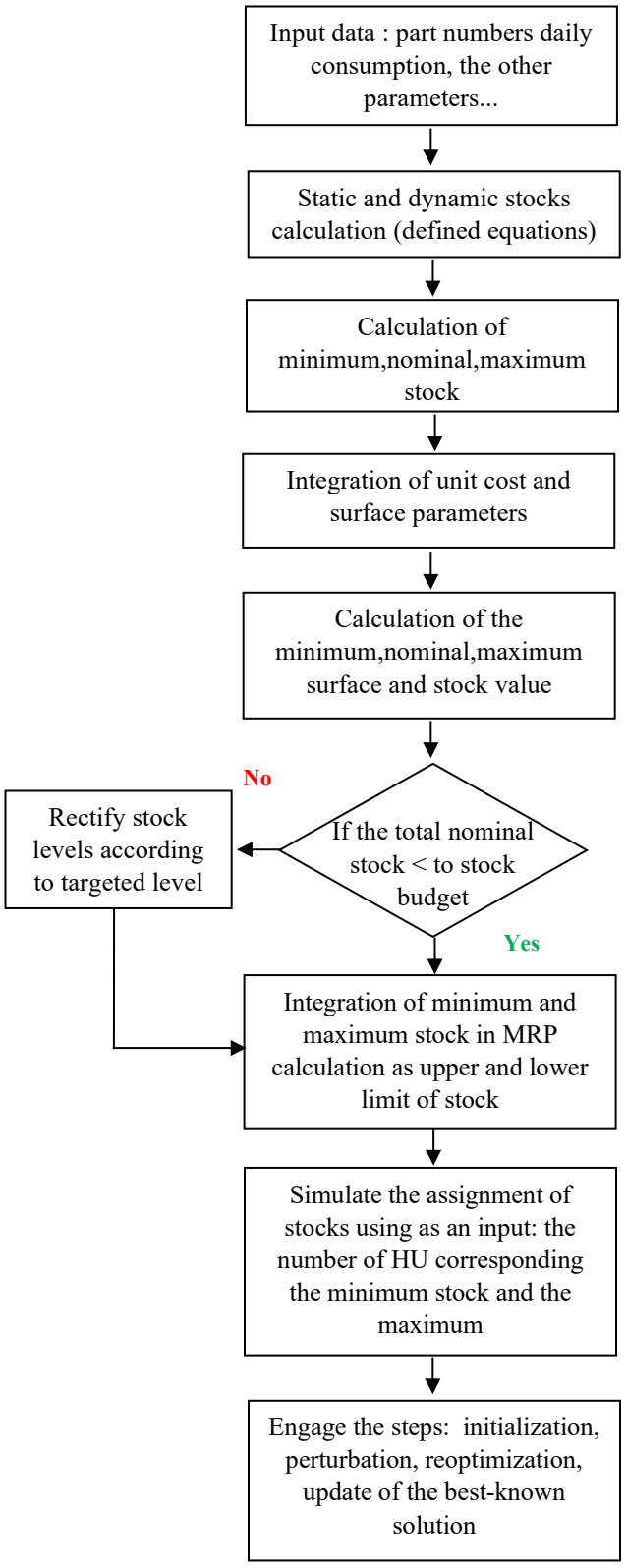


Figure 5: Segmentation and assignement demarche

The output of the segmentation demarche will be the input data for the storage location assignment demarche, the demarche starts with the initialization (input data), then engaging the first fit decreasing algorithm for the construction step (best known solution) based on the daily average consumption. Next, we will use a large neighborhood search as described below:

- Step 1 (perturbation): we randomly remove handling units from bins, and randomly empty a few bins completely. Then we sort the bins in a decreasing order in function of the current volume occupied in each one A_j.
- Step 2 (optimization): we select the removed HU, and reassign them into bin using a constructive heuristic, in our case we have chosen the wall building heuristic because of the flow constraints, we should place similar HUs next to each other as if we build a brick wall.
- Step 3 (update of the best-known solution): if the new solution obtained is better than the best known solution, we replace the current best known solution, if not, we go back to the first step and engage a new cycle of perturbation, reoptimization.

  The demarche of stock segmentation and storage location assignment is described below:

The action plan mentioned in the segmentation demarche tends to optimize six variables:

- Stock delivering (adapting the frequencies of delivery to minimize the number of days of finished good stock).
- Batch of production or supply (to synchronize the daily average consumption or daily average delivery).
- Variability of the demand (that must be contractual).
- Service Rate (of both customer and supplier).
- Quality level (of both supplier and customer delivered parts) an finally the safety stock.

For the case study, we will use the script code developed by Güneş Erdoğan (CLP Spreadsheet Solver) with some modification related to our context.

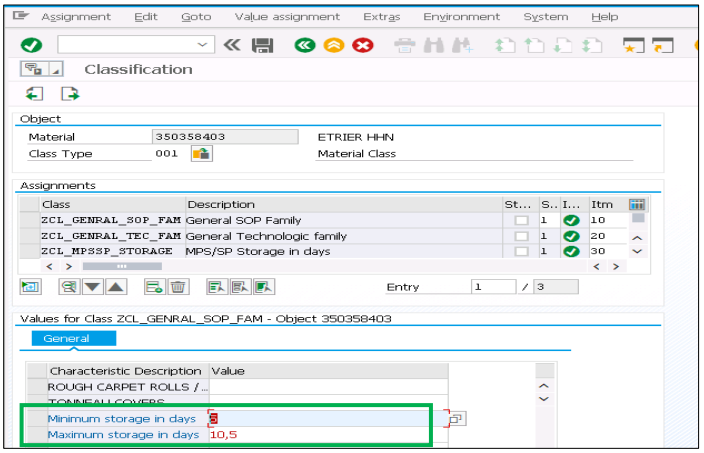


Figure 6: Master data transaction (classification tab)

							Av/Re	W13						W14	W15
Part number	Description	Minimum Batch size	Alarm Min.	Alarm Max.	Type of info.	Stock (SKU)	<=28/03	29/03	30/03	31/03	01/04	02/04	03/04	05/04 - 11/04	12/04 - 18/04
350658746	PART A	400	0	11,5	1.Stock Quantity	801,3	801	801	801	801	467	467	467	533	599
350658746	PART A	400			1.Stock Cover		14,4	13,4	12,4	11,4	10,4	9,4	8,4	9,94	11,31
350658746	PART A	400			1.Stock Value	76 235	76 235	76 235	76 235	76 235	44 462	44 462	44 462	50 745	57 028
350658746	PART A	400			2.Requirement						334			334	334
350658746	PART A	400			7.Purchase Init. Qty									400	400
350658746	PART A	400			7.Purchase /PU									1	1
350658746	PART A	400			7.Purchase New Qty									400	400

							Av/Re	W13					W14		
Part number	Part description	Minimum lot size	Alarm Min.	Alarm Max.	Type of info.	Stock (SKU)	<=30/03	31/03	01/04	02/04	03/04	05/04	06/04	07/04
351264852	Part B	120	2	11	1.Stock Quantity	1320	1 560	1 800	1 320	1 440	1 560	1 800	2 040	2 280
351264852	Part B	120			1.Stock Cover		10,6	10,8	11,6	11,2	10,8	11	10,86	10,71
351264852	Part B	120			1.Stock Value	51 249	60 567	69 885	51 249	55 908	60 567	69 885	79 202	88 520
351264852	Part B	120			2.Requirement				840					
351264852	Part B	120			3.Prod. init. Qty		240	240	360	120	120	240	240	240
351264852	Part B	120			3.Prod. /PU		2	2	3	1	1	2	2	2
351264852	Part B	120			3.Prod. New Qty		240	240	360	120	120	240	240	240

Figure 7: SP (1) and MPS (2)

4.5. Integration of alerts in MRP calculation

After the calculation of minimum and maximum alerts of stock, we integrate theme in SAP following the steps below:

- ✓ Access to the transaction of master data.
- ✓ Choose the classification tab.
- ✓ Add the alerts of stock in the corresponding field.

This data in the classification tab will be used to run the Material Resource Planning in automatic mode (MRP job) for the frozen period or Manuel smoothing of orders for the forecast horizon.

Our alerts were integrated in the supply VBA program (SP) and Master Production Schedule VBA program (MPS) already developed by the company using EXCEL VBA. The datasheet of SP includes the following data (Figure 7):

- The stock quantity and cover and value extracted from SAP.
- The requirement of the part in question into the frozen and forecast horizon (named "Part A" in the figure 7).
- The cover of the part in question in the end of the period (day, week).
- The purchase quantity in function of the batch size or its multiples.
- In each period (day, week) the quantity proposed by the system ensures that the coverage after acquisition of this quantity is between the minimum and maximum alert.
- If the cover exceeds the maximum alert of stock, the system doesn't propose any purchase quantity and jump to the next period.

The datasheet of MPS includes the following data (figure 7):

- All the data mentioned in SP.
- It works the same as SP, but instead of purchase quantity, we find the production quantity and instead of supplier batch size we find the production minimum size.

4.6. Limits of the proposed method

- ✓ If there is an error in the input data, consequently the results will not be consistent and right.
- ✓ If there is any discrepancy between the physical and the stock injected in the information system, the quantities proposed by MRP (either planned orders or purchase order) won't ensure the coverage of stock shown in the information system.
- ✓ The alerts of stock should be updated in a regular basis (every three months for example) knowing that the volumes could change in horizon forecast and the daily average consumption must follow the new volumes.
- ✓ The model of assignment considers zero distance between two HUs, and the volume and number of HU are already known before engaging the solver (off-line).

5. Case study: Moroccan industrial compan

4.7. Segmentation of stocks

The segmentation model developed was tested in a real case study conducted using data of a Moroccan multinational company specialized in automotive parts.

We have chosen 5 references of each family to show the details of each parameters and how we can interpret the results obtained using equations and logistic parameters defined before.

After integrating the ten references in the spreadsheet developed, we obtain the result below:

After realizing the simulation with the current data we conclude the following point by family of productions

For raw materials:

- To see the impact of stock due to transit time, stock to cover the frequency of delivery, and stock to covering quality of service let's take an example of the part number A : the nominal stock is equivalent to 8.2 days this stock include also the transit time equivalent to four days.

TOTAL MINIMUM STOCK	TOTAL NOMINAL STOCK	TOTAL MAXIMAL STOCK
in days	in days	in days
1,6	5,9	7,8

TOTAL MINIMUM STOCK	TOTAL NOMINAL STOCK	TOTAL MAXIMAL STOCK
in packaging	in packaging	in packaging
82	343	465

REFERENCE	DESIGNATION	Customer	Packaging	Production Batch	Transport Frequency	Quality Level		Demand Variability			Daily consumption	Firm lead time	Planning frequency	Production lead time+Transit time	Stock for specific issues	Stock due to delivering	Stock due to production batch	Stock due to quality level	Stock due to Service Rate	Stock due to customer variability	MINIMUM STOCK	NOMINAL STOCK	MAXIMAL STOCK	MINIMUM STOCK	NOMINAL STOCK	MAXIMAL STOCK	MINIMUM STOCK	NOMINAL STOCK	MAXIMAL STOCK
			in parts	in parts	in days	scrap rate	Customer PPM	Production Reliability %	Service rate on time in full	Customer variability	in parts	in days	in days	in days	in days	in days	in days	in days	in days	in days	in days	in days	in days	in qty	in qty	in qty	in packaging	in packaging	in packaging
350103233	Ref G	xxxxx	120	240	2,50	3,0%	80	60,00%	97,00%	10%	2200	6	5	1,5		2,8	0,0	0,2	2,3	0,5	2,5	5,8	7,0	5 399	12 654	15 404	45,0	105,5	128,4
350107523	Ref H	xxxxx	100	100	5,00	3,0%	120	98,00%	97,00%	20%	2000	6	5	1,5	1,0	4,0	0,0	0,2	0,3	1,1	1,4	6,5	9,0	2 871	13 062	18 062	28,7	130,6	180,6
300438218	Ref I	xxxxx	180	180	5,00	3,0%	100	98,00%	97,00%	20%	1150	6	5	1,5		4,0	0,0	0,2	0,3	1,1	0,4	5,5	8,0	501	6 361	9 236	2,8	35,3	51,3
300453929	Ref J	xxxxx	100	100	5,00	3,0%	100	98,00%	97,00%	20%	820	6	5	1,5		4,0	0,0	0,2	0,3	1,1	0,4	5,5	8,0	357	4 535	6 585	3,6	45,4	65,9
313008732	Ref K	xxxxx	100	200	5,00	3,0%	100	98,00%	97,00%	20%	480	6	5	1,5		4,0	0,0	0,2	0,3	1,1	0,4	5,5	8,0	209	2 655	3 855	2,1	26,5	38,5

MINIMUM STOCK	NOMINAL STOCK	MAXIMAL STOCK
in value	in value	in value
191 339 MAD	967 442 MAD	1 256 308 MAD

REFERENCE	Unit cost	MINIMUM STOCK	NOMINAL STOCK	MAXIMAL STOCK	PACKAGING Length (L)	PACKAGING Width (w)	PACKAGING High (H)	Stacking	Storage Type (Racks or Flat Storage)	Stock height	Surface per Packaging	MINIMUM SURFACE	NOMINAL SURFACE	MAXIMAL SURFACE	MINIMUM Rack Locations	NOMINAL Rack Locations	MAXIMAL Rack Locations
	in value	in value	in value	in value	in (mm)	in (mm)	in (mm)	in packaging		in m	in m²	in m²	in m²	in m²			
332233112	82	19 603	79 991	116 763	1200	1000	1100	5,0	Racks	6	1				1	2	3
332233122	128	38 296	127 653	200 990	1200	800	1000	5,0	FLAT	5	1	1	1	2			
332233132	58	93 239	190 144	219 273	1200	1000	1200	5,0	FLAT	6	1	5	8	10			
332233142	59	26 525	409 688	542 232	1200	1000	1400	5,0	FLAT	7	1	1	7	10			
332233152	59	13 675	159 966	177 050	1200	1000	1400	5,0	FLAT	7	1	1	4	4			

Figure 8: The datasheet of raw materials

TOTAL MINIMUM STOCK	TOTAL NOMINAL STOCK	TOTAL MAXIMAL STOCK
in days	in days	in days
1,6	5,9	7,8

TOTAL MINIMUM STOCK	TOTAL NOMINAL STOCK	TOTAL MAXIMAL STOCK
in packaging	in packaging	in packaging
82	343	465

REFERENCE	DESIGNATION	Customer	Packaging	Production Batch	Transport Frequency	Quality Level		Demand Variability			Daily consumption	Firm lead time	Planning frequency	Production lead time+Transit time	Stock for specific issues	Stock due to delivering	Stock due to production batch	Stock due to quality level	Stock due to Service Rate	Stock due to customer variability	MINIMUM STOCK	NOMINAL STOCK	MAXIMAL STOCK	MINIMUM STOCK	NOMINAL STOCK	MAXIMAL STOCK	MINIMUM STOCK	NOMINAL STOCK	MAXIMAL STOCK
			in parts	in parts	in days	scrap rate	Customer PPM	Production Reliability %	Service rate on time in full	Customer variability	in parts	in days	in days	in days	in days	in days	in days	in days	in days	in days	in days	in days	in days	in qty	in qty	in qty	in packaging	in packaging	in packaging
350103233	Ref A	xxxxx	120	240	2,50	3,0%	80	60,00%	97,00%	10%	2200	6	5	1,5		2,8	0,0	0,2	2,3	0,5	2,5	5,8	7,0	5 399	12 654	15 404	45	105	128,4
350107523	Ref B	xxxxx	100	100	5,00	3,0%	120	98,00%	97,00%	20%	2000	6	5	1,5	1,0	4,0	0,0	0,2	0,3	1,1	1,4	6,5	9,0	2 871	13 062	18 062	29	131	180,6
300438218	Ref C	xxxxx	180	180	5,00	3,0%	100	98,00%	97,00%	20%	1150	6	5	1,5		4,0	0,0	0,2	0,3	1,1	0,4	5,5	8,0	501	6 361	9 236	3	35	51,3
300453929	Ref D	xxxxx	100	100	5,00	3,0%	100	98,00%	97,00%	20%	820	6	5	1,5		4,0	0,0	0,2	0,3	1,1	0,4	5,5	8,0	357	4 535	6 585	4	45	65,9
313008732	Ref E	xxxxx	100	200	5,00	3,0%	100	98,00%	97,00%	20%	480	6	5	1,5		4,0	0,0	0,2	0,3	1,1	0,4	5,5	8,0	209	2 655	3 855	2	27	38,5

TOTAL MINIMUM STOCK	TOTAL NOMINAL STOCK	TOTAL MAXIMAL STOCK
in value	in value	in value
509 441 MAD	1 795 461 MAD	2 373 334 MAD

REFERENCE	Unit cost	MINIMUM STOCK	NOMINAL STOCK	MAXIMAL STOCK	PACKAGING Length (L)	PACKAGING Width (w)	PACKAGING High (H)	Stacking	Storage Type (Racks or Flat Storage)	Stock height	Surface per Packaging	MINIMUM SURFACE	NOMINAL SURFACE	MAXIMAL SURFACE	MINIMUM Rack Locations	NOMINAL Rack Locations	MAXIMAL Rack Locations
	in value	in value	in value	in value	in (mm)	in (mm)	in (mm)	in packaging		in m	in m²	in m²	in m²	in m²			
350103233	65,6	354 097,6	829 893,6	1 010 243,9	1200	800	1200	6,0	FLAT	7,2	1,0	7,7	17,3	21,1			
350107523	43,0	123 594,1	562 288,2	777 526,5	1200	800	1200	6,0	Racks	7,2	1,0				29	131	181
300438218	37,1	18 598,6	236 237,0	343 017,8	1200	800	1200	6,0	Racks	7,2	1,0				3	36	52
300453929	36,8	13 151,0	167 041,8	242 546,0	1200	800	1200	6,0	Racks	7,2	1,0				4	46	66
313008732	54,5	11 393,9	144 723,9	210 140,1	1200	800	1200	6,0	Racks	7,2	1,0				3	27	39

Figure 9: The datasheet of final products

- This duration could be reduced if we consider local suppliers of customers, that the distance between the place of order and the place of delivery is closer. This decision could have an impact on the order cost. Then, we have 5.3 days in stock because of the frequency of delivery to customer, if there is a possibility to raise the frequency from 2 trucks to 3 trucks per week, this quantity of stock days will be largely optimized. Consequently, the overall value of stock will be optimized reducing the impact on immobilized cash. Moreover, if we have a contract with suppliers that stipulate a global service rate greater than 90%, the stock related to service rate will be improved (currently 1,7 days). In the other hand, a minimum order quantity synchronized with daily consumption will solve many problems of returns of stocks non-consumed to the warehouse after production. This constraint was taken into consideration while developing the program under Excel VBA, so if we have a non-synchronization between these two values; this means that the resulting rate is greater than one week (generally 5 working days); the program underlines in red the minimum supplier batch. In the current simulation data, the resultant rate doesn't exceed one week or production. Regarding the space of storage required if the storage type required is: RACK (shelving structures), the spreadsheet gives the number of modules equivalent to nominal stock for the part number A is about 2 modules. For the flat type of stock. If the storage type required is: FLAT, the spreadsheet specifies the net surface needed in square meters (for the part number B and C, the surface needed is respectively 4 and 8 square meters). For the

total space needed (includes the stock net space and services aisles space), we apply a mark-up of 2.4.

- Another example of optimization we found, for the reference B. The stock due to the batch supply is 1.3 days, knowing that we consume 60 parts per day, and the batch supply is 800 parts. So, if we reduce the supplier batch, we could gain 1.3 days in stock of this reference.
- For the reference D, if we increase the supply frequency from 1 truck per month to 1 truck per week, we could gain 7.5 days in stock and the stock due to supply frequency will be 20.5 days. Furthermore, if we find a forwarder who can ensure the delivery in less than 18 days, for example 14 days, the stock due to supply frequency will be 16.5 stock days.

For finished goods:

- The important stock for the references H,I,J,K is the stock due to the frequency of customer deliveries, that implies a stock to cover the time between two deliveries, we have 4 days for each reference, if we increase the delivery frequency for example from one truck per week to 2 trucks per week, we could gain 2 days in stock
- The stock due to production service is huge for the reference G, if we oblige the production team to respect the production plan communicated by logistics, we could improve the reliability rate and gain 2.3 days in stock caused by the current production reliability rate.

Considering all this remarks, the action plan to reduce stocks should start by eliminating the important stock mentioned above. The spreadsheet also gives the corresponding space that should be revered for each reference (minimum, nominal and maximum); all these informations will serve as an input data for the assignment phase.

4.8. Assignment of stocks in storage locations

For the assignment phase, we define the following data structure according to the bin packing notation:

$$
Article \begin{cases} - HU(handling\ unit) \begin{cases} - Volume \begin{cases} Length \\ Width \\ Height \end{cases} \\ - Quantity \\ - Type \begin{cases} RW: raw\ materials \\ FG: finished\ goods \end{cases} \end{cases} \\ - DAC(daily\ average\ consumption) \\ - Storage\ type \begin{cases} Flat \\ Rack\ or\ shelving(RCK) \end{cases} \end{cases}
$$

$$
Bin \begin{cases} - Bin\ type \begin{cases} - Flat \begin{cases} - 1 \\ - 2 \\ - .. \end{cases} \\ - RCK \begin{cases} - 1 \\ - 2 \\ - .. \end{cases} \end{cases} \\ - Volume\ capacity \begin{cases} - Lenth \\ - Width \\ - Height \end{cases} \\ - Number\ of\ bins \end{cases}
$$

Figure 10: Simulation data structure

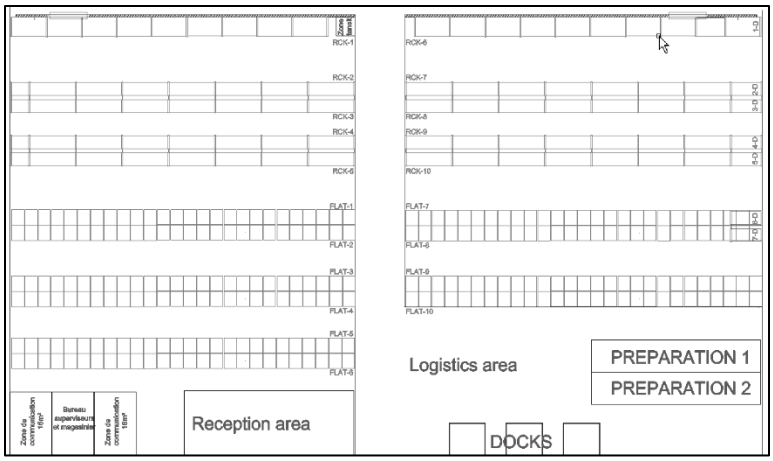


Figure 11: The layout of the warehouse

We consider the following instances:

- Considering the warehouse of the company (figure 11).
- The set of storage bin (Table 3).
- 10 reference of RW and FG (table 4 and 5) as an example. The two family of products are not compatible (can't be stored in the same bin). Compatibility between bins and references (storage type needed).

Table 3 : The set of Bins

Bin Type	Number of BINs	Width in meter (x)	Height in meter (y)	Length in meter(z)
FLAT	10	28	6	1,2
RCK	10	28	6	1,2

Table 4: The set of RW references

RW	Storage type	DAC	Number of HU	Width (meter) (x)	Height (meter) (y)	Length (meter) (z)
RW-1-A	FLAT	120	2	1	1,1	1,2
RW-2-B	RCK	60	5	0,8	1	1,2
RW-3-C	RCK	400	33	1	1,2	1,2
RW-4-D	FLAT	225	28	1	1,4	1,2
RW-5-E	FLAT	116	11	1	1,4	1,2
RW-6-F	RCK	1200	80	1	1,1	1,2
RW-7-G	FLAT	700	56	1	1,1	1,2

RW-8-H	FLAT	865	40	1	1,1	1,2
RW-9-I	FLAT	923	21	1	1,1	1,2
RW-10-J	RCK	1032	84	1	1,1	1,2

Table 5 : The set of FG references

RW	**Storage type**	**DAC**	**Number of HU**	**Width (meter) (x)**	**Height (meter) (y)**	**Length (meter) (z)**
FG-1-A	FLAT	2200	105	0,8	1,2	1,2
FG-2-B	RCK	2000	131	0,8	1,2	1,2
FG-3-C	RCK	1150	35	0,8	1,2	1,2
FG-4-D	RCK	820	45	0,8	1,2	1,2
FG-5-E	RCK	480	27	0,8	1,2	1,2
FG-6-F	FLAT	640	50	0,8	1,2	1,2
FG-7-G	FLAT	500	42	0,8	1,2	1,2
FG-8-H	FLAT	140	64	0,8	1,2	1,2
FG-9-I	FLAT	980	99	0,8	1,2	1,2
FG-10-J	FLAT	655	83	0,8	1,2	1,2

Table 6: Simulation results

Bin	**Volume occupied**	**References Assigned**	**Number of HU assigned**
Flat 1	100%	FG-9-I FG-10-J	175
Flat 2	100%	FG-6-F FG-1-A FG-10-J FG-7-G	175
Flat 3	99.94%	RW-1-A RW-4-D RW-7-G RW-8-H RW-9-I	145
Flat 4	54%	FG-6-F FG-8-H	94
Flat 5	10.48%	RW-5-E RW-9-I	13
Flat 6	0%	-	-
Flat 7	0%	-	-
Flat 8	0%	-	-
Flat 9	0%	-	-
Flat 10	0%	-	-
Rack 1	99%	FG-5-E FG-2-B FG-3-C	173
Rack 2	91%	RW-2-B RW-6-F RW-10-J	141
Rack 3	42%	RW-10-J RW-3-C	61
Rack 4	38%	FG-5-E FG-4-D	67
Rack 5 to Rack 10	0%	-	-

We take as an example the progress of assignment of BIN 1 Flat to show how the algorithm used works:

- Phase 1, 2: we place HU of the reference FG-9-I next to each other, knowing that the storage type required is “Flat”.
- Phase 3: the algorithm moves to the reference FG-10-J which is compatible with the FG-9-I and also requires a flat storage type.
- Phase 4, 5, and 6: the algorithm place similar items next to each other in order to form stock piles with the same references.

After engaging the assignment VBA solver developed initially by Güneş Erdoğan and modified by us for 1200 seconds and 67978 iterations, we obtain the following results:

After analyzing the simulation results, we have concluded the following points:

- The number of bins used: 9 (20 in total).
- Number of HU assigned: 1044 (all the HU were assigned correctly).

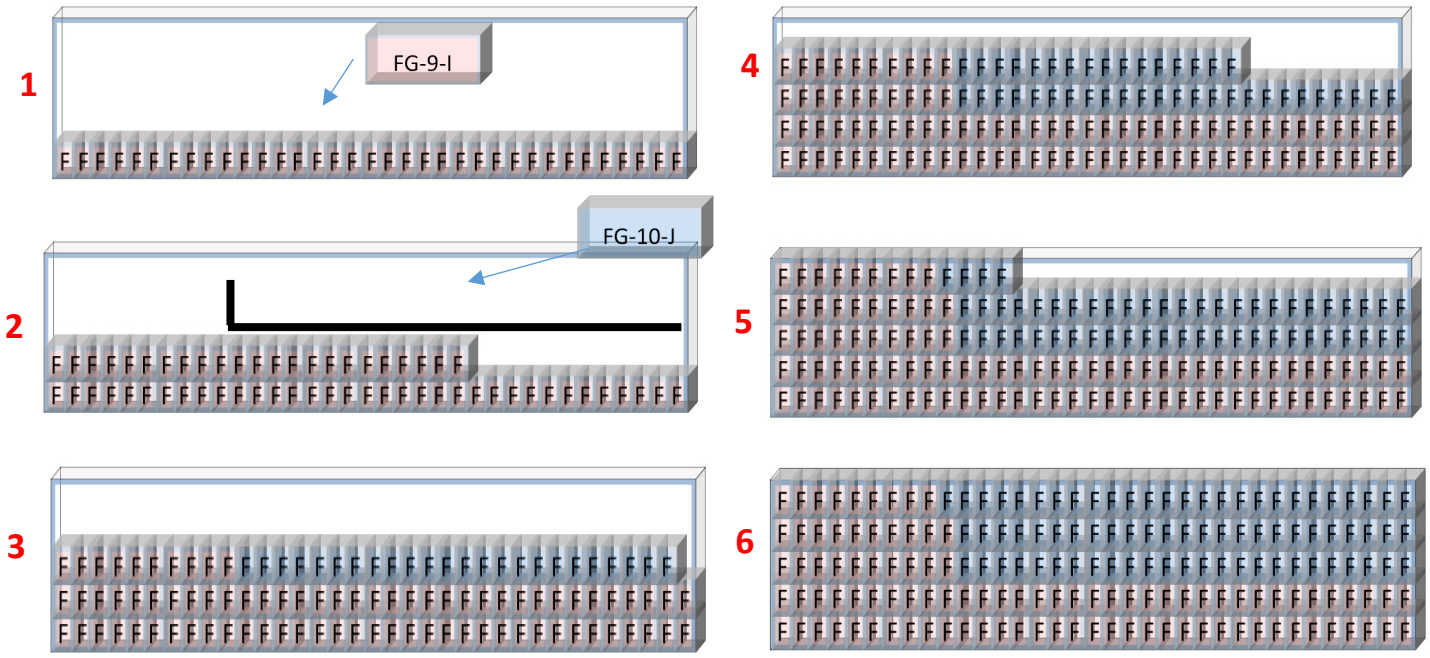


Figure 12: The assignment progress of the Bin 1 Flat

- The constraints of compatibility between raw materials and finished goods have been respected; in each bin we found only one family of products; Raw materials or Finished goods but not both. This means that there will be no congestion while preparing raw materials to be transferred to the linefeed of finished goods to be transferred to preparation area.
- The constraints of compatibility between the references to assign and the storage bin type (Flat or Rack) have been respected.
- The filling rate of bins is maximized, where there is no constraint of compatibility.
- The stocks are assigned following the physical flow, starting from the first bin, then the second, then the third… which keeps the warehouse organized.
- The wall building heuristic has given good quality results, if we look at the first part of assignment progress (figure 13), we observe that the assignment is done by placing HU's of similar articles next to each other to form the first level of stock (phase 1) then the second level (phase 2). In the third level, the heuristic moves to the article FG-10-j which is compatible with FG-9-I and place HU in the third level. Moving to the next levels, the piles formed respond to the notion: brick wall, we found a harmony of articles in each pile, which is a part of the visual management; it is easy to identify the references stored in the bin FLAT 1 as an example when we are present physically in the warehouse.

4.9. Synthesis

The spreadsheet developed defines the alerts of stocks in addition to the nominal stock in quantity, in days of stock and in monetary value for each part number. Furthermore, the required space for each reference equivalent to each alert of stock is defined considering the requested type of storage either massive or rack stock, and the tolerated stacking level. In the MRP calculation, we have integrated the upper and lower limit of stock in order to propose quantities of orders inside this interval of stock when running automatically (daily Job of MRP)

This tool facilitates the management of procurement and production planning by providing an interface including all logistics parameters throughout the supply chain impacting the EQO. The tool we have developed will facilitate the work of logistic decision-makers as it integrates the static and dynamic aspect of the stock while considering logistic management constraints. The assignment part allows to manager to check the feasibility of the stock replenishment policy defined in the segmentation phase compared to the storage locations available and the total budget authorized. Furthermore, it facilitates the management of physical flow; the spreadsheet is a logistic decision support tool. Among these advantages:

- Defining adequate procurement policy and checking the alignment of the stock value calculated with the value monthly budgeted by the top management.
- Defining a dynamic storage area for each item in the warehouse, if there is any change in the values of nominal stock, we anticipate the actions to be taken regarding the storage area.
- Detecting the anomalies and the discrepancies regarding the various logistic parameters of supply or sale (example: supplier minimum order and daily average consumption of products) and see the impact on the number of days in stock in real time mode.

- Ensuring a better interpretation of the stock constraints (FG and RM), by segmenting them into six variables, and then prioritizing the actions plan to get the optimal stock:
  - Stock delivering: Increase the frequency of the delivery downstream (PF) and upstream (RM).
  - Batch: Decrease the launches quantities (FG) and the minimum order (RM).
  - Variability of the demand: To master the variability of customer's needs (FG) / to respect and smooth the production schedule (RM).
  - Service Rate: To respect the production schedule and follow the reliability of the customer's loading (FG)/ Master the suppliers with the supplier service rate (RM).
  - Quality level: Quality action (decrease the defective parts rate…).
  - Safety stock: Safety stock defined following to a management decision or a customer contract decision.
- Check the correspondence of the financial target defined by the top management and the stock management policy defined.

6. Conclusion and perspectives

The research work in our hand is very complex from a modeling point of view; the context of stocks in the industrial domain; and integrating all the constraints of operational management of raw materials and also finished goods. In terms of the initial objectives set, we successfully determine a static and dynamic model of stock calculation. Particularly, define the optimal order quantities in a detailed mathematical and algorithmic demarche. Then propose an algorithm to solve the storage location assignment problem known as NP-Hard problem and include the allocation of space to references in function of the alerts of stock defined in the segmentation phase within a Moroccan multinational company respecting the constraint of flow. Moreover, we propose a multi criteria decision tool to support decision makers in order to take decisions based on current policy of stock replenishment, and show the variables that impact the level of stocks and should be revised or negotiated. Finally, the model proposed could be replicated in any automotive industry context since it encompasses all the constraints adopted in such an exacting context. For other industries, where there is less variables we can use the same model using a relaxation approach of some constraints for each context

For further research, after solving the space allocation problem in both types rack and flat, we have detected a new constraint in the case of flat stocks. The algorithm proposed doesn't include the order of consumption of products when assigning them in the available space; this gap provokes one of the main Muda. It includes the operations of movement of products to withdraw the requested product since all the products are stacked, and then returning the non-needed products to the stockpiles. For all these reasons, we need to solve the assignment problem of flat stock type references to the space in the warehouse considering the picking order (retrieval) of handling units while constituting stock piles and optimize rehandling operations. This problem is also NP-Hard in a strong sense and can be modeled also as a bin packing problem tree dimensions with constraints of retrieval.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

I would like to express my gratitude to my primary supervisor, of thesis Abdelfettah Sedqui, who guided and assisted me throughout this work of research. I would also like to thank the logistics manager of the multinational company for the data, support, and the trust. A special thanks for my wife and family who supported me during this work till the redaction of these lines. I would like to thank also the two anonymous reviewers whose constructive and helpful comments have helped considerably in improving the contents as well as the presentation of the paper.

References

[1] A. Laassiri, A. Sedqui, "Dynamic dimensioning of logistic resources: Case of stocks Moroccan multinational company," in 2020 5th International Conference on Logistics Operations Management (GOL), 1–7, 2020, doi:10.1109/GOL49479.2020.9314755.

[2] M.Z. Babai, Y. Dallery, "Inventory Management: Forecast Based Approach vs. Standard Approach," International Conference on Industrial Engineering and Systems Management (IESM05), 2005.

[3] G. Hadley, T.M. Whitin, Analysis Of Inventory Systems, 1963.

[4] F.W. Harris, "How Many Parts to Make at Once," Operations Research, **38**(6), 947–950, 1990, doi:10.1287/opre.38.6.947.

[5] Y. Yao, M. Dresner, "The inventory value of information sharing, continuous replenishment, and vendor-managed inventory," Transportation Research Part E: Logistics and Transportation Review, 44(3), 361–378, 2008, doi: 10.1016/j.tre.2006.12.001.

[6] D. Battini, A. Gunasekaran, M. Faccio, A. Persona, F. Sgarbossa, "Consignment stock inventory model in an integrated supply chain," International Journal of Production Research, **48**(2), 477–500, 2010, doi: 10.1080/00207540903174981.

[7] B. Korte, J. Vygen, J. Fonlupt, A. Skoda, Le probléme du bin-packing, Springer, Paris : 475–492, 2010, doi : 10.1007/978-2-287-99037-3_18.

[8] A. Laassiri, A. Sedqui, "Dynamic dimensioning of logistics resources: Case of production workshop: Analogy with the problem of bin-packing Moroccan multinational company," in 2019 International Colloquium on Logistics and Supply Chain Management (LOGISTIQUA), 1–8, 2019, doi:10.1109/LOGISTIQUA.2019.8907270.

[9] K. Cuthbertson, D. Gasparro, "The Determinants of Manufacturing Inventories in the UK," The Economic Journal, **103**(421), 1479–1492, 1993, doi: 10.2307/2234478.

[10] K.A. Cook, G.R. Huston, M. Kinney, Managing Earnings by Manipulating Inventory: The Effects of Cost Structure and Valuation Method, SSRN Scholarly Paper ID 997437, Social Science Research Network, Rochester, NY, 2012, doi:10.2139/ssrn.997437.

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 320-326 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Designs of Frequency Reconfigurable Planar Bow-tie Antenna Integrated with PIN, varactor diodes and Parasitic Elements

Mabrouki Mariem*, Gharsallah Ali

Physical Department, Faculty of Mathematical, Physical and Natural Sciences of Tunis, Tunis El Manar University, Campus Universities Tunis - El Manar, Tunis, 2092, Tunisia

ARTICLE INFO

Article history:
Received: 11 February, 2021
Accepted: 11 July, 2021
Online: 16 August, 2021

Keywords:
Bow-tie antenna
Frequency- reconfigurable antenna
Multi-band frequency antenna

ABSTRACT

This paper presents the designs and the simulations of proposed structures of electronically frequency reconfigurable planar bow-tie antenna. In the first part, a modified wide band self-complementary bow-tie antenna is designed and implemented. In the second part, varactor and PIN diodes are integrated in top side to adjust electronically the modified structure of bow-tie antenna over multi-band frequency. By adjusting PIN diode between the two states and by tuning the varactor diode inside these two states; the proposed antenna demonstrates two different operational frequencies. In ON state, the antenna covers a narrow frequency bands and in OFF state the antenna demonstrates a wide-band operational frequency.
Furthermore, a new structure of reconfigurable antenna implemented with PIN diode and two hexagonal parasitic elements is developed to realize a multi-band operational frequency band and to cover GPS and GMS bands. Simulated results show a return loss less than -10dB with a gain varied between 0.5 and 3.5dB.

1. Introduction

In the modern communication system, frequency reconfigurable antenna has been achieved an important extension as for as adjusted operated band and low cost.

Indeed, various structures designs are implemented to improve configurability. In [1], mechanical control method is employed to creat a different operated frequency band. Nevertheless, this methode proposals several problems in comparison with the electronical control method which can achieve an important exactitude and a very fast speed. Electronic control method is more commonly used [2]-[4] based to RF switches such as PIN diodes, MEMS or varactor diodes. Those RF switches can adjust the effective length of the antenna which can tune the frequency band of the antenna when the control bias is varied.

A frequency reconfigurable microstrip slot antenna is designed and implemented in [5]. The proposed antenna structure can switch between six different operation frequency bands using five RF PIN diodes. Those switches are integrated in the slot of the antenna to adjust the effective length of the slot. Measured results demonstrate a return loss less than -10dB at the different frequency bands, a gain equal to 4dB and a bidirectional radiation patterns.

In [6], A compact PIFA antenna is realized to cover UHF DVB-H frequency band. The antenna is implemented using three PIN and a varactor diodes. Three metallic strips are integrated between the radiation element and the ground plane via PIN diodes when the largest metallic strips is directly soldered to the ground plane. Moreover, a lateral short circuit is related to the ground plane via varactor. Simulation results demonstrate four configurations with the appeared of different frequency bands in each configuration.

Varactor and PIN RF switches were employed to adjust the frequency of reconfigurable bow-tie antenna operate over (3-6GHz). Measured results show a reflexion coefficient less than -10dB over the different wide operational frequency bands and a stable radiation patterns with a gain varied between 3.21dB and 5.42 dB [7].

In [8], authors proposed a new configuration which consist to integrate PIN diodes over the bow-tie arms to switch the antenna

*Corresponding Author: Mariem Mabrouki et al, Tunis-Tunisia, Mariem.Mabrouki.fst@gmail.com

between Bluetooth, Wimax and Wlan bands. Measurements demonstrated an important results.

In this paper, we propose different structures of frequency reconfigurable antenna which can switch between wideband and narrow band operational frequency. The proposed structures of reconfigurable antenna have been demonstrated a multi-bands operation in two states with acceptable results.

In the second section, we present the implementation, the simulation and the measurements of the proposed structure of SCBT antenna.

In section 3, a PIN and a varactor diodes are added in the top side of the SCBT antenna to realize a frequency reconfigurable antenna which is switched between a simple frequency band and two dual frequency bands in ON state and a wide-band operational frequency in OFF state.

Then, a proposed structure based on PIN diode and two hexagonal parasitic elements is developed. The two parasitic elements are integrated in the front and in the back side of the antenna structure to produce a new frequency bands over (1.49-1.7GHz) in ON state and over (1.77-1.85GHz) in OFF state.

2. Implementation of SCBT antenna

A wide-band performance is required in the modern wireless communication systems. To obtain this goal, several wideband antennas are implemented in the literature such as Bow-tie antenna which demonstrates an important performances over a wideband frequency [9]-[11].

In this section, a proposed structure of the SCBT antenna is designed and realized using FR4 substrate.

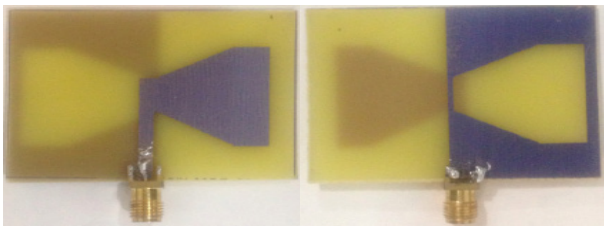


Figure 1: Prototype of proposed bow-tie antenna

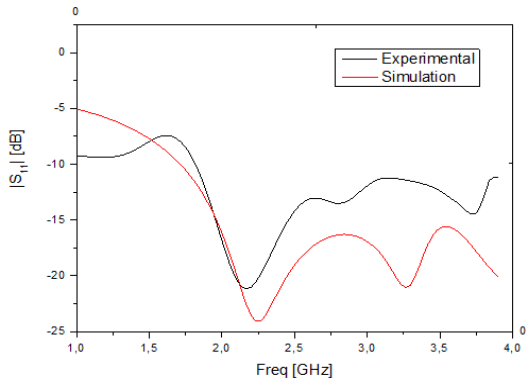
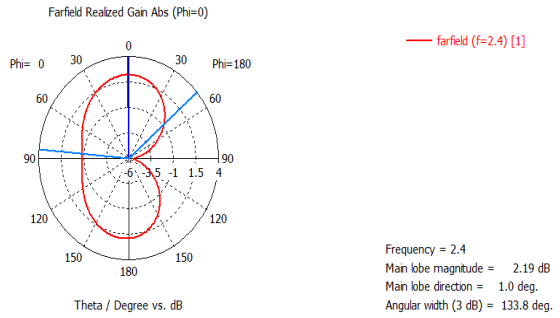


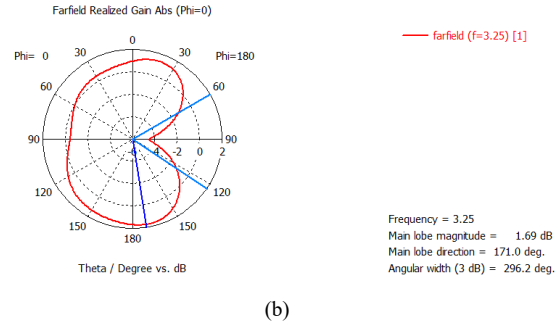
Figure 2: Bow-tie antenna Measurement and simulation

This proposed structure presents good performances such as wideband operation and simple implementation and feeding [12]. The antenna dimensions are calculated using the formulas in [12] and they are optimized via CST: L=54.4mm; W=41.07mm; Wt =2.4mm; Lt =23mm; Wp =25.6mm; Lp =21.5mm; Ls =28.6mm; Wr =0.6mm. Therefore, the size of the modified antenna structure is (50×40 mm2). The top and bottom views of the structure design are shown in figure 1.

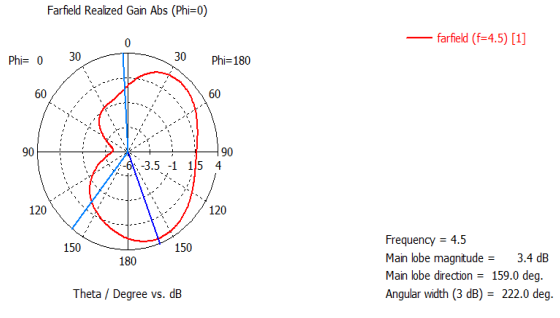
The simulation results are checked through experimental measures figure 2. Simulation return loss is validated with the experimental result and it shows a reflection coefficient less than -10 dB over (2-4GHz). The simulation results of gain patterns are presented in figure.3.



(a)



(b)



(c)

Figure 3: Gain patterns at f= 2.4, 3.25 and 4.5GHz

Three frequencies are chosen to determine the generated gain inside the operational band. Therefore, simulation results of modified SCBT antenna demonstrated a gain equal to 2.19dB, 1.69dB, 3.4dB at 2.4GHz, 3.25GHz and 4.5GHz respectively, with a bidirectional radiation patterns in E plane.

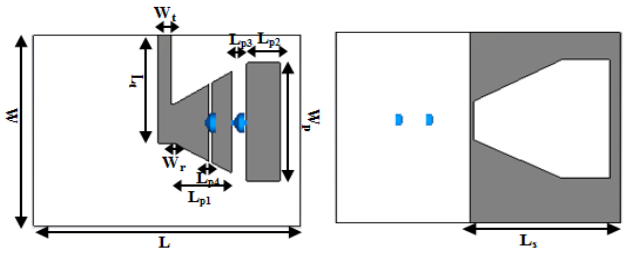
3. Frequency reconfigurable SCBT antenna using PIN and varactor diodes

3.1. Design and Simulation

Varactor and PIN diodes are implemented in the top side of the antenna structure to switch the antenna among different operational frequency bands figure 4. Therefore, the BAR 64-05w

PIN diode is used to control the antenna between the two states. When the ON state is selected, the antenna can achieve a narrow operational frequency band. Otherwise, when the OFF state is selected we can show a wideband operational frequency.

The varactor diode used is MA46585 with a capacitance varied between (0.14 pF to 2.2 pF). By adjusting the reverse bias of the varactor diode between (0V-30V), the varactor capacitances values are improved and the antenna can select the looked-for band frequency. The varactor is modeled in the simulation as a capacitor in series with a forward resistance and inductor. Antenna size is equal to (50x40) mm^2.



(a) (b)

Figure 4: Proposed frequency reconfigurable bow-tie antenna implemented with PIN and varactor diodes: (a) Front view (b) Back view

When the applied Bias control is (V=0.82V), PIN diode is modeled by the resistance R=2.7Ω. Thus, when the diode is tuned ON and by adjusting the varactor capacitances value, different narrow frequency bands are appeared. when the value of the varactor diode capacitance is equal to 0.14pf or 0.19pf, we can show a dual band operational frequency. However, when the capacitance value is fixed to 0.75 pf the antenna is pointed to a simple narrow frequency band.

Otherwise, when the no bias voltage is applied (V=0); the diode is modeled by the capacitor C= 0.22pf. Indeed, three wide bands are obtained when the varactor capacitances value is equal to 0.14pf, 0.19pf and 0.75pf.

The determined simulation results in the two states are presented in figure 5.

When the diode is switched ON and the varactor capacitance is adjusted to 0.75pf, simulation results show a reflection coefficient less than -10dB over a narrow frequency band (1.83-2GHz). Then, a dual band frequency is shown while the varactor capacitance is tuned to 0.19pf. The first band is varied among (2.2-2.5GHz) and the second band occupies the frequencies (3.5-4.2GHz), reflection coefficient in resonance frequencies is equal to -35dB and -18dB at 2.2GHz and 3.6GHz respectively. Two others dual band are achieved when the varactor capacitance is tuned to 0.14pf, the first band is varied between (2.2-2.5GHz) and the second band occupies the frequencies (4.1-4.4GHz). Return loss in resonance frequencies is equal to -60dB and -24dB at 2.1GHz and 4.3GHz respectively. However, when the PIN diode is switched OFF and the varactor capacitance is adjusted to 0.75pf, the reflection coefficient is less than -10dB over (3.5-5GHz). When the varactor capacitance is tuned to 0.19pf the antenna shows a wideband operation with a reflection coefficient less than -10dB. The frequency band is varied among (2.1-5GHz). When the varactor capacitance is adjusted to 0.14pf, the antenna presents wideband frequency operation over (2.1-4.4GHz). Figure 6 proves the gain patterns in ON state with varactor capacitances 0.75 pf, 0.19 pf and 0.14 pf and the gain patterns in OFF state with varactor capacitances 0.75 pf, 0.19 pf and 0.14 pf is illustrated in figure 7.

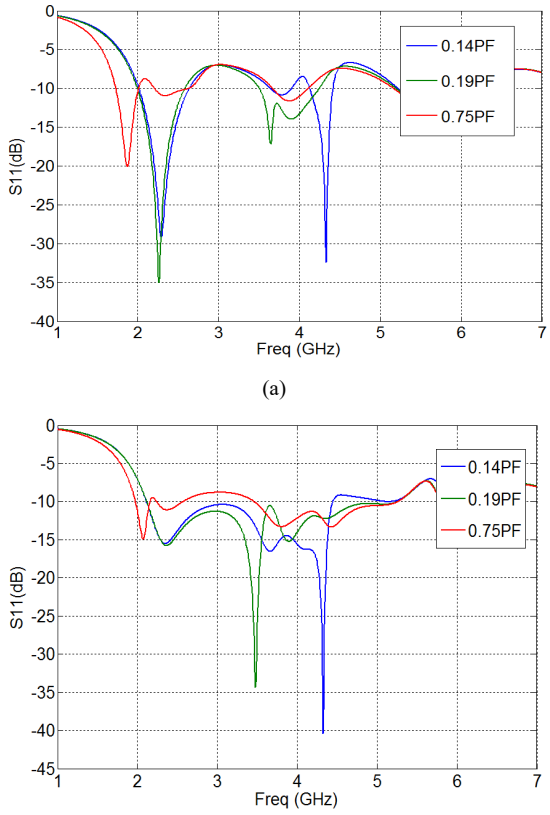
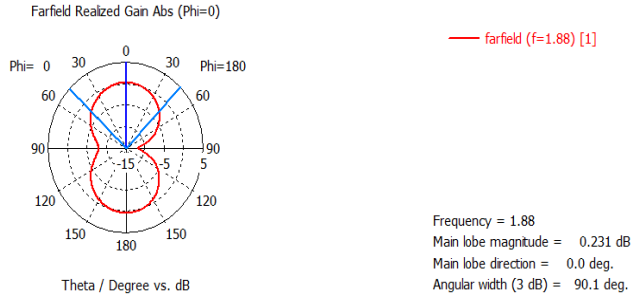
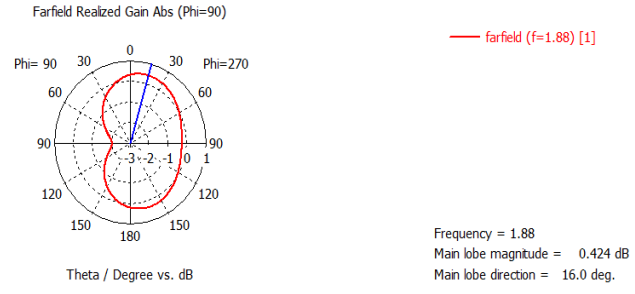


Figure 5: Return loss of reconfigurable antenna with different capacities values: (a) in ON state, (b) in OFF states

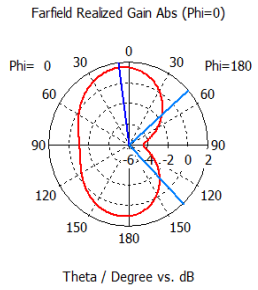


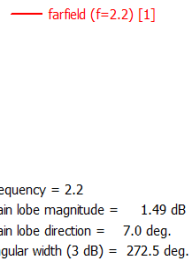
E Plane



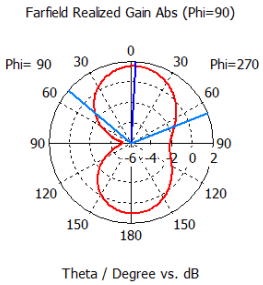
H Plane

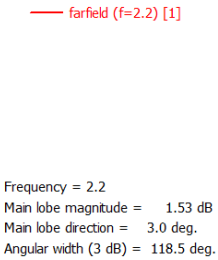
(a)



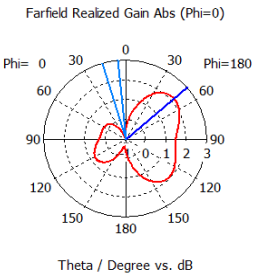


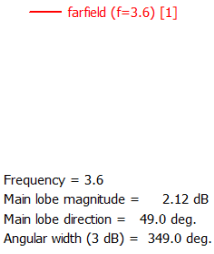
E Plane





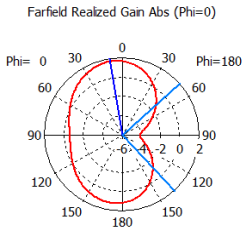
H Plane

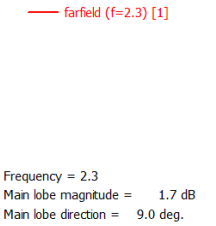




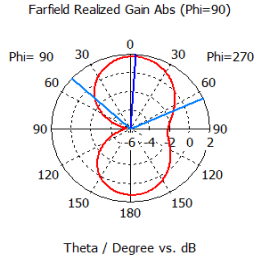
E Plane

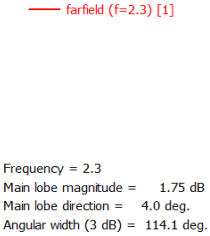
(b)



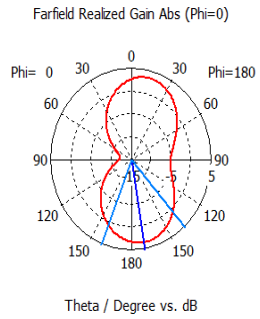


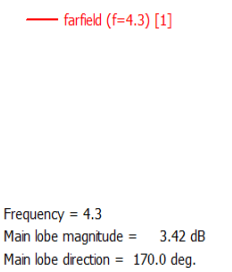
E Plane



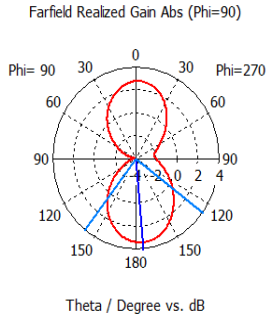


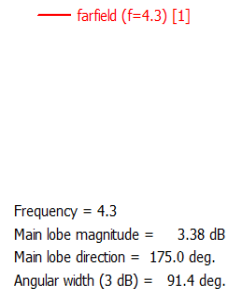
H Plane





E Plane

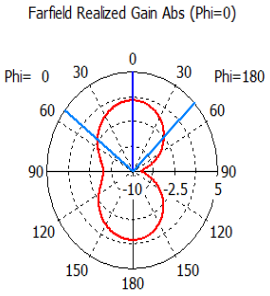


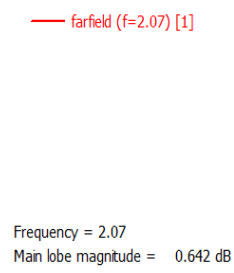


H Plane

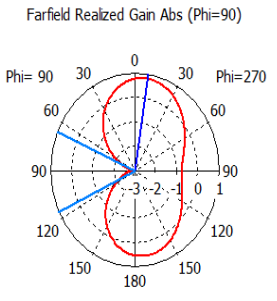
(c)

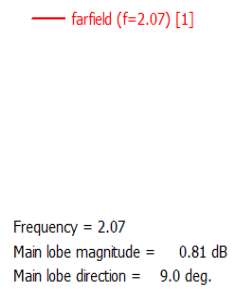
Figure 6: Gain patterns in ON state (a) C=0.75pf, (b) C=0.19pf, (d) C=0.14pf.





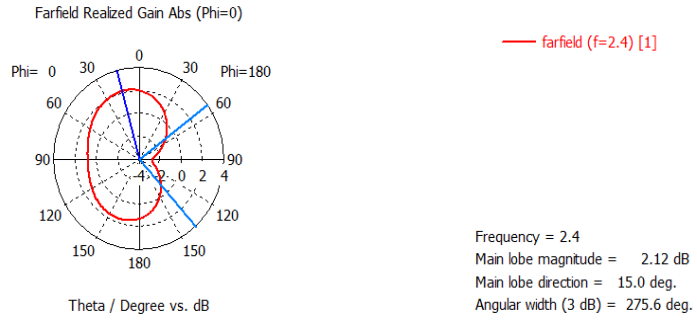
E Plane



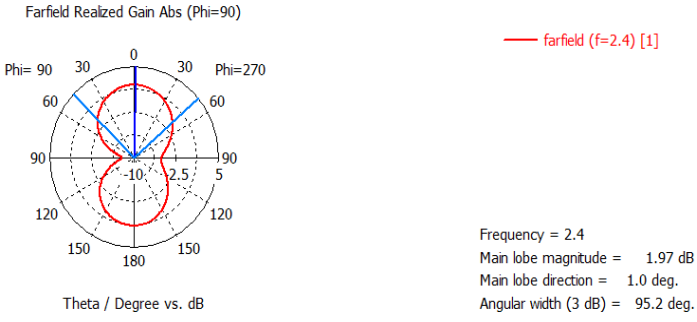


H Plane

(a)

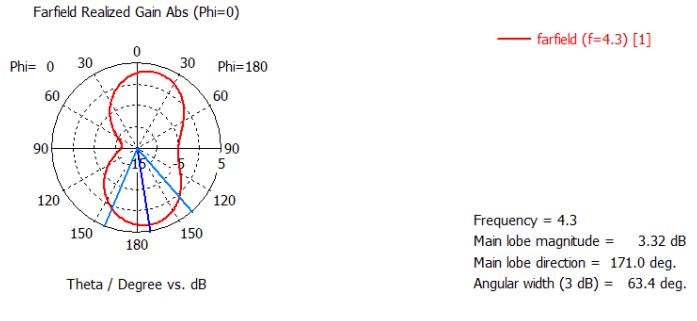


E Plane



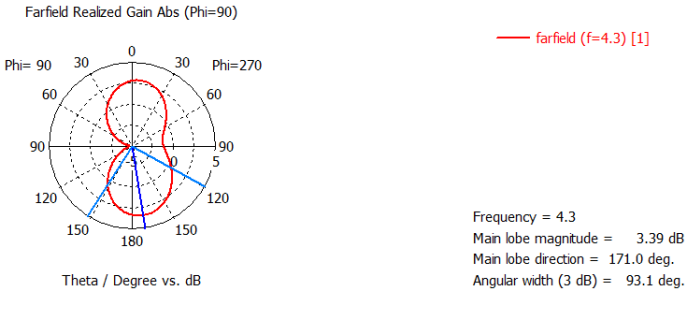
H Plane

(b)



(c)

E Plane



H Plane

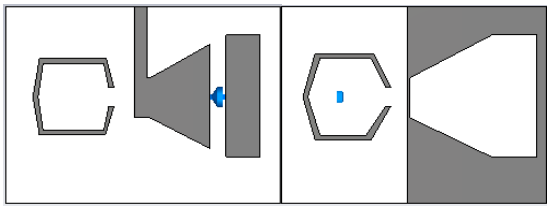
Figure 7: Gain patterns in OFF state (a) C=0.75pf, (b) C=0.19pf and (c) C=0.14pf

Gain patterns results in ON state present a gain equal to 0.231dB at 1.88GHz when the varactor capacitance is on 0.75pf, a gain varied between 1dB and 2.8dB when varactor capacitance is on 0.19pf and a gain around 3.42dB when varactor capacitance is on 0.14pf. In OFF state, the gain is varied between 0.64dB and 3.32 dB over the three wideband frequency. Consequently, bidirectional pattern has been shown and circular polarization has been produced.

4. Frequency reconfigurable SCBT antenna using PIN diode and two parasitic elements

4.1. Design and Simulation

In this section, the idea is to integrate two hexagonal parasitic elements in the front and in the back antenna structure which provide a new operational frequency bands. Using this proposed configuration, GPS and GMS band have been covered. Antenna structure is shown in figure.8.



(a) (b)

Figure 8: Proposed reconfigurable bow-tie antenna integrated with PIN diode and two parasitic elements: (a) Top view (b) Bottom view

We can show Multi-bands frequency operation when the two states are commuted. In figure 9, the reflection coefficients of the proposed antenna are illustrated.

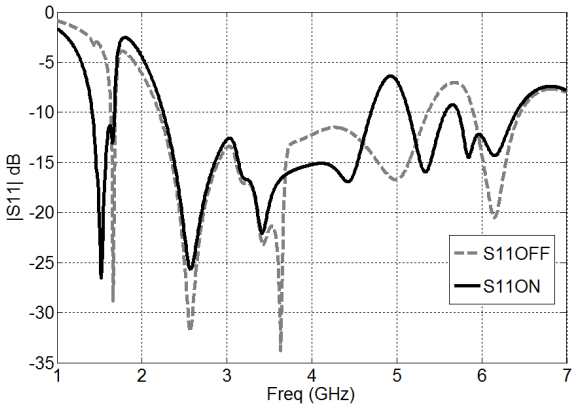
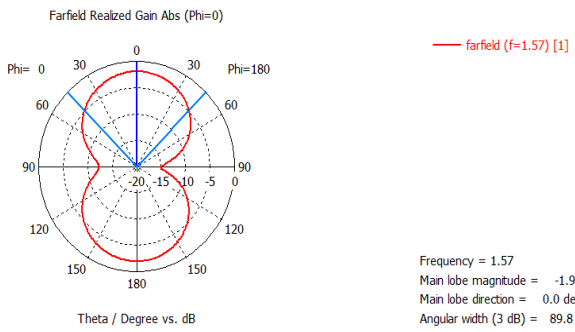


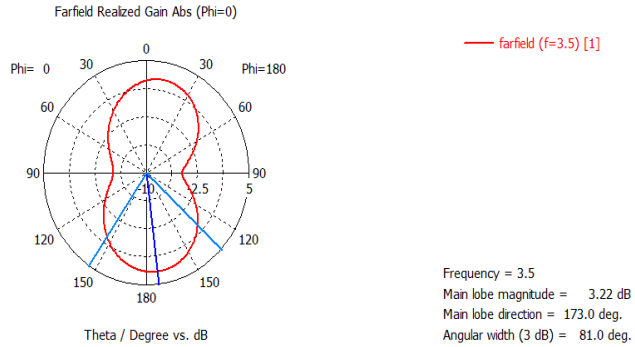
Figure 9: Reflection coefficient results

Simulation results demonstrate two different states with multi-bands operation frequency. The simulated reflection coefficients are less than -10dB at all frequency bands of the two states. In ON state, the improved frequency bands are (1.5-1.75GHz), (2.3-4.61GHz) and (5.15-5.95GHz). At the resonance frequencies 1.58GHz, 3.51GHz and 5.71GHz, the reflection coefficients are -25dB -22,5dB, -34dB respectively. When the OFF state is selected, the obtained frequency bands are (1.77-1.85GHz), (2.24-5.1GHz) and (5.8-6.5GHz) and the reflection coefficients at the resonances frequencies ,1.81 GHz, 2.41 GHz and 5.81 GHz, are -28dB, -32dB and -15dB.

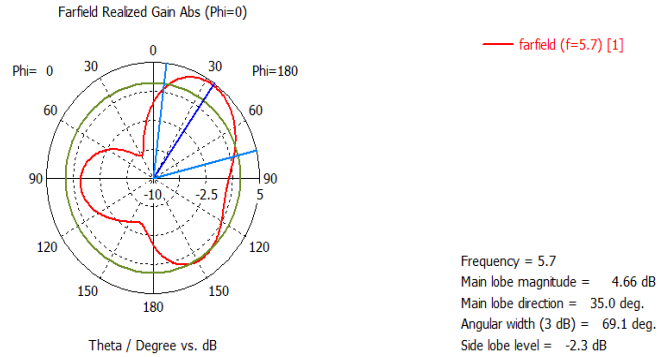
Simulated gain patterns in the two states are presented in figures 10-11.



(a)

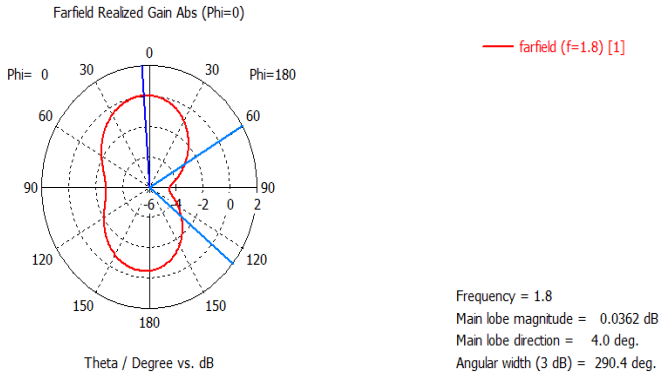


(b)

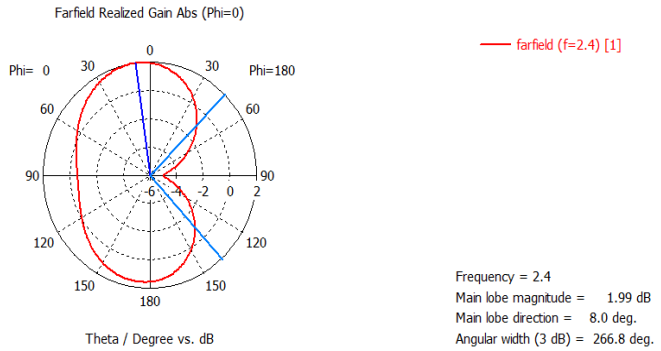


(c)

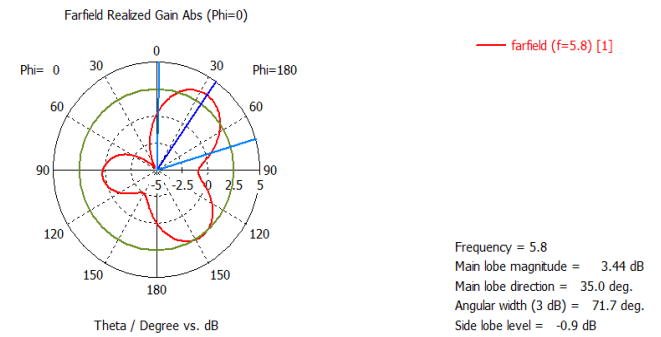
Figure 10: Gain patterns in ON state at 1.58GHz (a), (b) 3.51GHz, (c) 5.71GHz



(a)



(b)



(c)

Figure 11: Gain patterns in OFF state at 1.81GHz (a), (b) 2.41GHz, (c) 5.81GHz

Simulated gain patterns demonstrate the disparity of gain between 2.2dB and 4.7dB and a gain equal to -1.9dB at f=1.58GHz in ON state. when the OFF state is selected, generated gain is restricted to 3.55dB at all operational frequencies and a bidirectional pattern has been shown. Consequently, the proposed antenna provides a circular polarization.

5. Comparison study

Using the parasitic elements, we can show two states with a multi-band operational frequency; the structure is characterized by the simple design and a low control voltage (0v-3v). However, antenna structure using PIN and varactor diodes demonstrates six states which show a simple narrow band, dual-band and wide band operational frequency. The structure requires a two bias voltage with control voltage of (0v- 3v and 15v).

Table 1: Results Table

STATES	Varactor			Parasitic elements
	C1=0.14pf	C2=0.19pf	C3=0.75pf	
ON:	Dual-band [2.2-2.5 GHz] [4.1-4.4 GHz]	Dual-band [2.2-2.5 GHz] [3.5-4.2 GHz]	[1.83-2 GHz]	[1.49-1.7GHz] [2.29-4.6GHz] [5.1-5.9GHz]
OFF:	[2.1-4.4GHz]	[2.1-5GHz]	[3.5-5GHz]	[1.77-1.85GHz] [2.24-5.1GHz] [5.8-6.5GHz]

Reference	Type of switches	Number of switches	Antenna size (mm2)	Number of achieved Bands	Bandwidth of Each Subband
[5]	RF PIN	5	50x46	6 narrow bands	≈ 500MHz
[13]	RF PIN	5	50x100	5 narrow bands	≈ 300MHz
[14]	*PIN &varactor*	2	*30x70*	*6 narrow bands*	*100MHz<BW< 700MHz*
Proposed	*PIN &varactor*	2	*50x40*	*3narrow bands & 3wide bands*	*200MHz<BW< 3GHz*
	PIN& Parasiticelements	1	*50x40*	*2 states with multi-bands in each state*	*200MHz<BW< 3GHz*

6. Conclusion

In this paper, two structures of electronically reconfigurable antenna are designed and simulated. First proposed structure is developed based on PIN and varactor diodes. Thus, three frequency bands are appeared in each state with a good return loss and acceptable gain (1 to 3.5dB). Using this configuration, antenna can commute between a narrow band and a wide band operational frequency. Proposed antenna size is (50×40mm). The second structure of reconfigurable antenna is based on simple PIN diode and two hexagonal parasitical elements implemented in the top and the bottom side of the substrate. The proposed structure can cover GMS and GPS band and it can realize a multi-band operational frequency in each state. Simulation results show an important performance corresponding to the gain and reflection coefficient. Structure size is (50×40mm).

References

[1] Y. Tawk and C.G Christodoulou, "A new reconfigurable antenna design for cognitive radio," IEEE Antenna wireless prog.lett., **8**, 1378-1381, 2009, DOI: 10.1109/LAWP.2009.2039461.

[2] A. C. K. Mak, C. R. Rowell, R. D. Murch and C. L. Mak, "Reconfigurable Multiband Antenna Designs for Wireless Communication Devices, " IEEE Transactions on Antennas and Propagation, **55**(7), 1919-1928, 2007, doi: 10.1109/TAP.2007.895634.

[3] D.Peroulis, K. Sarabandi, and L. P. B. Katehi, "Design of reconfigurable slot antennas," IEEE Trans. Antennas Propag., **53**(2), 645–654, 2005, doi: 10.1109/TAP.2004.841339.

[4] W.H. Weedon, W.J. Payne and G. M. Rebeiz, "MEMS Switched reconfigurable antenna," In Procceeding of the IEEE International Symposuim on Antenna And Propagation, **3**, 654-657, 2001, doi: 10.1109/APS.2001.960181.

[5] H.A. Majid, M. Kamal A. Rahim, M. Rijal Hamid and M. F. Ismail, "A Compact Frequency-Reconfigurable Narrowband Microstrip Slot Antenna," IEEE Antennas And Wireless Propagation Letters, **11**, 616 - 619, 2012, doi: 10.1109/LAWP.2012.2202869 .

[6] F. Canneva. F. Ferrero, J.M. Ribero, R. Staraj, "Reconfigurable miniature antenna for DVB-H standard," IEEE Antennas and Propagation Society International Symposium," 2010, doi: 10.1109/APS.2010.5561955.

[7] T. Li, Huiquing Zhai, Long Li and changhong Liang, "Frequency – Reconfigurable Bow-tie antenna with a wide tuning Range," IEEE Antennas and Wireless Propagation Letters, **13**, 1549 - 1552, 2014, doi: 10.1109/LAWP.2014.2344676.

[8] T. Li, Huiquing Zhai, Long Li and changhong Liang, "Frequency-Reconfigurable Bow-Tie antenna for Bluetooth, Wimax and Wlan application," IEEE Antennas and Wireless Propagation Letters, **14**, 171 - 174 , 2015, DOI: 10.1109/LAWP.2014.2359199.

[9] A.A. Eldek, Atef Z. Elsherbeni, and Charles E. Smith, "Wideband Microstrip-fed Printed Bow-tie Antenna For Phased Array Systems," Microwave And Optical Technology Letters, **43**, 123–126, 2004, doi: 10.1002/mop.20396.

[10] T. Karacolak and Erdem Topsakal , "A Double-Sided Rounded Bow-Tie Antenna (DSRBA) for UWB Communication," IEEE Antennas And Wireless Propagation Letters, **5**, 446 - 449, 2006, doi: 10.1109/LAWP.2006.885013.

[11] K. P. Ray, "Design Aspects of Printed Monopole Antennas for Ultra-Wide Band Applications," International Journal of Antenna and Propagation, 1-8, 2008, doi: 10.1155/2008/713858.

[12] K.H. Sayidmarie, Yasser A. Fadhel, "A Planar Self-Complementary Bow-Tie Antenna for UWB Applications ," Progress In Electromagnetics ResearchC, **35**, 253-267, 2013, doi:10.2528/PIERC12103109

[13] Y. Choi , Ji-Hun Hong , Jong-Myung Woo, "Electrically and Frequency-Tunable Printed Inverted-F Antenna with a Perturbed Parasitic Element," Journal Of Electromagnetic Engineering And Science, **20**(3), 164-168, 2020, doi:10.26866/jees.2020.20.3.164.

[14] J. Lim, Gyu-Tae Back, Young-Il Ko, Chang-Wook Song, and Tae-Yeoul Yun, "A Reconfigurable PIFA Using a Switchable PIN-Diodeand a Fine-Tuning Varactor for USPCS/WCDMAM-WiMAX/WLAN," IEEE Transactions On Antennas And Propagation, **58**(7), 2404 - 2411, 2010, doi: 10.1109/TAP.2010.2048849.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 327-333 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Real Time RSSI Compensation for Precise Distance Calculation using Sensor Fusion for Smart Wearables

Kumar Rahul Tiwari*, Indar Singhal, Alok Mittal

STMicroelectronics, SRA-SAIL, Greater Noida, 201308, India

ARTICLE INFO

Article history:
Received: 14 July, 2021
Accepted: 02 August, 2021
Online: 16 August, 2021

Keywords:
Bluetooth® Low Energy
RSSI
Antenna Orientation
STEVAL-BCN002V1
Inertial Measurement Unit
Yaw Angle
Sensor Fusion

ABSTRACT

To effectively implement the social distancing or digital contact tracing in epidemic using an RSSI-based localization approach through Bluetooth beacon is one of the most widely used technologies, but simply using RSSI measurement is not more relevant because the RF signal is affected by several factors and the environment of usage. Traditional distance or positioning algorithms have large-ranging errors when applied for moving objects because they do not account for the device orientation and use fixed path loss models. Hence, the distance between the nodes cannot be obtained accurately by RSSI measurement in a dynamic environment. In this paper, we propose a solution to compensate for the RSSI loss in real-time by filtering out the noise and then accounting for the antenna orientation using a Beacon Packet. Antenna Orientation is determined using 9DoF (9 Degrees of freedom) IMU (Inertial Measurement Unit). The nodes simultaneously advertise their presence and scan for the presence of other similar beacons in their range. These nodes also deploy Low Power techniques during periods of inactivity to conserve battery power. Advertising is performed on three Bluetooth channels and no connection or response packet is required between the devices during advertising and scanning activities (ADV_NONCONN_IND). The addition of the Motion Sensor could also be used to optimize the battery life of the device.

1. Introduction

This paper is an extension of work originally presented in ICCCS Patna 2020 [1]. Smart Contact tracing with social distancing is one of the efficient ways to avoid the spread of contagious pandemics, such as COVID-19 and future pandemics. In such cases, it is an advantage to alert people to maintain a safe distance. Additionally, it is advantageous to record their physical contact wherever they are at work, in public places, or at a relative's home, etc. RSSI (Received Signal Strength Indicator) based distance calculation between two BLE (Bluetooth® Low Energy) based movable devices (nodes) are widely used technologies. However, RSSI measurements give lower accuracy due to variable attenuation (path loss) and fading effects with high variance [2]. Another factor that affects the RSSI is the antenna orientation and thus can affect the calculated distance between two transceivers. In an ideal scenario, the RSSI of the received signal does not very much. However, in a practical scenario the RSSI is affected by different factors: e.g., physical distance, reflections of objects, environmental parameters, movement of objects or change in the environment, antenna position and polarization etc.

In this paper, we describe the algorithm and later experimentally verified how effectively distance calculation can be done by real-time compensation of the RSSI loss using nested mathematical filters and by accounting for the relative antenna orientation between transmitter and receiver for digital contact tracing or social distancing related applications.

The solution developed is based on ST BlueNRG-2 SoC (Bluetooth Low Energy (BLE) system-on-chip), LSM6DSO and LIS2MDL (9DoF IMU Motion Sensor) for estimating device orientation. The assumption is being taken that orientation of the device is the orientation of antenna embedded in device from a fixed plane. Bluetooth® Low Energy technology is used due to several advantages like wide deployments in wearables and Low-Power consumption. etc. It can communicate with a smartphone which can configure device parameters from the app. This solution will also help to avoid huge dependency on smartphone(s) even for pausing/ resuming the logging of the beacon.

Antenna radiation pattern changes in a different orientation and this impact the RSSI of the signal, impacting parameters like distance based on RSSI. In most of the known solutions or papers,

*Corresponding Author: Kumar Rahul Tiwari, rahul.tiwari@st.com

orientation of antenna in RF devices are static, and relative antenna direction is not accounted.

2. Design Architecture

The concept has been proven using a system solution using the STEVAL-BCN002V1B development kit based on BlueNRG-2 SoC [3] [4]. This development kit hosts multiple sensor which includes gyroscope, magnetometer, accelerometer, Time-of-Flight, humidity, pressure, and microphone sensors. It can be powered by a common CR2032 coin cell. The development kit communicates with a Bluetooth® Low Energy enabled smartphone running the ST BLE Sensor app [5], available on iTunes and Google Play and stores. Following are major components in STEVAL-BCN002V1 as shown in Figure 1:

- BlueNRG-2: Programmable Bluetooth® LE 5.2 Wireless SoC
- LSM6DSO: 3D gyroscope and 3D accelerometer [6]
- LIS2MDL: high-performance, ultra-low power, 3-axis digital magnetic sensor [7]

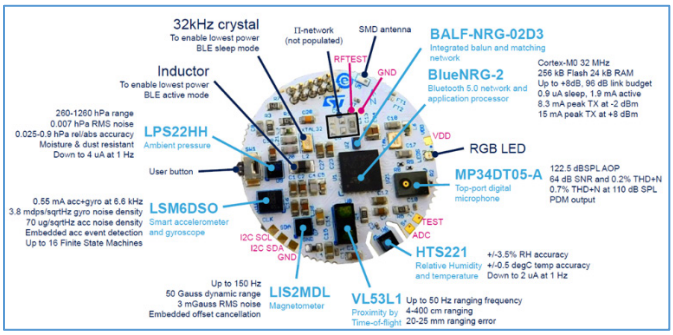


Figure 1: STEVAL-BCN002V1 known as BlueNRG-Tile

The MEMS sensor section of the STEVAL-BCN002V1 sensor node includes inertial and environmental MEMS sensors connected with the BlueNRG-2 via an I2C bus operating at 400 kHz. All sensors can generate interrupts, but only the interrupts from the LIS2MDL magnetometer and the LSM6DSO accelerometer and gyroscope are connected with the BlueNRG-2 through dedicated and independent lines.

BlueNRG-2 integrates a Bluetooth Low Energy radio (BLE), an ARM® Cortex®-M0 core, 24 kB of static RAM memory, 256 kB of Flash memory, SPI (max 1 MHz in slave and 8 MHz in master mode), two I2Cs (standard 100 kHz or fast 400 kHz), UART interfaces; two multi-function timers (MFT), a DMA controller, RTC and watchdog, and an ADC with PDM stream processor.

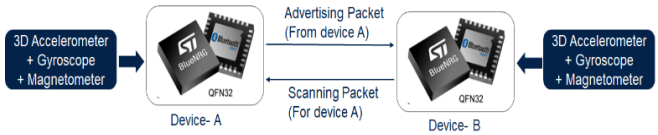


Figure 2: Sequential Advertising and Scanning by Transceiver

Once the BLE stack is initialized in STEVAL-BCN002V1 node, it will advertise non-connectable undirected beacon (ADV_NONCONN_IND) packets and sequentially scans packets from neighbor wearables as shown in Figure 2. RSSI approximation for the relative distance between two or more BLE devices/Nodes can be improved using sensor fusion of Accelerometer, Gyroscope and Magnetometer. The solution is useful in several applications especially in social distancing or digital contact tracing applications.

The impact of the RSSI variation by rotating antenna on same plane can be compensated by sending the Euler angle (Roll, Pitch, Yaw) of antenna obtained from sensor fusion through advertising packet as shown in Figure 3. In the proposed solution, only Yaw axis data (direction of the antenna) is broadcasted.

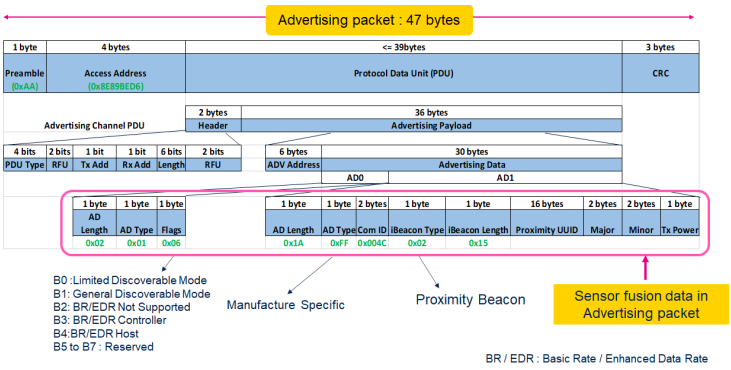


Figure 3: Advertising packet format with sensor fusion data

It is important to know that the radiation pattern is not only a function of the antenna itself, but it depends also on the overall system including PCB layout, ground plane size, space and mechanical surroundings. The Radiation pattern of the antenna as shown in Figure 4 is taken as a reference to observe the change in RSSI with different Euler angle, which is taken on X and Y plane keeping Z axis constant with reference to IMU.

The solution will overcome the drawback in which RSSI variations are huge even if the node is at same position but in a different direction or orientation. The radio frequency section of the STEVAL-BCN002V1 includes the following elements:

- BALF-NRG-02D3 ultra-miniature balun which integrates matching network and harmonics filter.
- A Pi-network which allows additional filtering and provides access points for testing. Note: This network is not populated, as the integrated balun provides the necessary matching.
- An SMD 2.4 GHz (ANT016008LCS2442MA1) antenna, which requires a certain clearance area on the PCB and specific passives for precise tuning.

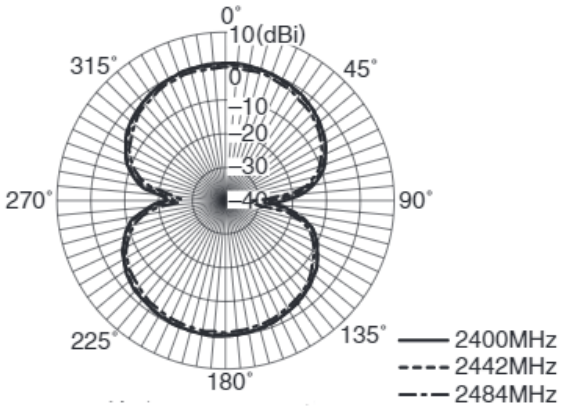


Figure 4: Antenna Radiation Pattern

3. Proximity Sensing and RSSI Compensation Algorithm using Sensor Fusion

Following paragraph explains in detail about the algorithm that is developed for calculating the proximity of two Bluetooth/RF Nodes using RSSI Measurement and compensating RSSI loss in real-time by embedding the dynamic node orientation in advertising packet. Node orientation will be calculated using 9DoF IMU Motion sensor

- Initially sensor fusion with 9 axis (3 axis each for accelerometer, gyroscope and magnetometer) is initialized, and sensor orientation is set, default orientation is ENU (East: X, North : Y, UP : Z)
- The output data rate for gyroscope and accelerometer should be equal to or greater than 100Hz, the magnetometer can be 40Hz
- Magnetometer and Gyroscope calibration is done as shown in Figure-5, Accelerometer calibration is not necessary for sensor fusion except for applications demanding very high orientation precision. Calibration image is captured from ST BLE Sensor App

Figure 5: Node rotation pattern for Calibration

- Roll, Pitch and Yaw as shown in Figure 6 is obtained as an output of sensor fusion. Next, only the Compass angle on the Yaw axis (antenna direction from a fixed point) is calculated using sensor fusion and is embedded in the last 2 bytes of advertising packet to advertise it regularly. Measurement of compass angle depends on the Output data rate of sensors too

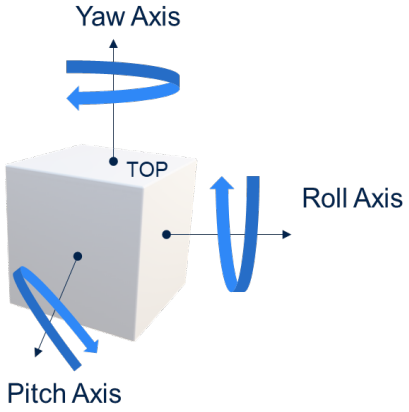


Figure 6: Roll, Pitch and Yaw axis

- Each BLE transceiver advertises and scans sequentially. Tuning the scan window, scan interval and advertising interval parameters can significantly impact battery life. The interval must be an integer multiple of 0.625ms (Time = N * 0.625ms) from 20ms to 10.24s. For ADV_NONCONN_IND (non-connectable undirected event), the minimum advertising interval can be 20ms. There is also a random delay generated by the Link Layer between 0 to 10ms in each advertising packet

The Controller maintains the list of these advertising data and provides the relevant information in one Advertising Report event from multiple devices

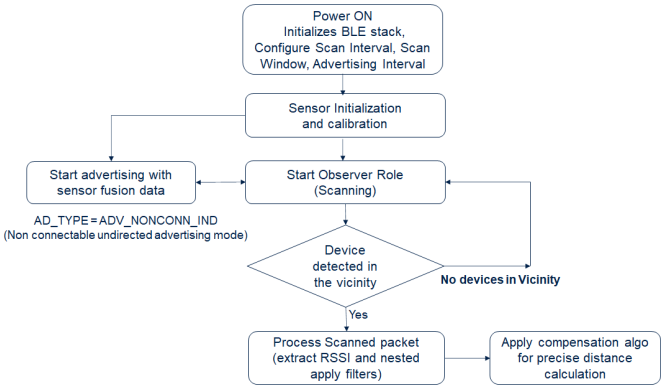


Figure 7: Flow chart of event generation during passive scanning

- RSSI strength is taken for the beacons received during scanning and nested filtering algorithm is implemented to filter out the variations. One of the filters applied in this application is Weighted Mean filter as in equation (1), where 0< αlpha <1. To determine the best, optimize constant (αlpha), several tests were performed with different constant values between 0 to 1. The results and known conditions were analyzed, and best optimized value was achieved between 0.6 to 0.7

$$\text{FILTER\_RSSI} = (int8\_t)(((1 - alpha) \times aCurrent\_RSSI) + (alpha \times aPrevious\_RSSI)); \quad (1)$$

- RSSI samples are filtered using nested filtering mechanism then transmit power and filter RSSI values are used to calculate distance [8] by using equation (2):

$$\text{Distance} = 10^{(Measured\ Power - RSSI)/(10 \times N)} \quad (2)$$

Measured Power: Factory-calibrated constant, it indicates average RSSI value at 1 meter

N: Environment dependent constant (or path loss exponent) which is in the range of 2 to 6

The path loss exponent indicates the rate at which the path loss increases with distance. Normally, N is defined between 2 to 6 as shown in Table 1 [9]

Table 1: Path loss exponent for different environments

Environment	Path loss exponent (n)
Free space	2
Urban area cellular radio	2.7 to 3.5

Shadowed urban cellular radio	3 to 5
Inside a building - line-of-sight	1.6 to 1.8
Obstructed in building	4 to 6
Obstructed in factory	2 to 3

- During the scanning, the transceiver extracts the yaw or compass angle of the neighbor transceiver and calculates the relative yaw angle
- Slope of RSSI vs Yaw angle is calculated as shown in Figure 12, assumed to be linear. Equation (3) for real-time compensation of RSSI is derived based on the slope and Relative Yaw angle

$$CompensatedRSSI = (int8\_t)(Filtered_{RSSI} + (Relative\ yaw \times Slope)) \quad (3)$$

Based on Compensated RSSI value, Compensated distance which incorporates the error due to mismatch antenna orientation and is calculated as in equation (4)

$$\text{Compensated distance} = 10^{(Measured\ Power - RSSI)/(10 \times N)} \quad (4)$$

- The Look-up table is maintained for 2m, 4m, 6m for yaw-angles ranging between 0 to 360 degrees as shown in Figure 8. The look-up table will be taken as a reference to calculate the Variation of RSSI for relative yaw angle from 0 to 90.

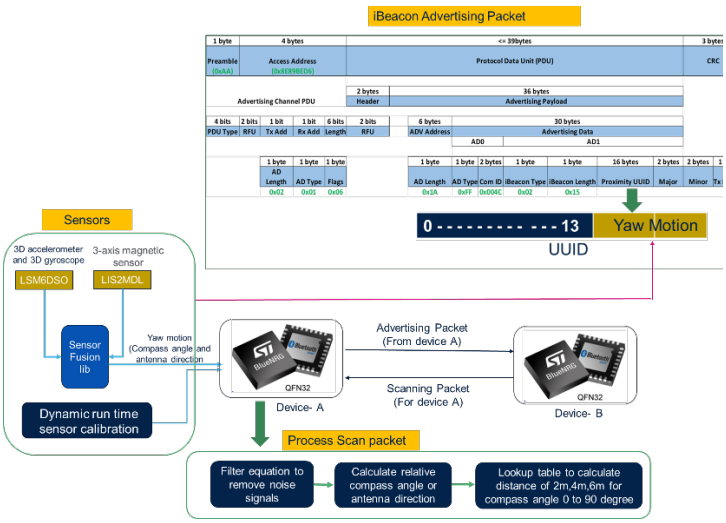


Figure 8: Flow chart of Proposed Solution

4. Adaptive Filtering of RSSI Values

Fluctuation in RSSI is influenced by environmental noise and values are varied with time due to multipath reflections [10]. Variations in RSSI on same distance can be optimized by applying nested filtering algorithms.

4.1. Weighted Mean Filter

Optimized constant value αlpha derived from experiments is multiplied with previous measured RSSI value and 1 – αlpha [11] to the current RSSI value received from same receiver or node. αlpha is for deciding the weightage between previous RSSI and current measurement, αlpha can be in between zero to one. Several tests were performed with different constant values and best results were obtained for αlpha between 0.6 to 0.7

$$Filtered\ RSSI = (1 - \alpha\text{lpha}) \times A + \alpha\text{lpha} \times B \quad (5)$$

A: Current RSSI value

B: Previous RSSI value from same receiver

4.2. Moving Average and Feedback Filter

After applying Weighted Mean filter on RSSI, output values are updated with current values. Feedback is provided to the input and parallelly moving average methodology are applied on the number of samples saved in RAM (Random Access Memory of the SoC), which further filtered the noise. Nested filtering algorithm reduces the variations to a large extent.

5. Power Optimization using Sensor Fusion

Another advantage of using Motion Sensor in Digital Contact Tracing application is to optimize the power for longer battery life. The Wearable will go automatically into low power mode if there is no movement or activity sensed by the motion sensor for certain time. Whenever there is some activity sensed, for example, wearable is worn by a person again; Device will immediately switch from sleep mode to normal mode using the Wake-up event generated by the LSM6DSO sensor. As the wakeup interrupt line is connected with motion sensor interrupt line and the sensor generates an interrupt as soon as the device moves from it's position. Action will be to stop the radio (scanning + advertisement) automatically when the devices are not being worn (for example at home) as shown in Figure 9. Similarly, the radio can be switched On if the devices are being worn.

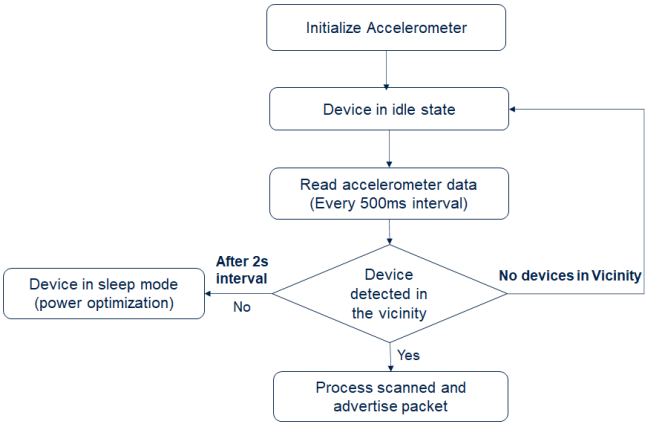


Figure 9: Device Power state in wake up or sleep mode

6. Test Analysis and Results

Various tests analysis performed on multi nodes in the network to observe RSSI variations due to antenna orientation [12]. In this paper, Sensor fusion impact is accounted to compensate RSSI and Distance between two nodes. Node-A is fixed at a center of the circle while Node-B is at 2m distance and rotated in approximately 2m radius as shown in Figure 10. STEVAL-BCN002V1 is taken as nodes shown in Figure 11. At each quarter movement, hundreds of filtered RSSI samples were measured to calculate the mean distance and error variance. With the change in

direction of Node B, it is assumed that antenna direction also changes as the or an antenna is embedded in the Node. The Direction of Antenna is measured in Yaw angle also can be counted as Compass angle. Compass angle measurement is logged as shown in Figure 12.

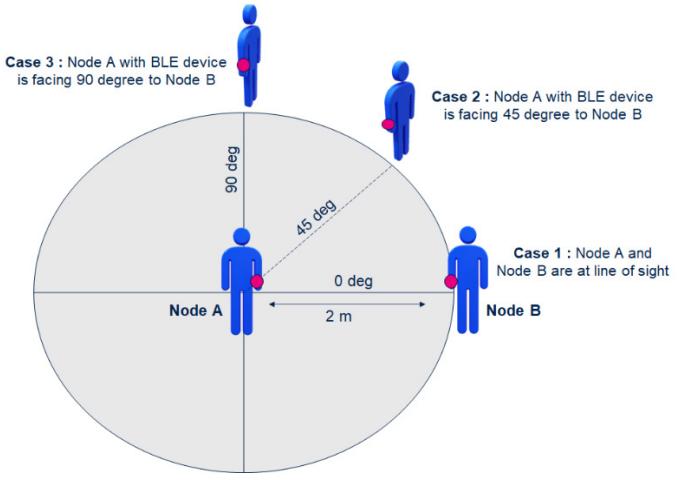


Figure 10: Node B direction with respect to fixed Node A

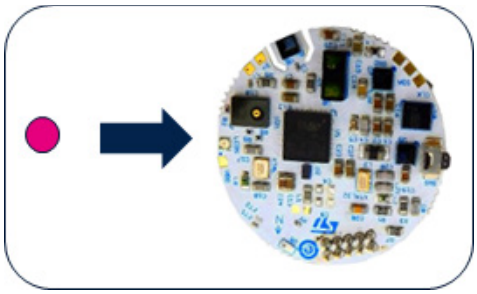


Figure 11: STEVAL-BCN002V1 as Node

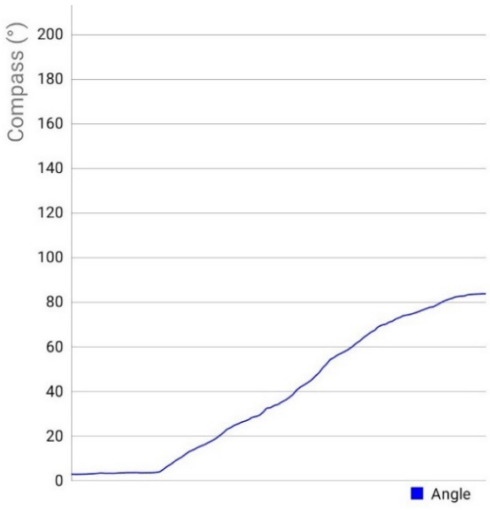


Figure 12: Compass Angle on Yaw axis is measured from 0 to 90 degree

Relative compass angle or Relative Yaw angle is the difference between the direction of two antennas in terms of angle. Even the nodes are at a fixed distance, RSSI strength is weakening as the orientation between two antennas are not aligned as shown in Figure 13.

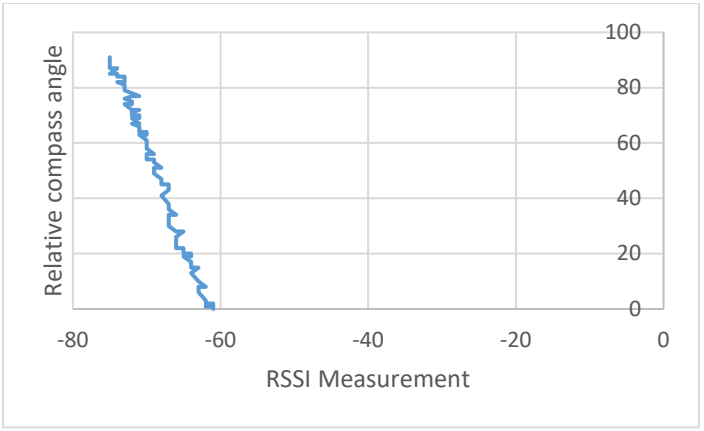


Figure 13: RSSI Measurement vs Relative compass angle on fixed position

Two experiment cases were defined to observe the impact on RSSI measurement and distance calculation, one before accounting Sensor Fusion in advertising packet and one after accounting it in packet. Total 100 samples were taken in each case.

6.1. RSSI Measurement on Node B before applying Compensation Algorithm

As shown in Figure 14 that before applying the compensation algorithm there is a huge impact in RSSI Measurement observed on Node B by just changing the relative antenna orientation or antenna direction between Node B and Node A, which further leads to calculating the wrong distance as per the path loss model, even if the actual distance between Node B and Node A is 1m. Similarly, Table 2 shows that as the Node B antenna direction is moving away from the line of sight of Node A, the calculated distance is increasing accordingly

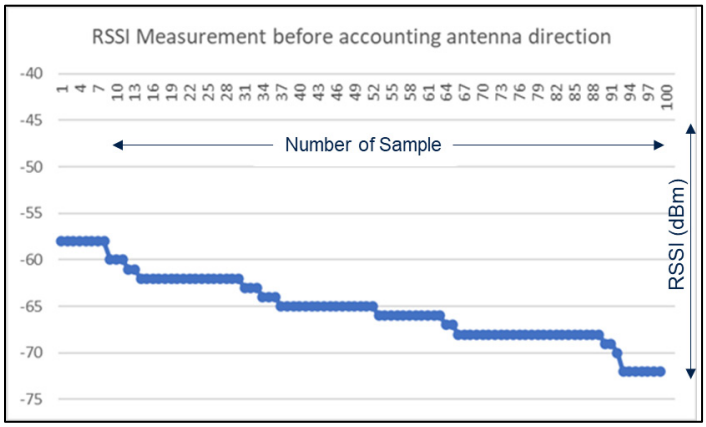


Figure 14: RSSI samples measurement without accounting relative antenna direction

6.2. RSSI Measurement on Node B after applying Compensation Algorithm

In this experiment, the Antenna direction in terms of compass angle is included in the advertising packet of Bluetooth® Low Energy node. Node-B extracts the antenna direction of Node-A to calculate the relative direction as shown in Figure 14, that after applying the compensation algorithm the impact in RSSI measurement observed on Node B by changing the relative

antenna orientation or antenna direction between Node B and Node A is compensated, which further leads to calculating precise distance calculation as per the path loss model. Similarly, Table 3 shows that calculated distance based on RSSI after accounting the effect of Antenna direction is close to 1m only

Table 2: Distance based on RSSI without compensation

Relative Yaw Angle (B - A) in degree	RSSI (without compensation) Observed	Distance based on RSSI (in meter)
0	-58	1.10201845
20	-60	1.33253754
45	-62	1.41827941
90	-64	1.58188643

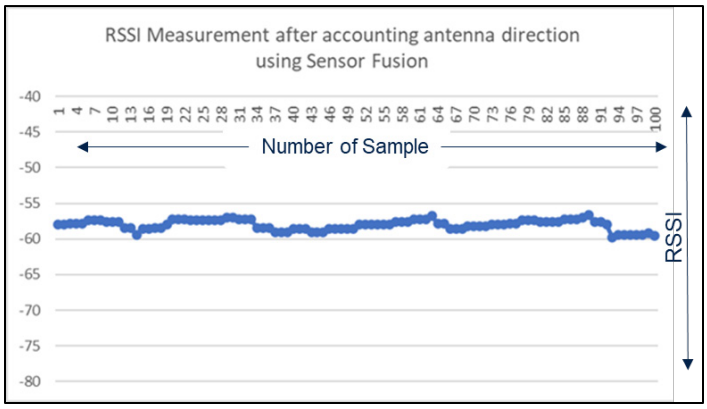


Figure 15: RSSI samples after accounting relative antenna direction

Table 3: Distance based on RSSI after compensation

Relative Yaw Angle (B - A) in degree	Compensated RSSI (dBm)	Compensated Distance (mm)
0	-58	1.10018454
20	-58	1.12301845
45	-57.2	1.07151930
90	-56	1.21618600

7. Current Consumption in Proposed Solution

The current consumption of the BlueNRG-2 can be accurately predicted under different conditions using the BlueNRG consumption tool [13]. As mentioned in paper that BlueNRG-2 SoC (Bluetooth® Low Energy application processor), LSM6DSO and LIS2MDL (9DoF IMU Motion Sensor) are the main component used in STEVAL-BCN002V1 for the proving the concept. Hence active phase and inactive phase current consumption of BlueNRG-2, LSM6DSO (accelerometer and gyroscope), LIS2MDL (magnetometer) is shown in Table 4.

Table 4: Current Consumption

Device	*Active Phase*	*Inactive phase (power not gated by MCU)*
BlueNRG-2	1.9 mA (Active mode)	0.9 µA (Sleep mode)
LSM6DSO	280 µA (50Hz)	3 µA (power-down)
LIS2MDL	475 µA (50Hz)	1.5 µA (power-down)

8. Conclusion

The use of Sensor Fusion for accounting Antenna Orientation in Real-time Location system is quite helpful to obtain the optimal design for digital contact tracing or social distancing related applications. Additionally, sensor fusion helps to optimize battery life by sensing the motion. If the node is stationary, it will direct the controller to disable all the radio activities and in the same way it will generate an event if any motion is sensed to wake up the device from sleep mode. The Antenna direction or orientation of one device is advertised using a Bluetooth beacon at regular intervals to other devices in the vicinity, so that the device while acting as a receiver can estimate the distance using RSSI by also accounting whether the device antenna is perfectly aligned or not. The Solution will be able to estimate the approximate RSSI for a specific position in a dynamic environment. The Solution will also help to avoid snooping of the beacon as the data which is being advertised is changing at regular intervals due to a change in antenna orientation.

Acknowledgment

We would like to thank STMicroelectronics for providing all the resources to accomplish this work.

References

[1] K.R. Tiwari, I. Singhal, A. Mittal, "Smart Social Distancing Solution Using Bluetooth® Low Energy," in 2020 5th International Conference on Computing, Communication and Security (ICCCS), 1–5, 2020, doi:10.1109/ICCCS49678.2020.9277175.
[2] M. Barralet, X. Huang, D. Sharma, "Effects of antenna polarization on RSSI based location identification," in 2009 11th International Conference on Advanced Communication Technology, 260–265, 2009.
[3] STMicroelectronics, "BlueNRG-2, Bluetooth® Low Energy wireless system-on-chip," (December), 2020.
[4] STMicroelectronics, "How to use the BlueNRG-Tile Bluetooth LE enabled sensor node development kit," (June), 1–44, 2020.
[5] STMicroelectronics, "STBLESensor Data brief BLE sensor application for Android and iOS," (May), 2019.
[6] STMicroelectronics, "LSM6DSO Datasheet," (January), 2019.

[7] STMicroelectronics, "LIS2MDIL Datasheet," (November), 1–41, 2018.
[8] F. Awad, A. Omar, M. Naserllah, A. Abu-Hantash, A. Al-Taj, "Access point localization using autonomous mobile robot," in 2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), 1–5, 2017, doi:10.1109/AEECT.2017.8257754.
[9] Saverio Grutta, "Application Note : Radio communication range estimation in ISM band," (November), 1–47, 2020.
[10] M. Ayadi, A. Ben Zineb, "Body Shadowing and Furniture Effects for Accuracy Improvement of Indoor Wave Propagation Models," IEEE Transactions on Wireless Communications, **13**(11), 5999–6006, 2014, doi:10.1109/TWC.2014.2339275.
[11] N. Chithirala, B. Natasha, N. Rubini, A. Radhakrishnan, "Weighted Mean Filter for removal of high density Salt and Pepper noise," in ICACCS 2016 - 3rd International Conference on Advanced Computing and Communication Systems: Bringing to the Table, Futuristic Technologies from Arround the Globe, 1–4, 2016, doi:10.1109/ICACCS.2016.7586326.
[12] A. Kishk, "Fundamentals of Antennas," 2009.
[13] STMicroelectronics, "STSW-BNRG001 Data brief BlueNRG current consumption estimation tool STSW-BNRG001," (December), 1–5, 2020.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 334-340 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Theoretical study for Laser Lines in Carbon like Zn (XXV)

Nahed Hosny Wahba[*,1], Wessameldin Salah Abdelaziz[1], Tharwat Mahmoud Alshirbeni[2]

[1]*Department of Laser Applications in Metrology, Photochemistry, and Agriculture, National Institute of Laser Enhanced Sciences, Cairo University, Giza, 12613, Egypt*

[2]*Physics Department, Faculty of Science, Cairo University, Giza 12613, Egypt*

ARTICLE INFO

Article history:
Received: 06 July, 2021
Accepted: 08 August, 2021
Online: 16 August, 2021

Keywords:
FAC
C-like Zn (XXV)
Coupled Rate Equation
Reduced Population
Radiative Decay
Doppler Broadening Equation
Gain Coefficient

ABSTRACT

The energy states, transitions chances, oscillator intensities, and collision intensities were computed with FAC (fully relativistic flexible atomic code) program. The calculated results were utilized for identification of the reduced population to sixty-nine thin structural states in C-like Zn (XXV) and indicates the gain coefficients with several electron densities (from 10^{+20} to 10^{+22} cm^3) and at a wide range of electron plasma temperatures (700,800,900,1000, &1100,1200,1300,1400,1500) eV. By using coupled rate equation to calculate the reduced population at different temperature and plotting that against electron densities; gives that at lower electron densities the reduced population proportional with reduced population till radiative decay happening; while at higher electron densities than 10^{+20} the radiative decay may be neglected in comparing with collisional depopulation so population states becomes independent and approximately the same. The gain coefficient was calculated by using the Doppler broadening equation of several transitions in Zn(XXV); these data plotted against electron density, and it was found that the gain was increased with temperature and producing the short wavelength laser , between 22 and 50 nm for the Zn^{30+}ion. The data was compared with the experimental calculations values collected by NIST and with the theoretical calculations of Bhatia, Seely & Feldman; where the calculated data differs from energy levels of Zn (XXV) comparing to experimental values in NIST at $(2p_{1/2}\,2p_{3/2})_1$ and $(2p_{1/2}\,2p_{3/2})_2$ by 0.05 and 0.04 successively; and it differs than the theoretical work of Bhatia at $(2p_{1/2}\,2p_{3/2})_1$ and $(2p_{1/2}\,2p_{3/2})_2$ by 0.05 Ryd and 0.04 Ryd successively also; which proved that our calculations are in well agreement with other works.

1. Introduction

X-ray lasers are a class of lasers in which gain has been demonstrated over various discrete wavelengths ranging from 3.56nm to 46.9nm. Because of the very short-duration and high-energy excitation pulses required to generate these lasers [1], [2], photo excitation method [3], Electron collisional pumping method (ECP), charge transfer technique, electron collisional recombination process and dielectronic recombination pumping are examples of X-ray pumping procedures which using picoseconds chirped pulse amplification (CPA) pulses [4]-[6]. Globally it's often observed that carbon is abundant element in astrophysical sits having the atmosphere. Emission lines of C-like ions are functionalized at prosopopoeia of the solar, astrophysical and melting plasmas whose illustrating needed exact atomic calculations; where the soft X-ray and XUV regions most of the data was found[7], [8]; thus Electron Collisional Pumping was functionalized to generate soft X-ray lasers after pumping methods[9], [10].

The process of pumping was illustrated as following:

$$X_l^{n+} + e \longrightarrow X_u^{n+}$$

where X_l^{n+} is a n-frequencies of atom ionization of the element X that pumping occurrence from lower level "l" to an excited level "u" in the same element atoms.

Theoretically there are more works done for computing the energy states, transition possibilities' and oscillator powers for Zn (XXV) [11]-[17]; while the gain for the same element not have

[*]Corresponding Author: Nahed Hosny Wahba, Department of Laser Applications in Metrology, Photochemistry, and Agriculture, National Institute of Laser Enhanced Sciences, Cairo University, Giza, 12613, Egypt
Email: nahedwahba77@gmail.com

more studies. The goal of this thesis is to utilize the atomic calculations such as energy states, oscillator powers and spontaneous radiative decay rates which calculated by using (FAC) program depending on Dirac equation for sixty nine thin-structure states to compute reduced populations and gain coefficients of C-like Zn excited states through a broad extent of electron densities (10^{+20} to10^{+23}) and at several electron temperatures (700, 800, 900, 1000, 1100, 1200, 1300, 1400 & 1500).These calculations might support the experimentalists for generating soft X-ray lasers.

2. Calculations equations used for Gain Coefficient determination

To calculate gain coefficient firstly energy levels, weighted oscillator strength and radiative rate for allowed transitions should be calculated; then the reduced population should calculated by solving coupled rate equation [18], [19]. After calculating the reduced population, it used to solve the Doppler broadening equation to obtain the gain coefficient.

Laser emission from Zn (XXV) ions plasma was investigated by studying the relation between several plasma temperatures and electron densities.

According to equation (1)

$$N_u\left[\sum_{l<u} A_{ul} + N_e\left(\sum_{l<u} C^d_{ul} + \sum_{l>u} C^e_{ul}\right)\right] = N_e\left(\sum_{l<u} N_l\, C^e_{lu} + \sum_{l>u} N_l\, C^d_{lu}\right) + \sum_{l>u} N_l\, A_{lu} \quad (1)$$

Since N_u and N_l are the fractional populations of levels u and l successively, A_{ul} represents Einstein coefficient for spontaneous radiative decay from u to l; N_e represents the electron density and C^e_{lu} and C^d_{ul} are the rate coefficients for collisional excitation and de-excitation successively. The actual population density N_u of the u^{th} state can be computed from relation (2) [20][21].

$$C^d_{ul} = C^e_{lu}\left[\frac{g_l}{g_u}\right]\exp\left[\frac{\Delta E_{ul}}{kT_e}\right] \quad (2)$$

Since g_l and g_u represents a statistical weights of lower and upper states, successively.

The electron impact excitation rates identified by the effective collision strengths γ_{ul} [20]Where;

$$C^d_{ul} = \frac{8.6287 * 10^{-6}}{g_u T_e^{1/2}}\gamma_{lu} \quad (3)$$

The measured population density N_u of the u^{th} was calculated [20],

$$N_U = N_u * N_l \quad (4)$$

Since N_L is the number of ions which achieved at the ionization stage L [20],

$$N_L = f_L\, {N_e}/{Z_{avg}} \quad (5)$$

Since N_e is the electron density, Z_{avg} is the average degree of ionization and f_L is the fractional abundance of the ionization levels were calculated [20]. Where the populations computed from Equation (1) is equal the unit;

$$\sum_{U=1}^{69} \frac{N_u}{N_l} = 1 \quad (6)$$

where the populations density calculated by Equation (1) is equal unit,

By computation the state's population density, the values N_u/g_u and N_l/g_l can be determined.

To prove that when inversion factor (F>0) gives positive gain equation (7) was used[22];

$$F = \frac{g_u}{N_u}\left[\frac{N_u}{g_u} - \frac{N_l}{g_l}\right] \quad (7)$$

Since N_u/g_u andN_l/g_l are the reduced populations of the upper state and lower state successively. Then Eq. (7) used to compute the gain coefficient (α) for Doppler broadening of the various transitions in the Zn (XXV) ion.

$$\alpha_{ul} = \frac{\lambda^3_{lu}}{8\pi}\left[\frac{M}{2\pi K T_l}\right]^{1/2} A_{ul} N_u\, F \quad (8)$$

Since M is the ion mass, λ_{lu} is the transition wavelength in (nm), and T_l is the ion temperature in eV.

3. Results and discussions

3.1. Energy states

With utilizing (FAC) [23] energy state measures for the $1s^2$ $2s^2$ 2pnl (n=3, l=s, p & d) and ml (m=4,l=s, p, d &f) configurations in C-like Zn^{+30} was obtained, this data presented in Tables (1); which shows the 69 energy levels of transition configurations:

Table (2) presented the comparison between our calculations of energy levels for Zn (XXV) the theoretical calculations by Bhatia, Seely and Feldman [12] and the actual results computed by NIST [24].

In table (2), the calculated data for energy levels of Zn (XXV) comparing to experimental values in NIST at $(2p_{1/2}\ 2p_{3/2})_1$ and $(2p_{1/2}\ 2p_{3/2})_2$ by 0.05 and 0.04 successively; and it differs than the theoretical work of Bhatia at $(2p_{1/2}\ 2p_{3/2})_1$ and $(2p_{1/2}\ 2p_{3/2})_2$ by 0.05 Ryd and 0.04 Ryd successively also; which proved that our calculations are in well agreement with other works.

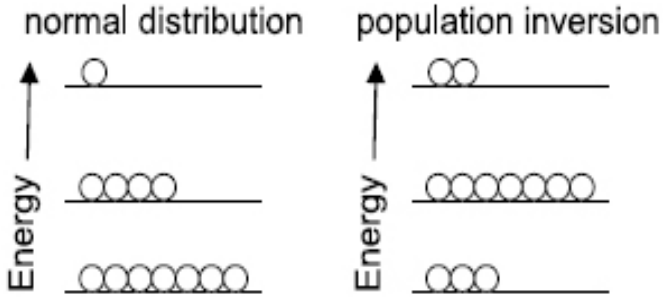
3.2. Level population

Where increasing the excited electrons in higher energy states than in ground state causes the production of Laser in the XUV and soft X-ray spectral area;

Table 1: Energy states and definitions for Zn (XXV)

index	State configuration	Energy in (Ryd)*	Index	State configuration	Energy in (Ryd)*
1	$(2p_0)_0$	0	36	$(2p_{1/2}\,4p_{3/2})_2$	133.053
2	$(2p_{1/2}\,2p_{3/2})_1$	1.3831	37	$(2p_{1/2}\,4p_{1/2})_0$	133.075
3	$(2p_{1/2}\,2p_{3/2})_2$	1.9445	38	$(2p_{1/2}\,4s_{1/2})_2$	133.754
4	$(2p_2)_2$	3.8985	39	$(2p_{3/2}\,4s_{1/2})_1$	133.886
5	$(2p_0)_0$	5.6409	40	$(2p_{1/2}\,4d_{3/2})_2$	133.979
6	$(2p_{1/2}\,3s_{1/2})_0$	97.617	41	$2p_{1/2}\,4d_{5/2})_2$	133.994
7	$(2p_{1/2}\,3s_{1/2})_1$	97.715	42	$(2p_{1/2}\,4d_{5/2})_3$	134.003
8	$(2p_{3/2}\,3s_{1/2})_2$	99.576	43	$(2p_{1/2}\,4d_{3/2})_1$	134.016
9	$(2p_{3/2}\,3s_{1/2})_1$	99.657	44	$(2p_{3/2}\,4p_{1/2})_1$	134.400
10	$(2p_{1/2}\,3p_{1/2})_1$	99.907	45	$(2p_{3/2}\,4p_{3/2})_3$	134.424
11	$(2p_{1/2}\,3p_{3/2})_2$	100.488	46	$(2p_{3/2}\,4p_{1/2})_2$	134.440
12	$(2p_{1/2}\,3p_{3/2})_1$	100.490	47	$(2p_{3/2}\,4p_{3/2})_1$	134.447
13	$(2p_{1/2}\,3p_{1/2})_0$	100.568	48	$(2p_{1/2}\,4f_{5/2})_3$	134.873
14	$(2p_{3/2}\,3p_{3/2})_1$	102.038	49	$(2p_{1/2}\,4f_{5/2})_2$	134.921
15	$(2p_{3/2}\,3p_{3/2})_3$	102.206	50	$(2p_{1/2}\,4f_{7/2})_3$	134.996
16	$(2p_{3/2}\,3p_{1/2})_1$	102.230	51	$(2p_{1/2}\,4f_{7/2})_4$	135.019
17	$(2p_{3/2}\,3p_{1/2})_2$	102.275	52	$(2p_{3/2}\,4p_{3/2})_2$	135.251
18	$(2p_{1/2}\,3d_{3/2})_2$	102.289	53	$(2p_{3/2}\,4p_{1/2})_0$	135.496
19	$(2p_{3/2}\,3p_{3/2})_2$	102.812	54	$(2p_{3/2}\,4d_{5/2})_4$	135.854
20	$(2p_{1/2}\,3d_{5/2})_3$	102.842	55	$(2p_{3/2}\,4d_{3/2})_2$	135.866
21	$(2p_{1/2}\,3d_{5/2})_2$	102.979	56	$(2p_{3/2}\,4d_{3/2})_3$	135.928
22	$(2p_{1/2}3d_{3/2})_1$	103.056	57	$(2p_{3/2}\,4d_{5/2})_2$	135.993
23	$(2p_{3/2}\,3p_{3/2})_0$	103.722	58	$(2p_{3/2}\,4d_{3/2})_1$	136.004
24	$(2p_{3/2}\,3d_{5/2})_4$	104.441	59	$(2p_{3/2}\,4d_{3/2})_0$	136.005
25	$(2p_{3/2}\,3d_{3/2})_2$	104.474	60	$(2p_{3/2}\,4d_{5/2})_3$	136.192
26	$(2p_{3/2}\,3d_{5/2})_3$	104.717	61	$(2p_{3/2}\,4d_{5/2})_1$	136.232
27	$(2p_{3/2}\,3d_{5/2})_2$	104.892	62	$(2p_{3/2}\,4f_{5/2})_1$	136.421
28	$(2p_{3/2}\,3d_{3/2})_1$	104.907	63	$(2p_{3/2}\,4f_{7/2})_4$	136.450
29	$(2p_{3/2}\,3d_{3/2})_0$	104.921	64	$(2p_{3/2}\,4f_{5/2})_2$	136.475
30	$(2p_{3/2}\,3d_{5/2})_3$	105.508	65	$(2f_{5/2}\,4f_{7/2})_3$	136.488
31	$(2p_{3/2}\,3d_{5/2})_1$	105.564	66	$(2p_{3/2}\,4f_{7/2})_5$	136.508
32	$(2p_{1/2}\,4s_{1/2})_0$	131.880	67	$(2p_{3/2}\,4f_{5/2})_4$	136.522
33	$(2p_{1/2}\,4s_{1/2})_1$	131.914	68	$(2p_{1/2}\,4f_{5/2})_1$	136.540
34	$(2p_{1/2}\,4p_{1/2})_1$	132.702	69	$(2p_{3/2}\,4f_{7/2})_2$	136.580
35	$(2p_{1/2}\,4p_{3/2})_1$	133.037			

* Ryd is Rydberg constant



Schematic diagram of population inversion (Source of figure: https://spie.org/publications/fg08_p94_lasers?SSO=1)

index	State Configuration	Our calculation (FAC)[a]	SS[b]	NIST[c]
1	$(2p_0)_0$	0	0	0
2	$(2p_{1/2}\,2p_{3/2})_1$	1.3831	1.4372	1.4370
3	$(2p_{1/2}\,2p_{3/2})_2$	1.9445	1.9866	1.9870
4	$(2p_2)_2$	3.8985	3.9122	...
5	$(2p_0)_0$	5.6409	5.3061	...
6	$(2p_{1/2}\,3s_{1/2})_0$	97.617	98.206	...
7	$(2p_{1/2}\,3s_{1/2})_1$	97.715	98.306	...
8	$(2p_{3/2}\,3s_{1/2})_2$	99.576	100.152	...
9	$(2p_{3/2}\,3s_{1/2})_1$	99.657	100.395	...
10	$(2p_{1/2}\,3p_{1/2})_1$	99.907	100.154	...
11	$(2p_{1/2}\,3p_{3/2})_2$	100.488	101.035	...
12	$(2p_{1/2}\,3p_{3/2})_1$	100.490	101.040	...
13	$(2p_{1/2}\,3p_{1/2})_0$	100.568	101.143	...
14	$(2p_{3/2}\,3p_{3/2})_1$	102.038	102.522	...
15	$(2p_{3/2}\,3p_{3/2})_3$	102.206	102.670	...
16	$(2p_{3/2}\,3p_{1/2})_1$	102.230	102.752	...
17	$(2p_{3/2}\,3p_{1/2})_2$	102.275	102.728	...
18	$(2p_{1/2}\,3d_{3/2})_2$	102.289	103.503	...
19	$(2p_{3/2}\,3p_{3/2})_2$	102.812	102.847	...
20	$(2p_{1/2}\,3d_{5/2})_3$	102.842	104.155	...
21	$(2p_{1/2}\,3d_{5/2})_2$	102.979	103.399	...
22	$(2p_{1/2}3d_{3/2})_1$	103.056	103.443	...
23	$(2p_{3/2}\,3p_{3/2})_0$	103.722	103.589	...
24	$(2p_{3/2}\,3d_{5/2})_4$	104.441	104.934	...
25	$(2p_{3/2}\,3d_{3/2})_2$	104.474	104.973	...
26	$(2p_{3/2}\,3d_{5/2})_3$	104.717	105.2..7	...
27	$(2p_{3/2}\,3d_{5/2})_2$	104.892	105.415	...
28	$(2p_{3/2}\,3d_{3/2})_1$	104.907	105.382	...
29	$(2p_{3/2}\,3d_{3/2})_0$	104.921	105.390	...

Thus the process of the reduced population densities was computed for sixty nine thin structure states starting from $1s^2\ 2s^2$ 2pnl (n=3, l=s, p&d) and ml (m=4, l=s, p, d &f) configurations. The determination was done by applying the coupled rate Eq. (1) simultaneously using MATLAB version 7.10.0 (R2010a) computer program [25][17].

Figure (1 to 4) illustrate the reduced population for states $(2p_{3/2}3s_{1/2})_2$, $(2p_{1/2}3p_{3/2})_1$, $(2p_{3/2}3p_{3/2})_3$, $(2p_{3/2}3p_{3/2})_1$, $(2p_{3/2}3d_{5/2})_4$, and $(2p_{3/2}3d_{3/2})_3$at various temperatures (800,900,1000,1100)eV; so it can explain the behavior of states populations' density for several ions; where at low electron densities the reduced population densities are proportional to the electron densities, and the excitation process for an excited state is followed immediately by radiation decay. These results were agreed with the results of Feldman et.al. [11,17,24]. At electron density 10^{+19} various peaks were appeared; which means that radiative transitions dominant the de-excitation due its higher energy and fast decay time.

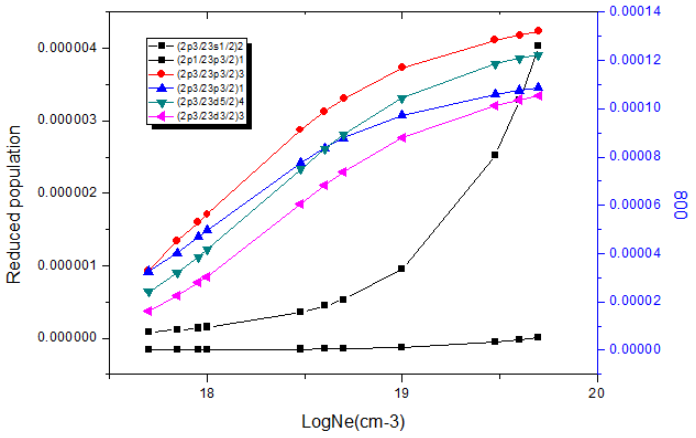


Figure 1: Reduced population of Zn^{+30} states at electron temperature 800eV.

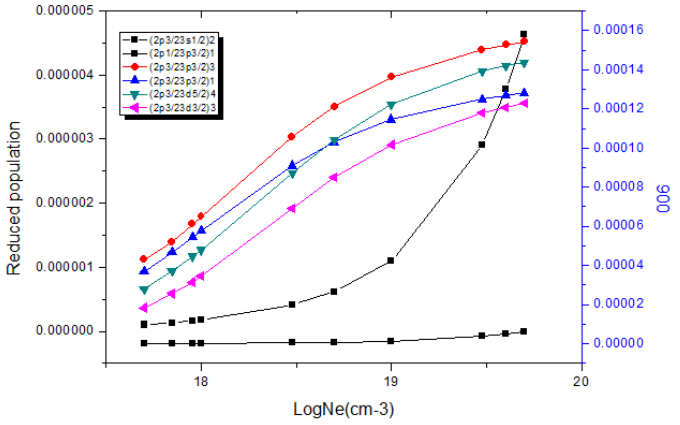


Figure 2: Reduced population of Zn^{+30} states at electron temperature 900eV.

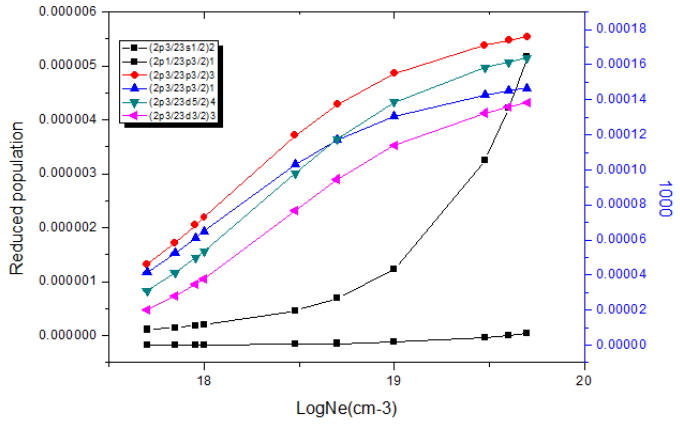
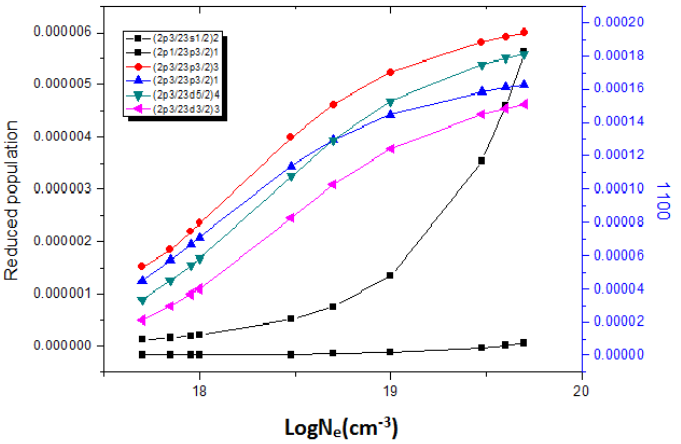
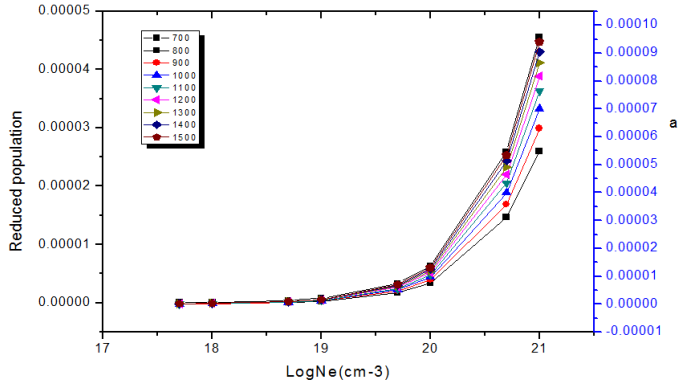


Figure 3: Reduced population of Zn^{+30} states at electron temperature 1000eV.





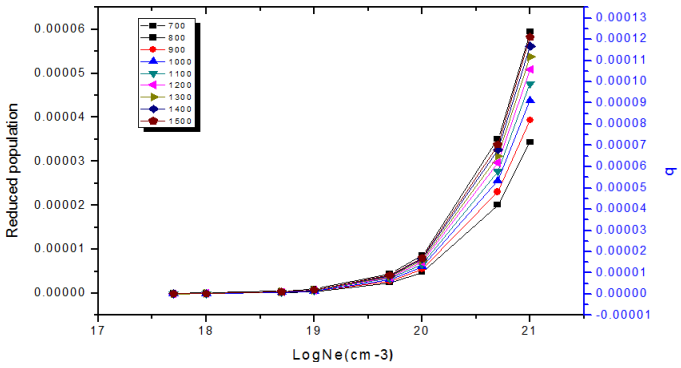


Figure 5: Reduced population of level (a) $(2p_{1\backslash2}3p_{1\backslash2})_1$, (b) $(2p_{1\backslash2}3p_{3\backslash2})_1$ for Zn (XXV) after electron collisional pumping as a function of the electron density at temperatures (700, 800, 900, 1000, 1100, 1200, 1300, 1400&1500) eV.

3.3. Radiative lifetime

atomic transfer probability is related to the life time u of an exited state

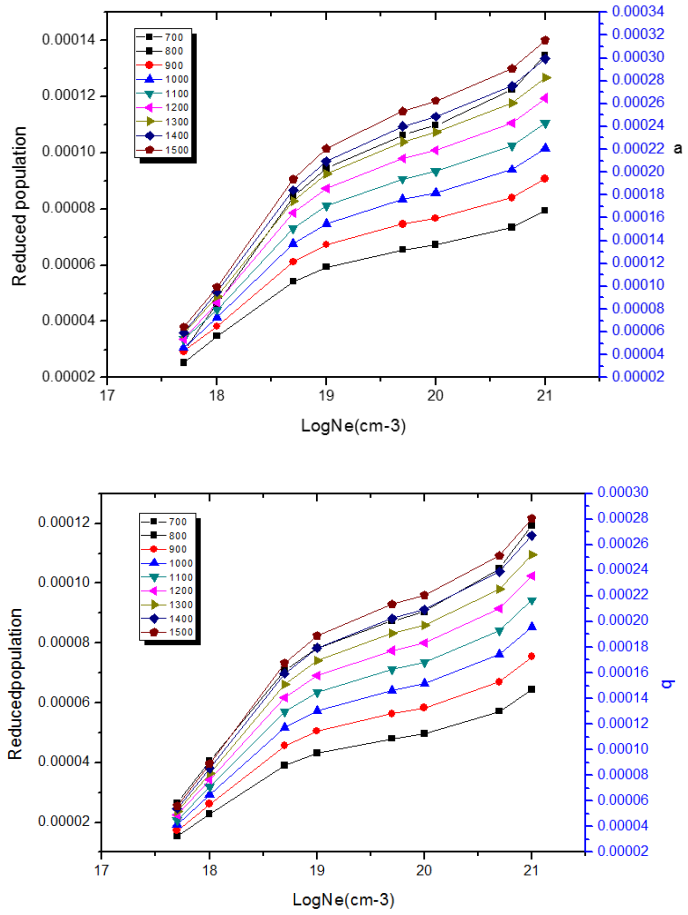


Figure 6: Reduced population of level (a) $(2p_{3/2}3p_{3/2})_3$, (b) $(2p_{3/2}3p_{3/2})_1$ for Zn (XXV) as a function of the electron density at different electron temperatures (700,800, 900, 1000, 1100, 1200, 1300, 1400&1500) eV.

$$\tau_u = \frac{1}{\sum_l A_{ul}} \quad (9)$$

Table 3. Illustrate the results of $(2p_{1/2}3d_{3/2})_2$--$(2p_{1/2}3p_{3/2})_1$, $(2p_{3/2}3d_{5/2})_4$—$(2p_{1/2}3p_{3/2})_1$and $(2p_{3/2}3d_{5/2})_4$—$((2p_{1/2}3p_{3/2})_2$radiative life time is longer than the lifetime of the lower state.

Configuration	τ_u(sec)	τ_l(sec)
$(2p_{3/2}3p_{3/2})_3$--$(2p_{1/2}3p_{1/2})_1$	8.926e-10	7.495e-10
$(2p_{1/2}3d_{3/2})_2$--$(2p_{1/2}3p_{3/2})_1$	9.894e-10	3.493e-12
$(2p_{3/2}3p_{3/2})_1$—$(2p_{1/2}3p_{3/2})_1$	2.578e-13	3.493e-12
$(2p_{3/2}3d_{5/2})_4$—$(2p_{1/2}3p_{3/2})_1$	2.104e-9	3.493e-12
$(2p_{3/2}3d_{5/2})_4$—$((2p_{1/2}3p_{3/2})_2$	2.104e-9	2.282e-10
$(2p_{3/2}3d_{3/2})_3$—$(2p_{1/2}3d_{3/2})_2$	5.597e-14	9.894e-10

3.4. Inversion factor

According to equation (7) the reduced population for lower states and upper states was calculated and demonstrate in the equation to calculate the inversion factor and it's found that the inversion factor is larger than zero. By using electron collisional pumping process the pumping quanta can be transferred to other state as a result of collision process, and this cause population inversion from the upper states to the lower states; whence this population inversion achieved appositive gain via F>0[21].

3.5. Gain coefficient

The gain process is the measure of the part of medium energy transferred to the emitted radiation which causes the amplification of the emitted radiation leading to strength optical power.

To calculate the gain the MATlab version the program was used to solve the coupled rate equation; this by using A_{ul} (spontaneous decay rates), C^e_{lu} (electron collisional excitation rate coefficients) and C^d_{ul} (electron collisional deexcitation rate coefficients).

$$> 0.$$

Finally the Doppler broadening equation was solved for various transitions to give the gain coefficient; then by plotting the relation between gain and electron density at different temperature to obtain the most intense laser transitions.

The figures (7, 8, 9, 10 & 11) illustrates the proportional relation between gain and electron density; and also have proportional relation between gain and temperature. According to the collected data it's found that the largest gain occur at temperature (1100eV) which give gain height of (13.522cm^{-1}) at wavelength (50nm); this transition is at $(2p_{3/2}3p_{3/2})_1$—$(2p_{1/2}3p_{3/2})_1$ which refers to them by (16<>9). The smallest gain occur at temperature (800eV) which give gain height of (2.5530cm^{-1}) at wavelength (22.79nm); this gain transition at $(2p_{3/2}3d_{5/2})_4$—$(2p_{1/2}3p_{3/2})_1$which describe them as (22<>10); the gain of these transition at (22<>10) and at (16<>9) was plotted against electron densities at different temperatures. See Figure (11).

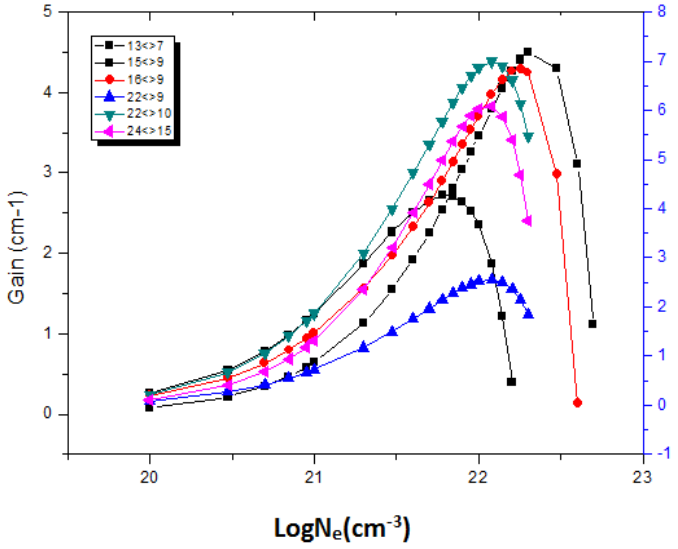


Figure 7: Electron density versus Gain coefficient at temperature 800 eV.

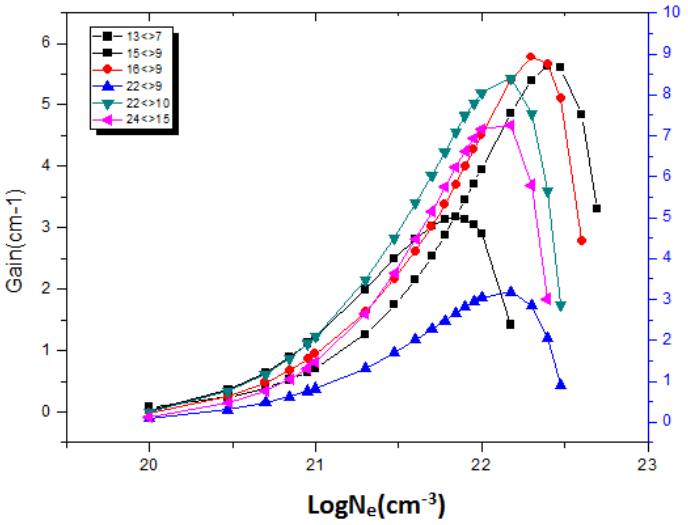


Figure 8: Electron density versus Gain coefficient at temperature 900 eV.

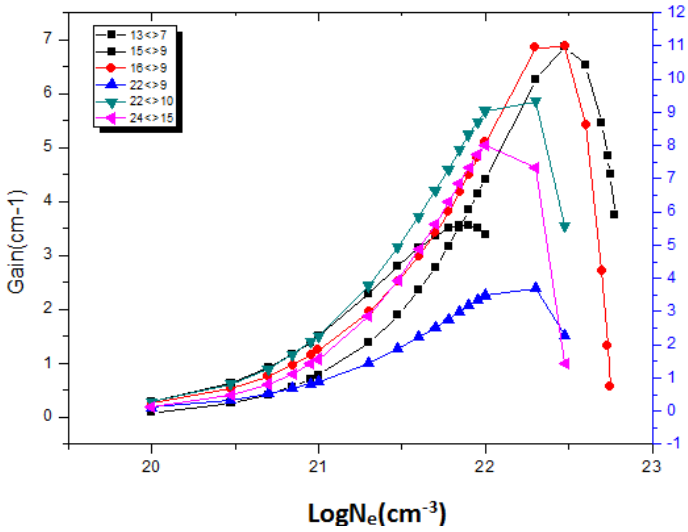


Figure 9: Electron density versus Gain coefficient at temperature1000 eV.

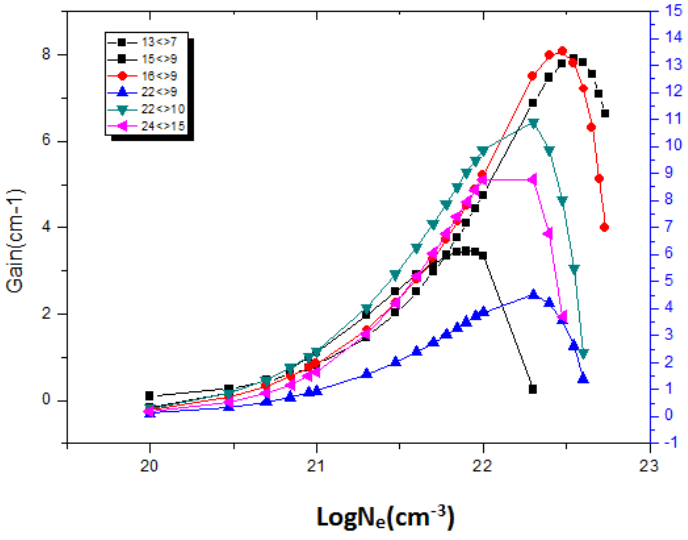
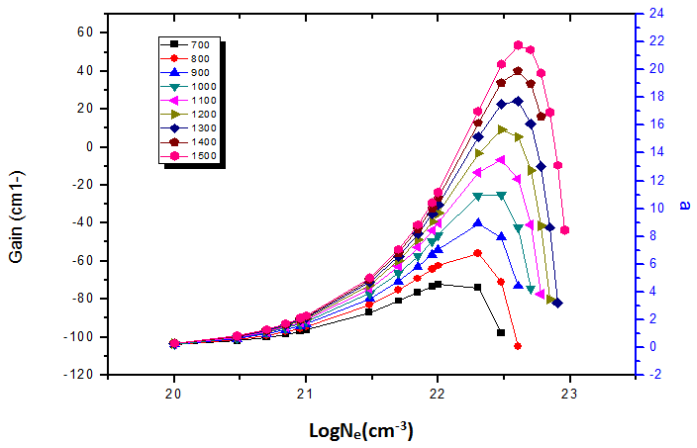


Figure 10: Electron density versus Gain coefficient at temperature1100 eV.



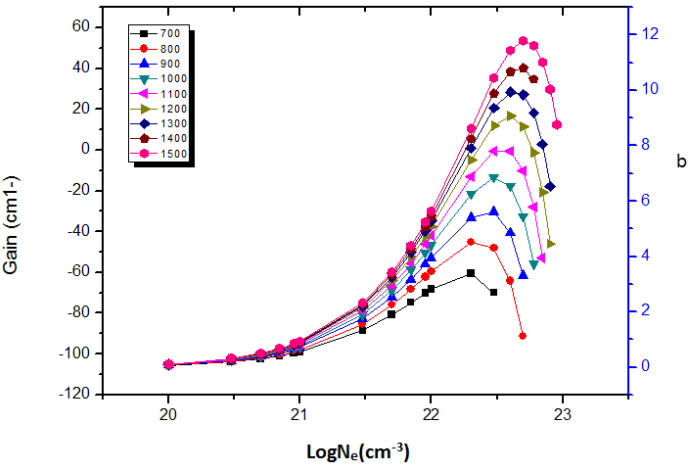


Table 4: configuration states, wavelength and maximum gain coefficient at various temperatures.

Configuration	λ(nm)	Gain(α)(cm-1) Temperature eV								
		700	800	900	1000	1100	1200	1300	1400	1500
$(2p_{3/2}3p_{3/2})_3$--$(2p_{1/2}3p_{1/2})_1$	35.34	3.372	4.504	5.635	6.866	7.909	9.067	9.979	10.814	11.795
$(2p_{1/2}3d_{3/2})_2$--$(2p_{1/2}3p_{3/2})_1$	50.4	3.409	4.266	5.006	5.595	6.118	6.560	6.899	7.156	7.387
$(2p_{3/2}3p_{3/2})_1$—$(2p_{1/2}3p_{3/2})_1$	50	4.811	6.847	8.918	10.977	13.522	15.676	17.869	19.888	21.82
$(2p_{3/2}3d_{5/2})_4$—$(2p_{1/2}3p_{3/2})_1$	22.79	1.938	2.553	3.182	3.711	4.504	5.144	5.811	6.438	7.173
$(2p_{3/2}3d_{5/2})_4$—$(2p_{1/2}3p_{3/2})_2$	22.8	5.469	6.998	8.410	9.327	10.919	12.19	13.325	14.229	15.226
$(2p_{3/2}3d_{3/2})_3$—$(2p_{1/2}3d_{3/2})_2$	36.9	4.685	6.077	7.258	8.017	6.788	10.381	10.978	11.526	12.068

4. Conclusions

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors thanks both Prof. Dr. Souad ElFeky, Dr. Nagy Emara, and NILES their promotion and support.

References:

[1] William.T.Silfvast, Cambridge University press, second edi, 2000.
[2] B.N. Wellegehausen, B., Eichmann, H., Meyer, S., Momma, C., Mossavi, K., Welling, H., &Chichkov, "Generation of coherent VUV and XUV radiation. In ICONO'95: Fundamentals of Laser-Matter Interaction.," International Society for Optics and Photonics., **2796**, 132–139, 1996.
[3] A. Verma, R. Kumar, A. Parashar, "Enhanced thermal transport across a bi-crystalline graphene–polymer interface: an atomistic approach," Physical Chemistry Chemical Physics, **21**(11), 6229–6237, 2019. doi: 10.1039/C9CP00362B
[4] R.E. King, G.J. Pert, S.P. McCabe, P.A. Simms, A.G. MacPhee, C.L.S. Lewis, R. Keenan, R.M.N. O'Rourke, G.J. Tallents, S.J. Pestehe, "Saturated x-ray lasers at 196 and 73 Å pumped by a picosecond traveling-wave excitation," Physical Review A, **64**(5), 53810, 2001. doi: 10.1103/PhysRevA.64.053810
[5] A. V Vinogradov, I.I. Sobel'man, E.A. Yukov, "Population inversion of transitions in neon-like ions," Soviet Journal of Quantum Electronics, **7**(1), 32, 1977.
[6] B.A. Norton, N.J. Peacock, "Population inversion in laser-produced plasmas by pumping with opacity-broadened lines," Journal of Physics B: Atomic and Molecular Physics, **8**(6), 989, 1975.
[7] G. Tachiev, C.F. Fischer, "Breit-Pauli energy levels and transition rates for the carbonlike sequence," Canadian Journal of Physics, **79**(7), 955–976, 2001. doi: 10.1139/p01-059
[8] K.M. Aggarwal, F.P. Keenan, A.Z. Msezane, "Oscillator strengths for transitions in C-like ions between K XIV and Mn XX," Astronomy & Astrophysics, **401**(1), 377–383, 2003.
[9] V.A. Bhagavatula, "Soft x-ray population inversion by resonant photoexcitation in multicomponent laser plasmas," Journal of Applied Physics, **47**(10), 4535–4537, 1976.
[10] J. Nilsen, P. Beiersdorfer, S.R. Elliott, T.W. Phillips, B.A. Bryunetkin, V.M. Dyakin, T.A. Pikuz, A.Y. Faenov, S.A. Pikuz, S. Von Goeler, "Measurement of the Ly-α Mg resonance with the 2s→ 3p Ne-like Ge line," Physical Review A, **50**(3), 2143, 1994.
[11] U. Feldman, G.A. Doschek, J.F. Seely, A.K. Bhatia, "Short wavelength laser calculations for electron pumping in Be I and BI isoelectronic sequences (18≤ Z≤ 36)," Journal of Applied Physics, **58**(8), 2909–2915, 1985.
[12] A.K. Bhatia, J.F. Seely, U. Feldman, "Atomic data and spectral line intensities for the carbon isoelectronic sequence (Ar XIII through Kr XXXI)," Atomic Data and Nuclear Data Tables, **36**(3), 453–494, 1987, doi:https://doi.org/10.1016/0092-640X(87)90012-X.
[13] A. Verma, A. Parashar, "Molecular dynamics based simulations to study the fracture strength of monolayer graphene oxide," Nanotechnology, **29**(11), 115706, 2018, doi:10.1088/1361-6528/aaa8bb.
[14] A. Verma, A. Parashar, "Molecular dynamics based simulations to study failure morphology of hydroxyl and epoxide functionalised graphene," Computational Materials Science, **143**, 15–26, 2018.
[15] V. Singla, A. Verma, A. Parashar, "A molecular dynamics based study to estimate the point defects formation energies in graphene containing STW defects," Materials Research Express, **6**(1), 15606, 2018. doi: 10.1088/2053-1591
[16] A. Verma, A. Parashar, M. Packirisamy, "Atomistic modeling of graphene/hexagonal boron nitride polymer nanocomposites: a review," Wiley Interdisciplinary Reviews: Computational Molecular Science, **8**(3), e1346, 2018. doi: 10.1088/25192018
[17] U. Feldman, J.F. Seely, G.A. Doschek, A.K. Bhatia, "3 s–3 p laser gain and x-ray line ratios for the carbon isoelectronic sequence," Journal of Applied Physics, **59**(12), 3953–3957, 1986.
[18] U. Feldman, A.K. Bhatia, S. Suckewer, "Short wavelength laser calculations for electron pumping in neon-like krypton (Kr XXVII)," Journal of Applied Physics, **54**(5), 2188–2197, 1983.
[19] U. Feldman, J.F. Seely, A.K. Bhatia, "Scaling of collisionally pumped 3 s-3 p lasers in the neon isoelectronic sequence," Journal of Applied Physics, **56**(9), 2475–2478, 1984.
[20] W.H. Goldstein, J. Oreg, A. Zigler, A. Bar-Shalom, M. Klapisch, "Gain predictions for nickel-like gadolinium from a 181-level multiconfigurational distorted-wave collisional-radiative model," Physical Review A, **38**(4), 1797, 1988. doi: 10.1137/083627
[21] A. V Vinogradov, V.N. Shlyaptsev, "Calculations of population inversion due to transitions in multiply charged neon-like ions in the 200–2000 Å range," Soviet Journal of Quantum Electronics, **10**(6), 754, 1980.
[22] I.I.S. Man, Introduction to the theory of atomic spectra, International series of Monographs in Natural Philosophy, 40, Pergamon Press.
[23] [FAC Code. http://kipac-tree.stanford.edu/fac].
[24] NIST [http:///F:/NIST/NIST%20ASD%20Levels%20Output32.htm].
[25] R.D. Neidinger, "Introduction to automatic differentiation and MATLAB object-oriented programming," SIAM Review, **52**(3), 545–563, 2010. 10.1137/080743627

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 341-347 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Power Saving MAC Protocols in Wireless Sensor Networks: A Performance Assessment Analysis

Rafael Souza Cotrim[1], João Manuel Leitão Pires Caldeira[1,2,3,*], Vasco Nuno da Gama de Jesus Soares[1,2,3], Pedro Miguel de Figueiredo Dinis Oliveira Gaspar[4,5]

[1]*Polytechnic Institute of Castelo Branco, Castelo Branco, 6000-084, Portugal*

[2]*Instituto de Telecomunicações, Covilhã, 6201-001, Portugal*

[3]*InspiringSci, Castelo Branco, 6000-767, Portugal*

[4]*Deparment of Electromechanical Engineering, University of Beira Interior, Rua Marquês d'Ávila e Bolama, Covilhã, 6201-001 Portugal*

[5]*Centre for Mechanical and Aerospace Science and Technologies (C-MAST), Rua Marquês d'Ávila e Bolama, Covilhã, 6201-001, Portugal*

ARTICLE INFO

Article history:
Received: 13 June, 2021
Accepted: 08 August, 2021
Online: 26 August, 2021

Keywords:
Wireless Sensor Networks
WSN
MAC protocols
Energy Efficiency
Performance Assessment
Simulation

ABSTRACT

Wireless sensor networks are an emerging technology that is used to monitor points or objects of interest in an area. Despite its many applications, this kind of network is often limited by the fact that it is difficult to provide energy to the nodes continuously, forcing the use of batteries, which restricts its operations. Network density may also lead to other problems. Sparse networks require stronger transmissions and have little redundancy while dense networks increase the chances of overhearing and interference. To address these problems, many novel medium access control (MAC) protocols have been developed through the years. The objective of this study is to assess the effectiveness of the T-MAC, B-MAC, and RI-MAC protocols in a variable density network used to collect data inside freight trucks carrying fruits that perish quickly. This article is part of the PrunusPós project, which aims to increase the efficiency of peach and cherry farming in Portugal. The comparison was done using the OMNET++ simulation framework. Our analysis covers the behavior and energetic properties of these protocols as the density of the network increases and shows that RI-MAC is more adaptable and consumes less energy than the alternatives.

1. Introduction

The world's growing reliance on technology has increased the necessity for data collection. Wireless sensor networks (WSNs) are one of the many technologies that have appeared to fulfill such a niche, allowing greater versatility on what data is collected and how. WSN's applications include improving emergency response [1], control urban lighting [2], control precision irrigation systems [3], monitor patients in healthcare facilities [4], and many others.

A WSN is a network comprised of nodes that collect data about the environment and send it to a data collection system. Although flexible, WSNs have several limitations. Usually, nodes do not have long-range communication capabilities and rely on a gateway, also called a base station, to send the data to its destination. They are also made to be cheap and compact, limiting the hardware that may be used. Finally, nodes do not have a reliable power source, forcing them to use batteries. These factors combined limit a node's battery life, so care must be taken to reduce consumption to a minimum.

The power used in performing the necessary measurements is hard to modify, being mostly dependent on the physical mechanism used to acquire the data. On the other hand, idle listening and transmitting data are some of the most power-intensive tasks on a WSN [5], therefore most optimization efforts have tried to tackle these factors. These attempts have achieved variable success through techniques such as duty cycling [6], the use of separate communication channels to wake up nodes [7],

*Corresponding Author: João Manuel Leitão Pires Caldeira, jcaldeira@ipcb.pt

https://dx.doi.org/10.25046/aj060438

reworking the medium access control (MAC) protocol [8], and many others [9].

Attempts to implement novel MAC protocols have been particularly prevalent [10] because the MAC sublayer controls when transmissions are sent and is responsible for avoiding collisions, which force data to be retransmitted. A protocol's behavior can be more adaptable than then a mechanism implemented on the radio level, allowing for the specialization of protocols into certain specific domains such as mobility [11] and others. Finally, data reduction, adaptive sampling, and data prediction techniques required additional coordination between nodes and may depend on the physical nature of the variable. These facts coupled with the availability of tools for simulations are the reason why this paper will focus on novel MAC protocols and their impact on energy consumption.

The environment and node density may also exacerbate the detriments of idle listening and retransmission [12]. Sparse networks have a much greater average distance between nodes, reducing interference between transmission and overhearing, but increasing the minimum power of the radios. On the other side of the scale, dense networks contain large quantities of nodes confined in a relatively small area. At its limit, a dense WSN may behave like a fully connected network. This increases the probability of collisions and means that many nodes around a transmitter will over-hear the radio signals. Because of these differences, sparse and dense networks have different requirements for optimal operations.

The purpose of this paper is to compare MAC protocol for use in variable density wireless sensor networks. More specifically, their use on freight trucks to gather data about cargo temperature and other properties during transport. The scenario was developed following a survey of MAC protocols [13] and as part of the PrunusPós project [14], which aims to extend the shelf life of peaches and cherries in the Beira Interior region in Portugal. These fruits are highly seasonal and deteriorate rapidly after harvest. Storage under controlled temperature and humidity can slow down their decay, but even slight variations may compromise this process. In the proposed scenario, the sensors have been integrated into the crates or other containers used to store and transport the product, which allows for them to provide continuous feedback on the ambient conditions. Such a system facilitates individualized data collection from the moment the fruits are packaged to their delivery.

This granularity is desirable because it allows historic data to be tracked even though products in the same store shelf can come from many different producers and take many different paths while flowing through national and local distribution networks. For example, if a problem happens to a specific batch of products, the companies involved could look through the data to determine exactly where the lapse in their process has happened. This individualized approach also means that other sensors could be added to track variables and phenomena which are more localized than temperature, meaning that such a system could be adapted for a variety of products.

One of the problems of the proposed setup is that the number of nodes in the space can vary according to the size of the container and how full the truck is. The main objective of this work is to analyze the behavior of several MAC protocols under a variety of node densities that could be expected in this application. For our comparison, we have simulated the Timeout MAC (T-MAC) [15], Berkley MAC (B-MAC) [16], and Receiver Initiated MAC (RI-MAC) [17] protocols. Together, these protocols cover the main types of MAC protocols available today and allow us to identify the weaknesses of each approach when compared to the others. The simulations have been done using the OMNeT++ discrete simulation framework [18], which is a robust simulation tool that has been used in previous research [19,20], is updated frequently, and has extensive documentation.

The rest of this paper is organized as follows: Section 2 overviews the related work and explains the basic operations of the protocols compared. Section 3 explains how the simulation environment was set up goes over the results of the simulations, detailing the behavior of each protocol. Finally, section 4 concludes the paper and provides directions for future work.

2. Related Work

There have been many MAC protocols developed over the years for specific applications. Some of the first ones that were tailor-made for WSNs focused on making defining schedules with active and inactive parts for each node, creating a period where whole sections of the network could sleep. Sensor MAC (S-MAC) [21] and T-MAC [15] are the most notable in this category. Both work by giving nodes wake-up/sleep schedules and synchronizing them as they enter the network. T-MAC, however, attempts to reduce the time a node remains awake by utilizing a timeout period. If a node does not receive any transmissions during a timeout window, it will assume all data has been sent and will go back to sleep. Otherwise, it restarts the timer and continues listening to the medium. S-MAC's and T-MAC's procedures are visualized in Figure 1.

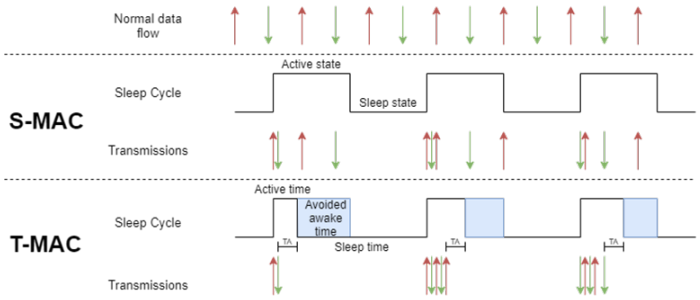


Figure 1: S-MAC and T-MAC comparison, adapted from [15]

The Demand Wakeup MAC protocol (DW-MAC) [22] is an alternative synchronous protocol, however, it does not utilize the active part of the schedule to send and receive messages. It replaces CTS/RTS messages with scheduling messages (SCH), which the nodes use to choose a moment during the sleeping section of the frame where they can communicate without the risk of interferences. This scheduling is done based on the time the SCH frame was received, meaning that no two messages can be scheduled for reception by the same node at the same time.

Although effective, these approaches require a synchronization mechanism to prevent schedule drift, which adds complexity and extends the time the radio module is active. They are also less effective when multiple schedules are being used in the network,

especially in very dense ones [15], a phenomenon called virtual clustering. Finally, one of the main flaws of synchronous protocols is that all the nodes wake up and contend for the medium at the same time. DW-MAC shifts when the data is sent, but the nodes still need to contend for the medium when the active part of the schedule begins. This means that there is a burst of activity in the beginning and, in networks restricted to small areas, that leads to only a few nodes being able to communicate at a time, with the rest of the network waiting for the medium to become free once more.

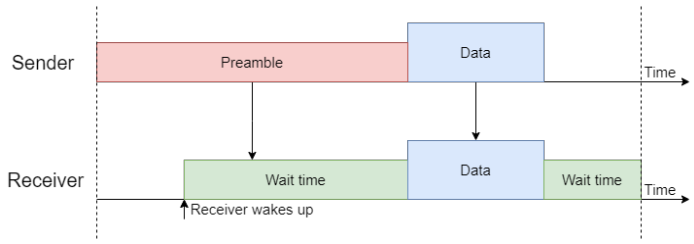


Figure 2: LPL communication example, adapted from [23].

The B-MAC protocol [16] improves on some of these concerns. It is a specialized MAC protocol that reduces energy consumption by allowing nodes to have independent activity schedules. Nodes that have data to transmit send preambles slightly longer than the sleeping period of the receiver. When the destination node wakes up, it samples the medium and, if it detects a preamble, it remains awake. Once the preamble has ended, the sender transmits the data with the destination identifier. This process is known as Low Power Listening (LPL) and it allows nodes to have completely independent schedules. The procedure is illustrated in Figure 2. LPL has been shown to considerably reduce energy consumption when compared to other mechanisms. B-MAC addresses many of the problems synchronous protocols have by not requiring a schedule, which eliminates the necessity for synchronization mechanisms and means that it is not affected by the formation of virtual clusters. However, B-MAC's long preamble leads to the same problem the other cited protocols have where a few nodes monopolize the medium, preventing nearby nodes from transmitting data in the meantime.

More recent protocols have explored other paradigms. As shown in Figure 3, in the RI-MAC protocol [17], the receiver initiates the data transfer by sending a beacon message to indicate to the sender nodes that it is available to receive data. This reduces the time a node occupies the medium and increases the maximum throughput. It also avoids sending the long preamble messages associated with LPL and other asynchronous strategies. However, this change in procedure can lead to problems when the communications channel is asymmetric, meaning that messages being sent in one of the communication directions has a lower chance of being received because of interference or other factors.

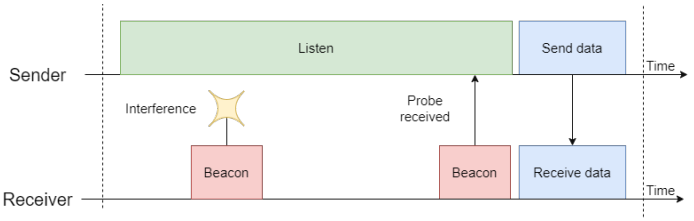


Figure 3: RI-MAC communication example, adapted from [8].

In cases where communication channels are very asymmetric, the Asymmetric MAC protocol (Asym-MAC) can reduce their impact [8]. Asym-MAC is very similar to RI-MAC, but it has two modes. Each communication attempt is started in R-mode, which operates like RI-MAC. However, if the sender does not receive a probe message within a certain period, indicating that it is being lost frequently, the communication will change to T-mode, where the sender initiates the communication. This prevents repeated loss of the beacon and restores communications in asymmetric channels, but Asym-MAC is slightly worse than RI-MAC in terms of energy consumption when the level of asymmetry is very low.

Much like Asym-MAC, the A-MAC protocol attempts to improve on RI-MAC [24]. It utilizes a different link-layer primitive, the backcast, to allow multiple nodes to be probed at the same time and reliably detect when it receives more than one reply at the same time, allowing it to better decide when to go back to sleep. A-MAC also allows nodes to utilize multiple frequencies to communicate, which increases the total throughput of the network and means that beacon messages may be segregated to a different frequency band to prevent interferences. While A-MAC is more effective than RI-MAC, it requires radios with memory-mapping and other features to work properly. A-MAC can still be used with other radios; however, it is less efficient and requires workarounds depending on the architecture of the hardware.

To test the effectiveness of various strategies in the proposed context, one protocol of each type was chosen. T-MAC was selected over the other alternatives because it makes various improvements without leading to additional drawbacks. B-MAC was chosen because it is one of the most robust asynchronous protocols. It has been used in multiple real-world applications and there are reliable implementations for TinyOS, an operational system for embedded systems. Despite its effectiveness, A-MAC's hardware requirements often conflict with available equipment, which is made to be cheap and easily replaceable. On the other hand, Asym-MAC's gains in asymmetric communication channels are not applicable in the proposed scenario. Outside interferences are dampened because the truck acts as a Faraday cage and there are no identifiable internal factors that could cause a high level of asymmetry. Considering these factors, RI-MAC was selected to be added to the simulations.

3. Performance Assessment

3.1. Network Settings

Figure 4 illustrates the proposed scenario. Nodes were integrated into the containers used to carry cargo inside a truck. These sensors measure the temperature regularly and transmit the data to a gateway that uses the truck's radio to send the data to its destination. In real-world applications, the nodes could also measure other parameters to guarantee the safety and quality of the products. The density of the network in this scenario can vary according to the size of the container, how full the truck is, and how the boxes were arranged. To reduce complexity, the parameters of the protocols are not adjusted depending on how the truck is loaded, meaning that protocols must be flexible to accommodate a wide range of densities.

To measure the effectiveness of each protocol, the scenario was built on the OMNET++ [18] simulator and the INET framework

[25] was used to handle wireless communications. The default B-MAC implementation from INET was used in these simulations. RI-MAC was implemented following the structure outlined in [17]. The original paper describing T-MAC leaves many questions unsolved about how the protocol should work [26], so our implementation was based on the one present in the Castalia Simulator [27], which is built on top of OMNET++.

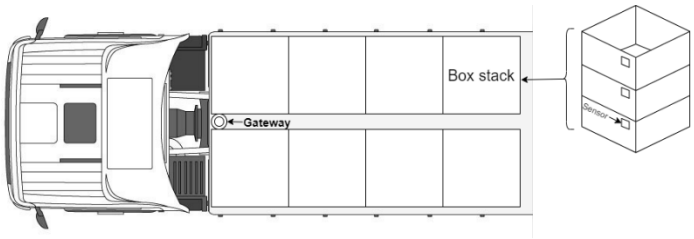


Figure 4: Proposed scenario.

As shown in Table 1, a variable number of nodes was scattered in a 2.2m by 15.75 m area to simulate the restricted environment where they would be deployed. Table 2 contains the power consumption of the various radio states used in the simulations, which were chosen to model the ESP8266EX Wi-Fi microchip [28]. For the purpose of this article, the energy consumed performing the measurements was ignored so that only the power spent by the normal function of each protocol is measured.

Table 1: Simulation parameters.

Simulation Parameter	Value	Unit
Area height	2.2	m
Area width	15.75	m
Sapling interval	100	s
Data length	32	B

Table 2: Radio power consumption based on a 3.3V power supply.

State	Power Consumption	Unit
Idle	15	mA
Receiving	50	mA
Transmitting	120	mA
Sleep	10	uA

Table 3 contains the main variables relating to the operations of the protocols studied. These values were previously acquired through other simulations designed to discover the optimal parameters for a network with 5 nodes. A small network was used so that the effects of increasing the number of nodes in the network without adjusting the parameters would be more noticeable.

Table 3: MAC variables

MAC Variable	Value	Unit
T-MAC frame duration	0.7	s
T-MAC timeout interval	0.03	s
B-MAC slot duration	0.17	s
RI-MAC sleep interval	0.85	s

The protocols were evaluated according to the number of delivered packages, their success rate, total energy consumption, energy spent per packet, number of over-heard packets, and their overall adaptability to the increasing network density. Other factors such as the latency of transmissions inside the network were not considered because external variables such as the delay of the communications between the truck's radio and the system that stores the acquired results would overshadow these small aspects in real-world applications.

In order to get representative and meaningful results, each simulation scenario was executed 20 times. The results presented for each performance metric represent the average values calculated from the obtained results. Only the average values are represented in the graphs, as the standard deviations were negligible.

3.2. Results Analysis

Firstly, T-MAC, B-MAC, and RI-MAC were compared in terms of delivery success rate. Figure 5 shows the number of delivered packages for each density. In the base case with only 5 nodes, all protocols have a high success rate, however, their behavior starts to diverge as the number of nodes increases. T-MAC maintains a very high delivery ratio until the number of messages saturates its initial capacity, after which it becomes erratic. Success rates pick up again after 40 nodes because nodes are spending more time awake due to timeout extensions, which increases the network capacity. At the 50 to 55 range the number of delivered packets peaks because the repeated timeout extensions make nodes remain awake all the time, which maximizes the time they have to transmit. However, the protocol has reached its limit after this point and any additional messages only cause degradation of the service due to interference, leading to a drastic decline in capacity.

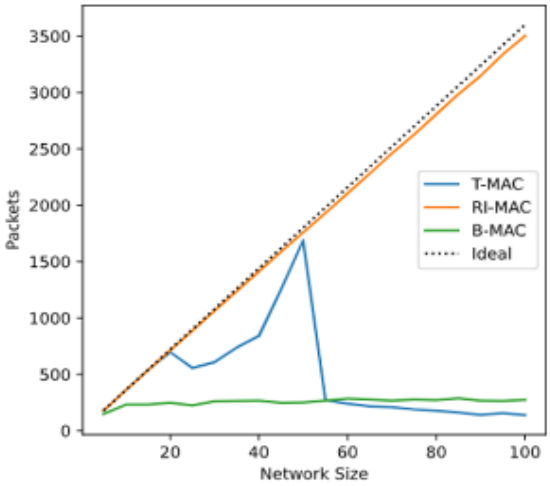


Figure 5: Total packets delivered according to network density.

B-MAC also does not work effectively outside its ideal conditions, the absolute number of delivered packages remained stable throughout the experiment. This is likely because of how the protocol saturates the medium while transmitting a preamble. A centralized node can easily interferer with the communications anywhere else in such a limited space, reducing total throughput. Finally, RI-MAC had the best overall results. Figure 6 shows that it consistently delivered almost all the packets and showed minimal service degradation as the density of the network increased. This is because nodes block the medium for shorter durations and less frequently than the other protocols, leaving room for a greater load in the network.

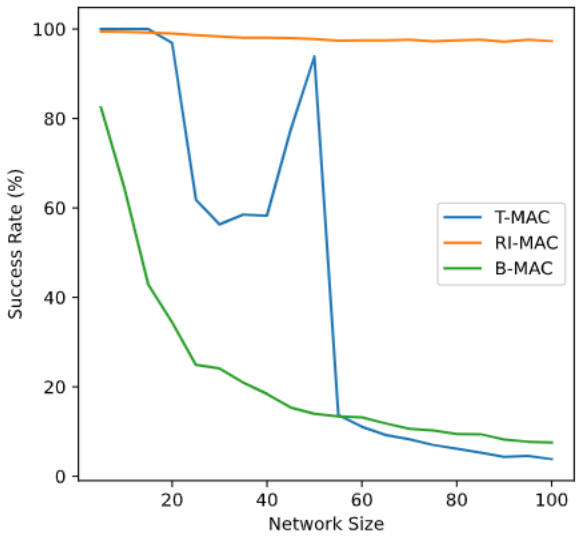


Figure 6: Delivery success rate according to network density.

In terms of power consumption, the most visible difference between the protocols occurs after the network reached 50 nodes. Figure 7 shows that, because of the increasing amount of network events (transmissions), nodes running T-MAC spend an increasing amount of time awake, which leads to more energy consumption. At a certain point, nodes are not able to sleep between wake-up signals. Figure 8 illustrates how this happens. This extra awake time increases the number of transmissions the protocol is capable of handling but also causes a substantial increase in energy consumption.

After 55 nodes the extensions become so frequent that the nodes remain permanently awake, maximizing energy consumption. Because of that, the energy consumed by the network increases linearly with each added node after this point as shown by Figure 9. However, the added consumption does not translate into extra capacity. At 55 nodes, the protocol starts to become overloaded, and each additional node increases the chances of interference, which forces nodes to retransmit data, increasing the chances of interference further. This leads to a feedback loop that severely hinders the protocol's operations.

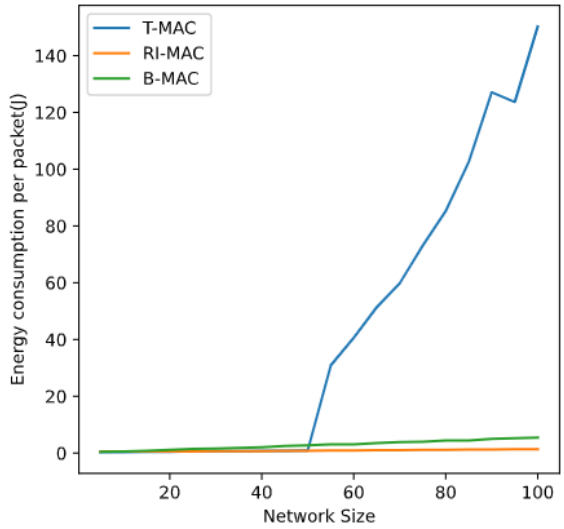


Figure 7: Energy consumption per packet according to network density.

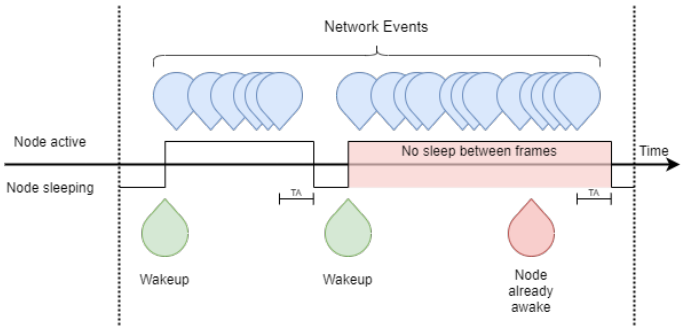


Figure 8: Continuous extension of the T-MAC timeout period leads to increased energy consumption and no sleep between wakeup signals.

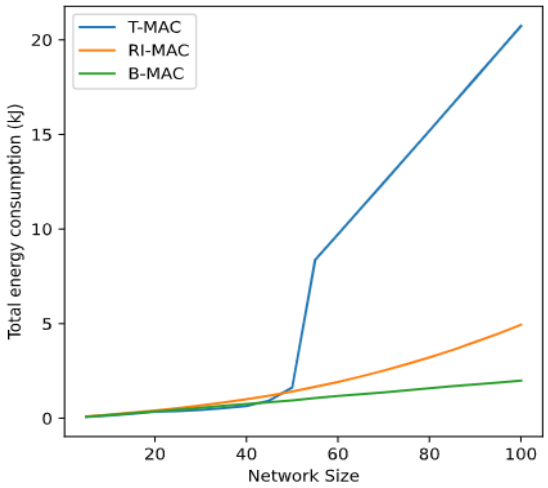


Figure 9: Total network consumption according to network size.

B-MAC and RI-MAC do not suffer from the same problem, their power consumption grows smoothly with the number of nodes in the network. RI-MAC spends more power in absolute terms, but its performance compared to the number of delivered packages is much better. In contrast, B-MAC's added expenditure does not translate into usable network capacity.

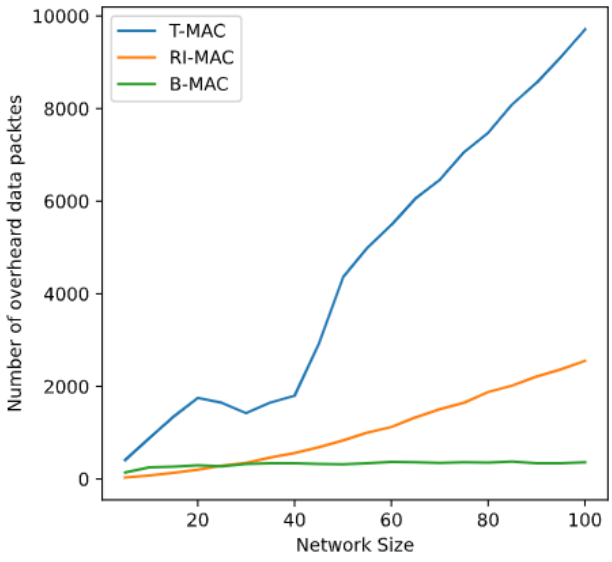


Figure 10: Average number of data packets overheard by nodes depending on network size.

Figure 10 puts the previous results into perspective. The average number of over-heard messages, data packets a node received that were not addressed to it, can also be used to characterize the behavior of a protocol. Ideally, a node would never receive a message addressed to another node to conserve energy. As this statistic did not consider beacon messages, B-MAC comes close to the optimal case. However, this only occurs because the absolute number of packets B-MAC sent was constant. If the protocol was more flexible, a similar phenomenon to what happened with RI-MAC would have been seen. As the number of nodes in the network increase, not only does the same happens to the number of messages, but also to the chances of a node waking up and accidentally receiving a packet sent to another one. This is one of the causes of the increasing power consumption per packet sent seen in Figure 7.

Finally, T-MAC's results are consistent with the increase in total energy consumption. Unlike RI-MAC, it displays a mostly linear increase in overhearing rate because nodes wake up at the same time, meaning that all nodes within the range of a transmission always receive the packet being sent. The gap in the graph is caused by the same issues explored in the analysis of the previous graphs.

4. Conclusion and Future Work

This paper presents the results from a series of simulations designed to study the performance of various MAC protocols in networks with variable node density, especially denser ones. In this instance, the proximity between nodes makes the network behave similarly to a fully connected one. The scenario was set up to model their use inside delivery trucks with the intent of monitoring perishable goods in transit as part of the PrunusPós project. This initiative aims to reduce the losses farmers and distributors of peaches and cherries incur every year due to the fragility of these fruits.

The protocols were evaluated in terms of delivery success rate, energy consumption, overhearing, and flexibility as node densities increase. The results show that RI-MAC, a protocol based on receiver-initiated communications, had the best reliability and lowest consumption per package in a wide range of network densities. Its flexibility is ideal for networks with highly variable density and where continuous adjustment of protocol parameters may be challenging. The growth of power expenditure is also minimal with every node, indicating the networks with more nodes are possible with a limited energy budget.

Furthermore, it is possible to see the various shortcoming of LPL based protocols such as B-MAC and synchronous protocols such as T-MAC. Their behavior is good under the conditions they were optimized for; however, they can quickly lose effectiveness when outside the initial bound. The length of beacon messages and the synchronized wakeup time make them unsuited for extremely dense networks. Future studies should focus on confirming the presented findings in a real testbed to uncover the finer details of RI-MAC's behavior in a network with variable density. Protocols that take into account the number of neighbors a node has may also offer an avenue for research.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This study is within the activities of project "PrunusPós - Optimization of processes for the storage, cold conservation, active and/or intelligent packaging and food quality traceability in post-harvested fruit products", project n. º PDR2020-101-031695, Partnership n.º 87, initiative n.º 175, promoted by PDR 2020 and co-funded by FEADER within Portugal 2020. P.D.G. thanks the support of Fundação para a Ciência e Tecnologia (FCT) and C-MAST - Centre for Mechanical and Aerospace Science and Technologies, under project UIDB/00151/2020. V. N. G. J. S. and J. M. L. P. C. acknowledge that this work is funded by FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/50008/2020. The authors would also like to acknowledge the company InspiringSci, Lda for its interest and valuable contribution to the successful development of this work.

References

[1] K. Lorincz, D.J. Malan, T.R.F. Fulford-Jones, A. Nawoj, A. Clavel, V. Shnayder, G. Mainland, M. Welsh, S. Moulton, "Sensor Networks for Emergency Response: Challenges and Opportunities," IEEE Pervasive Computing, **3**(4), 16–23, 2004, doi:10.1109/MPRV.2004.18.

[2] X. Liu, P. Hu, F. Li, "A street lamp clustered-control system based on wireless sensor and actuator networks," in Proceedings of the 10th World Congress on Intelligent Control and Automation, IEEE: 4484–4489, 2012, doi:10.1109/WCICA.2012.6359237.

[3] R.G. Vieira, A.M. Da Cunha, L.B. Ruiz, A.P. De Camargo, "On the design of a long range WSN for precision irrigation," IEEE Sensors Journal, **18**(2), 773–780, 2018, doi:10.1109/JSEN.2017.2776859.

[4] J.M.L.P. Caldeira, J.J.P.C. Rodrigues, P. Lorenz, "Intra-Mobility Support Solutions for Healthcare Wireless Sensor Networks–Handover Issues," IEEE Sensors Journal, **13**(11), 4339–4348, 2013, doi:10.1109/JSEN.2013.2267729.

[5] G.J. Pottie, W.J. Kaiser, "Wireless integrated network sensors," Communications of the ACM, **43**(5), 51–58, 2000, doi:10.1145/332833.332838.

[6] J. Ma, W. Lou, Y. Wu, X.-Y. Li, G. Chen, "Energy Efficient TDMA Sleep Scheduling in Wireless Sensor Networks," in IEEE INFOCOM 2009 - The 28th Conference on Computer Communications, IEEE: 630–638, 2009, doi:10.1109/INFCOM.2009.5061970.

[7] S. Singh, C.S. Raghavendra, "PAMAS - Power aware multi-access protocol with signalling for Ad Hoc networks," Computer Communication Review, **28**(3), 5–25, 1998, doi:10.1145/293927.293928.

[8] M. Won, T. Park, S.H. Son, "Asym-MAC: A MAC protocol for low-power duty-cycled wireless sensor networks with asymmetric links," IEEE Communications Letters, **18**(5), 809–812, 2014, doi:10.1109/LCOMM.2014.032014.132679.

[9] G. Anastasi, M. Conti, M. Di Francesco, A. Passarella, "Energy conservation in wireless sensor networks: A survey," Ad Hoc Networks, **7**(3), 537–568, 2009, doi:10.1016/j.adhoc.2008.06.003.

[10] S. Hayat, N. Javaid, Z.A. Khan, A. Shareef, A. Mahmood, S.H. Bouk, "Energy efficient MAC protocols," Proceedings of the 14th IEEE International Conference on High Performance Computing and Communications, HPCC-2012 - 9th IEEE International Conference on Embedded Software and Systems, ICESS-2012, 1185–1192, 2012, doi:10.1109/HPCC.2012.174.

[11] Q. Dong, W. Dargie, "A Survey on Mobility and Mobility-Aware MAC Protocols in Wireless Sensor Networks," IEEE Communications Surveys & Tutorials, **15**(1), 88–100, 2013, doi:10.1109/SURV.2012.013012.00051.

[12] F. Jia, Q. Shi, G.M. Zhou, L.F. Mo, "Packet delivery performance in dense wireless sensor networks," 2010 International Conference on Multimedia Technology, ICMT 2010, 12–15, 2010, doi:10.1109/ICMULT.2010.5629537.

[13] R. Cotrim, J.M.L.P. Caldeira, V.N.G.J. Soares, Y. Azzoug, "Power Saving MAC Protocols in Wireless Sensor Networks: A Survey," TELKOMNIKA, 2021.

[14] R.R. Nacional, PrunusPós, 2021.

[15] T. Van Dam, K. Langendoen, "An adaptive energy-efficient MAC protocol for wireless sensor networks," SenSys'03: Proceedings of the First International Conference on Embedded Networked Sensor Systems, 171–180, 2003, doi:10.1145/958491.958512.
[16] J. Polastre, J. Hill, D. Culler, "Versatile low power media access for wireless sensor networks," in Proceedings of the 2nd international conference on Embedded networked sensor systems - SenSys '04, ACM Press, New York, New York, USA: 95, 2004, doi:10.1145/1031495.1031508.
[17] Y. Sun, O. Gurewitz, D.B. Johnson, "RI-MAC: A Receiver-Initiated Asynchronous Duty Cycle MAC Protocol for Dynamic Traffic Loads in Wireless Sensor Networks," in Proceedings of the 6th ACM conference on Embedded network sensor systems - SenSys '08, ACM Press, New York, New York, USA: 1, 2008, doi:10.1145/1460412.1460414.
[18] OpenSim Ltd, OMNeT++ Discrete Event Simulator, 2021.
[19] A.A. Ibrahim, O. Bayat, "Medium Access Control Protocol-based Energy and Quality of Service routing scheme for WBAN," HORA 2020 - 2nd International Congress on Human-Computer Interaction, Optimization and Robotic Applications, Proceedings, 9–14, 2020, doi:10.1109/HORA49412.2020.9152849.
[20] M. Nabi, M. Blagojevic, M. Geilen, T. Basten, T. Hendriks, "MCMAC: An optimized medium access control protocol for mobile clusters in wireless sensor networks," SECON 2010 - 2010 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, 2010, doi:10.1109/SECON.2010.5508200.
[21] Wei Ye, J. Heidemann, D. Estrin, "An energy-efficient MAC protocol for wireless sensor networks," in Proceedings.Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, IEEE: 1567–1576, 2005, doi:10.1109/INFCOM.2002.1019408.
[22] Y. Sun, S. Du, O. Gurewitz, D.B. Johnson, "DW-MAC: A low latency, energy efficient demand-wakeup MAC protocol for wireless sensor networks," in Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing - MobiHoc '08, ACM Press, New York, New York, USA: 53, 2008, doi:10.1145/1374618.1374627.
[23] M. Buettner, G. V Yee, E. Anderson, R. Han, "X-MAC: A Short Preamble MAC Protocol for Duty-Cycled Wireless Sensor Networks," in Proceedings of the 4th international conference on Embedded networked sensor systems - SenSys '06, ACM Press, New York, New York, USA: 307, 2006, doi:10.1145/1182807.1182838.
[24] P. Dutta, S. Dawson-Haggerty, Y. Chen, C.J.M. Liang, A. Terzis, "A-MAC: A versatile and efficient receiver-initiated link layer for low-power wireless," ACM Transactions on Sensor Networks, **8**(4), 1–14, 2012, doi:10.1145/2240116.2240119.
[25] INET Framework.
[26] Y. Tselishchev, A. Boulis, L. Libman, "Experiences and lessons from implementing a wireless sensor network MAC protocol in the Castalia simulator," IEEE Wireless Communications and Networking Conference, WCNC, 2010, doi:10.1109/WCNC.2010.5506096.
[27] T. Boulis, D. Pediaditakis, Castalia.
[28] T. Boulis, D. Pediaditakis, ESP8266EX Datasheet, 31, 2020.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 348-354 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Modelling and Simulation of Fuzzy-based Coordination of Trajectory Planning and Obstacle Avoiding for RRP-Typed SCARA Robots

Ngoc-Anh Mai*

Le Quy Don Technical University, Hoang Quoc Viet str. 236, Hanoi, Vietnam

ARTICLE INFO

Article history:
Received: 14 April, 2021
Accepted: 13 August, 2021
Online: 26 August, 2021

Keywords:
Hierarchical chart
Fuzzy-based coordination
RRP-typed SCARA robot

ABSTRACT

In this article, a fuzzy-based solution of coordination between behaviors of trajectory planning and obstacle avoiding in a RRP-typed SCARA robot control is presented. The first idea of the proposed solution is to divide a robot's complex behavior into simpler parallel behaviors. The second key idea is a fuzzy-based coordination between these behaviors to make smooth robot motions without collision. The modelling and simulation on Matlab are executed to test the performance of the proposed solutions under basic circumstances.

1. Introduction

The Selective Compliance Assembly Robot Arm (SCARA robot) was firstly invented in 1978 [1]. Since then, SCARA robot has been developed for different Degree of Freedom (DoF) such as 3-DoF [2], [3], 4-DoF [4], 5-DoF [5,6], and 6-DoF [7], [8]. In particular, SCARA robot with 3-DoF has become one of the most applied industrial robots due to its simple and basic kinematic structures.

In the basic group of 3-DoF industrial robots, there are many different categories of kinematic structures including articulated manipulators [9], Spherical manipulators [10], SCARA manipulators [11], cylinder manipulators [12], cartesian manipulators [13]. Among these categories, RRP-typed SCARA robots have been strongly invested because of flexible trajectory planning solutions with rapidly exploring random algorithms.

Fuzzy logic is an intelligent control tool that simplifies the complexity of nonlinear control through IF-THEN rules. Thanks to this advantage, many fuzzy logic solutions are proposed for controlling 3-DoF SCARA robot such as trajectory tracking control [2], position control [14], path planning control [15], obstacle avoidance [16], [17]. The Matlab simulation results of [2] show that it is possible to apply fuzzy logic for the RRP-typed SCARA controller to reduce the loop trajectory errors. From the result in [14], it is proven that the designed fuzzy logic controller help the RRP SCARA robot move smoother for tracking trajectory but without obstacle avoidance. Similarly, the proposed design in [15] confirms that fuzzy logic application makes the RRP-typed SCARA robot movement faster than the conventional PD controller in path planning. Furthermore, [16] and [17] proposed fuzzy-based solutions for obstacle avoidance of the robotic manipulators. The simulation results demonstrate that fuzzy-based obstacle avoidance provides SCARA robots better solutions of local planning without any collisions. The computational complexity, however, increases due to the repeated use of the nonlinear functions of the fuzzy logic.

Different to the above-mentioned results, in [18] a fuzzy-based basic solution is proposed for coordinating obstacle avoidance and path planning behaviors of 6-DoF humanoid mobile robots. The main ideas of the solution is to divide a complex robot behavior into simpler behaviors and organize the behaviors in a hierarchical chart so that an upper class behavior consists of some behaviors in the lower class. The main ideas in [18] is reused in this research but for a 3-DoF robot to reduce the computational complexity of the robot control system. The proposed fuzzy-based coordination between parallel behaviors makes the robot movement smooth according to the changing obstacle distance. The article is presented as follows: First, the kinematic structure of RRP-typed SCARA robot is introduced. Then, a hierarchical chart of behaviors of trajectory planning and obstacle avoidance are analyzed. After that, a modular diagram of SCARA robot control system is given with the separable modules for trajectory planning and collision avoidance and fuzzy-based coordination. Finally, the Matlab simulations are analyzed to test the system performance.

*Corresponding Author: Ngoc-Anh Mai, Email: maingocanh.atc@mta.edu.vn

2. Kinematics

2.1. Kinematic structure

The structure of RRP-typed SCARA robot with 3-DoF includes 2 revolute (R) (or rotating joints) and one prismatic (P) (or translating joint) as shown on the left side in Fig.1.

These joints are operated by 3 independent actuators to control the pose (including position and direction) of the End-Effector (E) following a desired trajectory.

The basic structure of RRT-typed SCARA robot consists of 4 parts: Base B0, revolute joint R1, revolute joint R2 and prismatic joint P3 as shown on the right side in Fig. 1.

The kinematic parameters in Fig. 2 are defined as follows:

+ d_0 is the height from base B0 to joint R1 along Z_0 axis;

+ a_1 and a_2 are the lengths of the arms 1 and 2 along axes X_1 and X_2, respectively;

+ θ_1 and θ_2 are the rotation angles of joints R1 and R2 around axes Z_1 and Z_2, respectively;

+ d_3 is the distance from joint P3 to the end-effector E.

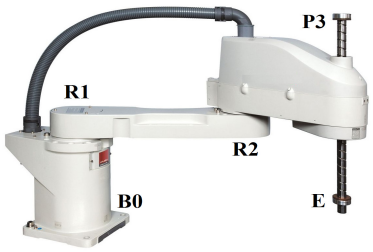


Figure 1: Basic structure of RRT-typed SCARA robot

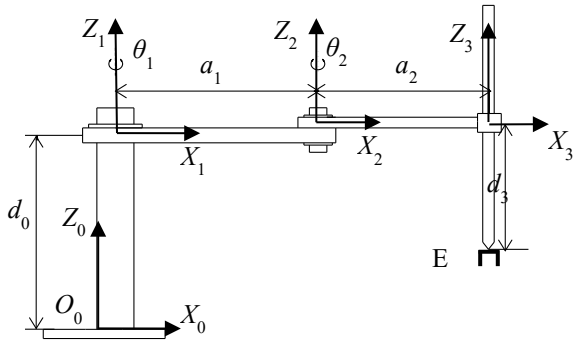


Figure 2: Kinematic structure of the SCARA robot

2.2. Homogeneous transformation

According to the Denavit-Hartenberg (D-H) of homogeneous transformation rules in [19], the D-H parameters of the SCARA robot can be setup in the table shown in Table 1. Using the D-H parameter table, the transformation matrices are as follows:

$$\mathbf{T}_1^0 = \begin{bmatrix} c_1 & -s_1 & 0 & 0 \\ s_1 & c_1 & 0 & 0 \\ 0 & 0 & 1 & d_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$\mathbf{T}_2^1 = \begin{bmatrix} c_2 & -s_2 & 0 & a_1 \\ s_2 & c_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

$$\mathbf{T}_3^2 = \begin{bmatrix} 1 & 0 & 0 & a_2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -d_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

where:

- $\mathbf{T}_1^0$, $\mathbf{T}_2^1$, $\mathbf{T}_3^2$ are respectively matrices of homogeneous transformation between the coordinate systems from $OXYZ_0$ to $OXYZ_1$, from $OXYZ_1$ to $OXYZ_2$, and from $OXYZ_2$ to $OXYZ_3$.

- c_1, s_1, c_2, s_2 are respectively the symbols of $\cos(\theta_1)$, $sin(\theta_1)$, $\cos(\theta_2)$, $sin(\theta_2)$.

To calculate the homogeneous transformation matrix from the origin O_0 to the end-effector E, we perform the matrix multiplication as follows:

$$\mathbf{T}_3^0 = \mathbf{T}_1^0\mathbf{T}_2^1\mathbf{T}_3^2 = \begin{bmatrix} \mathbf{R}_3^0 & \mathbf{p}_3^0 \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (4)$$

The detail computation of equation (4) is carried out as follows:

$$\mathbf{T}_1^0\mathbf{T}_2^1 = \begin{bmatrix} c_1 & -s_1 & 0 & 0 \\ s_1 & c_1 & 0 & 0 \\ 0 & 0 & 1 & d_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} c_2 & -s_2 & 0 & a_1 \\ s_2 & c_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \mathbf{T}_2^0 \quad (5)$$

$$\mathbf{T}_2^0 = \begin{bmatrix} c_{12} & -s_{12} & 0 & a_1 \cdot c_1 \\ s_{12} & c_{12} & 0 & a_1 \cdot s_1 \\ 0 & 0 & 1 & d_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6)\mathbf{T}_2^0\mathbf{T}_3^2 =$$

$$\begin{bmatrix} c_{12} & -s_{12} & 0 & a_1 \cdot c_1 \\ s_{12} & c_{12} & 0 & a_1 \cdot s_1 \\ 0 & 0 & 1 & d_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & a_2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -d_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7)$$

$$\mathbf{T}_3^0 = \begin{bmatrix} c_{12} & -s_{12} & 0 & a_1 \cdot c_1 + a_2 \cdot c_{12} \\ s_{12} & c_{12} & 0 & a_1 \cdot s_1 + a_2 \cdot s_{12} \\ 0 & 0 & 1 & d_0 - d_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (8)$$

$$\mathbf{R}_3^0 = \begin{bmatrix} c_{12} & -s_{12} & 0 \\ s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (9)$$

$$\mathbf{p}_3^0 = \begin{bmatrix} x_E \\ y_E \\ z_E \end{bmatrix} = \begin{bmatrix} a_1 \cdot c_1 + a_2 \cdot c_{12} \\ a_1 \cdot s_1 + a_2 \cdot s_{12} \\ d_0 - d_3 \end{bmatrix} \quad (10)$$

Table 1: Denavit-Hartenberg parameters of the SCARA robot

Joint	θ_i	α_i	a_i	d_i
1	θ_1^*	0	0	d_0
2	θ_2^*	0	a_1	0
3	0	0	a_2	$-d_3^*$
* means variable				

where

- $\mathbf{T}_3^0, \mathbf{R}_3^0, \mathbf{p}_3^0$are transformation matrix, orientation matrix, and position matrix of the end-effector E in comparison with base B0.

- c_{12}, s_{12} notes for $\cos(\theta_1 + \theta_2), sin(\theta_1 + \theta_2)$, respectively.

- x_E, y_E, z_E are respectively the coordinates of the end-effector E projected on the axes X, Y, Z of $OXYZ_0$.

2.3. Inverse kinematic computation

According to [20], the inverse kinematics problem consists of the determination of the joint variables corresponding to a given end-effector position and orientation. The solution to this problem is of fundamental importance in order to transform the motion specifications, assigned to the end-effector in the operational space, in to the corresponding joint space motions that allow an execution of the desired motion.

In this study, the inverse kinematics problem requires to find out the 2 joint variables concerning angles θ_1, θ_2 and the displacement d_3 based on the given pose of the end-effector E, that means $\mathbf{p}_3^0$ and $\mathbf{R}_3^0$ are known.

To solve this problem, the top-viewed projections of the SCARA robot, as shown in Figure 3, allows formula (10) to mathematically express in a different way as follows:

$$x_E^2 + y_E^2 = (a_1 \cdot c_1 + a_2 \cdot c_{12})^2 + (a_1 \cdot s_1 + a_2 \cdot s_{12})^2$$
$$= a_1^2 + a_2^2 + 2 \times a_1 \times a_2 \times (c_1 \times c_{12} + s_1 \times s_{12}) x_E^2 +$$
$$y_E^2 = a_1^2 + a_2^2 + 2 \cdot a_1 \cdot a_2 \cdot c_2 \quad (11)$$

From formula (11) we have:

$$c_2 = \frac{x_E^2 + y_E^2 - a_1^2 - a_2^2}{2 \cdot a_1 \cdot a_2} \quad (12)$$

$$\theta_2 = \operatorname{acos}\left(\frac{x_E^2 + y_E^2 - a_1^2 - a_2^2}{2 \cdot a_1 \cdot a_2}\right) \quad (13)$$

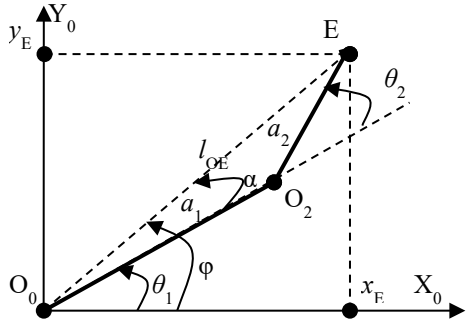


Figure 3: Top-viewed Projections on plane $OXYZ_0$

To calculate angle θ_1, the geometric calculation method can be used under the projections on the plane $OXYZ_0$ as follows:

Let consider triangle O_0O_2E and use mathematic computations:

+ The length from O_0 to E is calculated:

$$l_{OE} = \sqrt{(x_E^2 + y_E^2)} \quad (13)$$

+ The angle α between O_0O_2 and O_0E is calculated as follows:

$$\alpha = \operatorname{acos}\left(\frac{({a_1}^2 + l_{OE}^2 - {a_2}^2)}{2 \cdot a_1 \cdot l_{OE}}\right). \quad (14)$$

+ Angle θ_1 for rotating link 1 around axis Z1 is calculated:

$$\theta_1 = \varphi - \alpha = \operatorname{atan}\left(\frac{y_E}{x_E}\right) - \operatorname{acos}\left(\frac{({a_1}^2 + l^2 - {a_2}^2)}{2 \cdot a_1 \cdot l}\right). \quad (15)$$

Using the bottom expression in formula (10) we get:

$$d_3 = d_0 - z_E. \quad (16)$$

Let draw from equations (15), (13) and (16) we have:

$$\begin{bmatrix}\theta_1 \\ \theta_2 \\ d_3\end{bmatrix} = \begin{bmatrix}\operatorname{atan}\left(\frac{y_E}{x_E}\right) - \operatorname{acos}\left(\frac{({a_1}^2 + l^2 - {a_2}^2)}{2 \cdot a_1 \cdot l}\right) \\ \operatorname{acos}\left(\frac{x_E^2 + y_E^2 - a_1^2 - a_2^2}{2 \cdot a_1 \cdot a_2}\right) \\ d_0 - z_E\end{bmatrix} \quad (17)$$

In summary, based on formula (17) it is possible to simulate RRT-typed SCARA robot movement by giving the coordinates of waypoints (or path points) to form a desired trajectory.

In the next section, the calculation based on these waypoints for planning trajectory is presented. After that, the calculation for avoiding obstacle is analyzed. Based on these calculations, both kinds of behaviors concerning planning trajectory and avoiding obstacle are organized in a hierarchical chart before applying fuzzy logic for coordinating them.

3. Planning trajectory and avoiding obstacle

3.1. Planning trajectory

According to [19], to reduce the complexity of motion control, a complex path can be replaced with a sequence of *n* waypoints to ensure motion along the desired trajectory$[p_1..p_i, p_j..p_n]$.

To travel over all *n* waypoints at certain instants of time, it should be better to compute the actuating commands $\mathbf{v}_T = (v_T, \omega_T)$ for tracking each segment of the desired trajectory between two adjacent waypoints (p_i, p_j) by assigning the initial and final velocities.

Let briefly depict the computation by interpolating polynomials in a segment of trajectory between two adjacent waypoints (p_i, p_j) at two instants of time t_i and t_j as follows:

It is noticed that, if the order of interpolating polynomials increase, the nature of the desired trajectories is reduced and the numerical accuracy for computation polynomial coefficients decreases. For this reason, cubic polynomials are chosen in the following computation.

The generic equation of interpolating polynomials with velocity constraints at the two waypoints are:

$$s(t) = k_0 + k_1 \cdot t + k_2 \cdot t^2 + k_3 \cdot t^3. \quad (18)$$

$$v(t) = \dot{s}(t) = k_1 + 2k_2 \cdot t + 3k_3 \cdot t^2. \quad (19)$$

where

+ $s(t)$ is a cubic polynomial of the path;

+ $v(t)$ is a cubic polynomial velocity respected to $s(t)$;

+ k_i, i = 0...3, are polynomial coefficients depended on the arbitrary motion parameters at the two considering waypoints.

To solve the polynomials, the user has to specify the desired velocities at each points as follows:

+ at the initial time t_i at p_i: the position value is given $s(t_i) = \theta_i$; the velocity value is given $\dot{s}(t_i) = \dot{\theta}_i$;

+ at the final time t_j at p_j: the position value is given $s(t_j) = \theta_j$; the velocity value is given $\dot{s}(t_j) = \dot{\theta}_j$.

Equations (18) and (19) are defined as follows:

$$\theta_i = k_0 + k_1 \cdot t_i + k_2 \cdot {t_i}^2 + k_3 \cdot {t_i}^3. \quad (20)$$

$$\dot{\theta}_i = k_1 + 2 \cdot k_2 \cdot t_i + 3 \cdot k_3 \cdot {t_i}^2. \quad (21)$$

$$\theta_j = k_0 + k_1 \cdot t_j + k_2 \cdot t_j^2 + k_3 \cdot t_j^3. \quad (22)$$

$$\dot{\theta}_j = k_1 + 2 \cdot k_2 \cdot t_j + 3 \cdot k_3 \cdot t_j^2. \quad (23)$$

Expressing equations from (20) to (23) in a matrix form, we have the following matrix:

$$\begin{bmatrix} \theta_i \\ \dot{\theta}_i \\ \theta_j \\ \dot{\theta}_j \end{bmatrix} = \begin{bmatrix} 1 & t_i & t_i^2 & t_i^3 \\ 0 & 1 & 2 \cdot t_i & 3 \cdot t_i^2 \\ 1 & t_j & t_j^2 & t_j^3 \\ 0 & 1 & 2 \cdot t_j & 3 \cdot t_j^2 \end{bmatrix} \cdot \begin{bmatrix} k_0 \\ k_1 \\ k_2 \\ k_3 \end{bmatrix}. \quad (24)$$

To determine the polynomial coefficients k_i, equation (24) can be changed into the following matrix:

$$\begin{bmatrix} k_0 \\ k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} 1 & t_i & t_i^2 & t_i^3 \\ 0 & 1 & 2 \cdot t_i & 3 \cdot t_i^2 \\ 1 & t_j & t_j^2 & t_j^3 \\ 0 & 1 & 2 \cdot t_j & 3 \cdot t_j^2 \end{bmatrix}^{-1} \begin{bmatrix} \theta_i \\ \dot{\theta}_i \\ \theta_j \\ \dot{\theta}_j \end{bmatrix} =\cdot \begin{bmatrix} k_0 \\ k_1 \\ k_2 \\ k_3 \end{bmatrix}. \quad (25)$$

Equation (25) enables the control system planning the desired trajectory independently in pairs of waypoints. The computed velocity command at the waypoints has the following form:

$$\mathbf{v}_T = (v_T, \omega_T). \quad (26)$$

where

- $v_T$is linear velocity for tracking the planned trajectory.

- ω_T is angular velocity for tracking the planned trajectory.

The detail calculation of the velocities at the given waypoints can be seen in [20].

3.2. Avoiding obstacle

In the scope of educational goal, let consider an obstacle with a cylinder shape shown in Figure 4. The obstacle stays in segment of path assuming (p_i, p_j). In this situation, the obstacle have to be avoided by following an additional path to make a suitable curvature in segment (p_i, p_j).

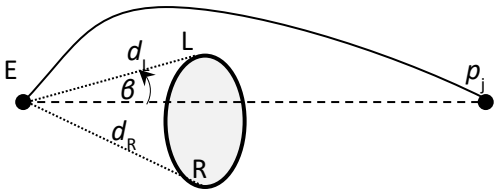


Figure 4: Additional path for avoiding obstacle

Let note the distance to the left obstacle edge L be d_L, the distance to the right edge R be d_R. and the angle deviation β from the direction to p_j to the nearest obstacle edge.

The actuating command for avoiding obstacle are follows:

$$\mathbf{v}_A = (v_A, \omega_A). \quad (27)$$

$$v_A = f_{vA}(min(d_L, d_R) . cos(\beta)). \quad (28)$$

$$\omega_A = f_{\omega A}\left(\frac{sin(\beta)}{min(d_L, d_R)}\right).\text{sign}(d_L - d_R). \quad (29)$$

where

- v_A and ω_A are linear velocity and angular velocity for avoiding obstacle, respectively.

- f_{vA} and $f_{\omega A}$ are user-defined non-linear functions related to d_L, d_R and β for calculating the linear velocity and angular velocity, respectively.

4. Modelling the system with fuzzy-based coordination

Let call the behaviors be planning trajectory PT and avoiding obstacle AO. Then, the AO behavior is divided into 2 simpler behaviors Turning-Left (TL) and Turning-Right (TR) to control the end-effector avoiding a collision by moving to the left or the right edge, respectively. These behaviors are organized in a hierarchical chart shown in Figure 5 and the robot control system can be designed in modules as the block diagram shown in Figure6.

Module PT defines a behavior for planning trajectory and provides a motion command $\mathbf{v}_T$ to module FC.

Module AO selects a suitable behavior between behavior TL and TR based on the results of obstacle detection and measurement in module DM. Then, an actuating command $\mathbf{v}_A$ for avoiding obstacle is computed and sent to module FC.

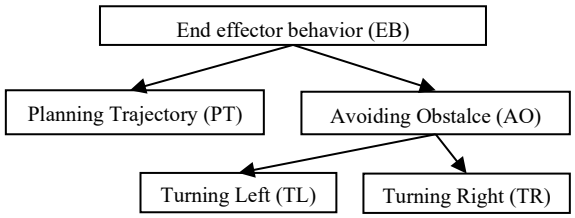


Figure 5: Hierarchical chart for the end-effector behavior

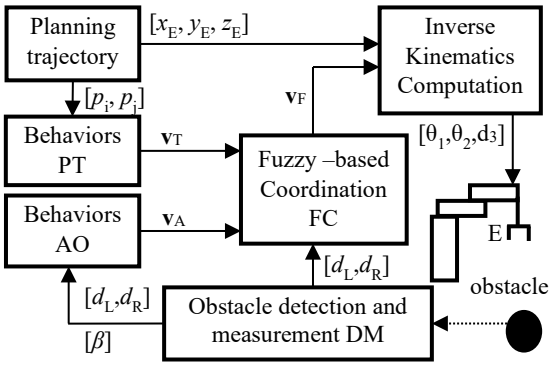


Figure 6: Block diagram of the control system with fuzzy-based coordination

Module FC carries out a fuzzy-based coordination between two behaviors $\mathbf{v}_T$ and $\mathbf{v}_A$ to give a suitable actuating command so that the end-effector can avoid the obstacle and reach to path point p_j.

The fuzzy-based behavioral coordination mechanism is computed as follows:

$$\mathbf{v}_E = (v_E, \omega_E). \tag{30}$$

$$\begin{cases} v_E = k_T \cdot v_T + k_c \cdot v_A \\ \omega_E = k_T \cdot \omega_T + k_c \cdot y_A \end{cases}. \tag{31}$$

where k_T and k_C are weights of tracking trajectory and avoiding collision, which are determined by the fuzzy rules as in Table 2.

Table 2: Fuzzy-based estimation of k_T and k_C

Fuzzy-based values of k_T and k_C		β		
		small	*medium*	*big*
min (d_L,d_R)	*far*	k_T big, k_C big	k_T big, k_C medium	k_T big, k_C small
	middle	k_T medium, k_C big	k_T medium, k_C medium	k_T medium, k_C small
	near	k_T small, k_C big	k_T small, k_C medium	k_T small, k_C small

In the next section, the simulations on Matlab are executed under the guide in [21] and [22] to test the control system with the fuzzy-based coordination.

5. Simulation

5.1. GUI interface of simulation

The robot system is tested by a simulation on Matlab. The robot's motions concerning trajectory and obstacle are tracked and displayed on the GUI interface shown in Figure7.

The left area on the GUI interface displays the coordinates of the three joints concerning the motion and obstacle parameters relative to position and dimensions of height h, long diameter r_1 and short diameter r_2. The middle area of the GUI interface is a 3D illustration of a truth-based trajectory. The right area is a 2D view of the robot's workspace and control buttons.

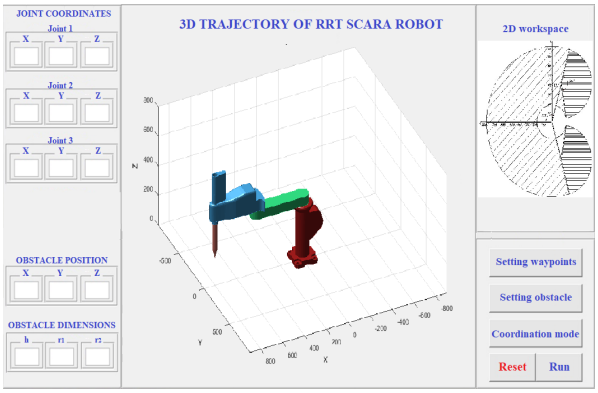


Figure 7: GUI interface of the simulation on Matlab

The five control functions are:

+ "Setting waypoints" button is used for setting the coordinates of each waypoint.
+ "Setting obstacle" button is used for setting the position and dimensions of the obstacle.
+ "Coordination mode" button is used for selecting a coordination mode including conventional coordination without fuzzy logic rules and fuzzy-based coordination.
+ "Reset" button is used for resetting all parameters.
+ "Run" button is used for starting the program.

5.2. Simulation goals and results

Two simulation goals are to test the conventional coordination without fuzzy rules to avoid obstacles and to test the capability of obstacle avoidance under the fuzzy-based coordination.

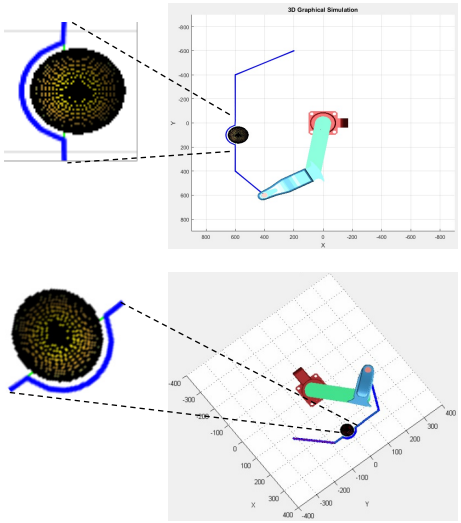


Figure 8: Avoiding obstacle without fuzzy-based coordination

The first simulation results are shown in Figure 8. During following the given waypoints, the robot avoids the obstacle without fuzzy-based coordination. The second simulation results are shown in Figure 9. During following the given waypoints, the robot avoids the obstacle without fuzzy-based coordination.

It is noticed that, the recoded trajectories in Figure 8 are not smooth at the points for changing from a behavior of trajectory planning to a behavior of obstacle avoiding. Otherwise, the recoded ones in Figure 9 become smooth at the points for changing

from a behavior of trajectory planning to a behavior of obstacle avoiding due to the fuzzy-based coordination.

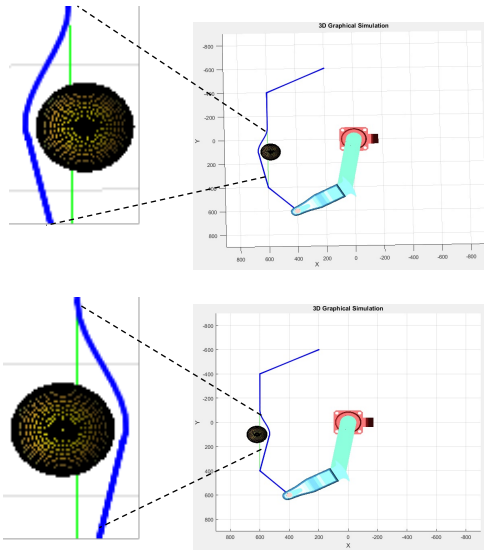


Figure 9: Avoiding obstacle with fuzzy-based coordination

The simulation results in Figure 9 prove that the fuzzy-based coordination make the robot motions non-stop before avoiding obstacle. That means the singularity problem of robotic manipulators can be avoided.

6. Conclusion

The proposed fuzzy-based coordination help the robot safely move and avoid the singularity problem of robotic manipulators.

Matlab simulations are performed to test the performance of the proposed system. Simulation results demonstrate that fuzzy-based coordination helps the robot move more smoothly to eliminate singularity that may appear at dead-points.

In further research, the proposed solution will be tested on a real robot to evaluate the accuracy of trajectory planning after applying the fuzzy-based coordination.

Conflict of Interest

The author declares that there are no conflicts of interest.

References

[1] K. Yamafuji, "Development of SCARA robots," Journal of Robotics and Mechatronics, **31**(1), 10–15, 2019, doi:10.20965/jrm.2019.p0010.

[2] S.M. Raafat, S.M. Mahdi, "Improved Trajectory Tracking Control for a Three Axis SCARA Robot Using Fuzzy Logic," In IJCCCE, **16**(January), 11–19, 2016.

[3] I.S. Karem, T.A.J. A.Wahabt, M.J. Yahyh, "Design and Implementation for 3-DoF SCARA Robot based PLC," Al-Khwarizmi Engineering Journal, **13**(2), 40–50, 2017, doi:10.22153/kej.2017.01.002.

[4] U.T. Nasional, C. Engineering, "Design and development of 4-DoF SCARA robot for educational purposes," Teknologi, **54**(1), 193–215, 2011, doi:10.11113/jt.v54.810.

[5] F. Cao, J. Liu, C. Zhou, Y. Zhao, Z. Fu, W. Yan, "A Novel 5-DOF Welding Robot Based on SCARA," in in 10th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2016–2019, 2016, doi:10.1109/ICIEA.2015.7334444.

[6] M. Abu Qassem, I. Abuhadrous, H. Elaydi, "Simulation and Interfacing of 5 DOF Educational Robot Arm," in Proceedings - 2nd IEEE International Conference on Advanced Computer Control, ICACC 2010, 569–574, 2010, doi:10.1109/ICACC.2010.5487136.

[7] N.A. Mai, X.B. Duong, "Algorithm for improving feeding rates of industrial welding robot TA 1400 in combination with a turntable frame," Computer Science and Cybernetics, **3**(3), 285–294, 2020, doi:10.15625/1813-9663/36/3/14968.

[8] N.A. Mai, X.B. Duong, "Voice Recognition and Inverse Kinematics Control for a Redundant Manipulator Based on a Multilayer Artificial," Hindawi Robotics, **2021**, 1–10, 2021, doi:10.1155/2021/5805232.

[9] E.C. Agbaraji, H.C. Inyiama, C.C. Okezie, "Dynamic Modeling of a 3-DOF Articulated Robotic Manipulator Based on Independent Joint Scheme," Physical Science International Journal, **15**(1), 1–10, 2017, doi:10.9734/PSIJ/2017/33578.

[10] T. Taunyazov, M. Rubagotti, A. Shintemirov, "Constrained Orientation Control of a Spherical Parallel Manipulator via Online Convex Optimization," IEEE/ASME Transactions on Mechatronics, **23**(1), 252–261, 2018.

[11] K. Wei, B. Ren, "A Method on Dynamic Path Planning for Robotic Manipulator Autonomous Obstacle Avoidance Based on an Improved RRT Algorithm," MDPI Special Issue Smart Sensors for Mechatronic and Robotic Systems, **18**(2), 571–585, 2018, doi:10.3390/s18020571.

[12] A. Prasad, B. Sharma, J. Vanualailai, "Motion Control of a Pair of Cylindrical Manipulators in a Constrained 3-dimensional Workspace," in 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), 75–81, 2017, doi:10.1109/APWConCSE.2017.00022.

[13] A. Suarez, M. Perez, G. Heredia, "Cartesian Aerial Manipulator with Compliant Arm," MDPI Special Issue Aerial Robotics for Inspection and Maintenance, **11**(3), 1001–1020, 2021.

[14] M.Z.A.- Faiz, A.M. Abbass, "Design and Implementation of Fuzzy Logic Controller for IVAX SCARA Robot Using Real Time Window Target," JCSET, **3**(10), 373–381, 2013.

[15] D. Prabu;, S. Kumar;, R. Prasad, "Dynamic Control of Three-Link SCARA Manipulator using Adaptive Neuro Fuzzy Inference System," in 2008 IEEE International Conference on Networking, Sensing and Control, 1609–1614, 2008.

[16] P.G.. Zavlangas, S.G. Tzafestas, "Industrial Robot Navigation and Obstacle Avoidance Employing Fuzzy Logic," Intelligent & Robotic Systems, **27**(1), 85–97, 2000, doi:10.1023/A:1008150113712.

[17] Y. Chen, Y. Wang, "Obstacle avoidance path planning strategy for robot arm based on fuzzy logic," in 12th International Conference on Control Automation Robotics & Vision (ICARCV), 5–7, 2012, doi:10.1109/ICARCV.2012.6485438.

[18] H.C. Nguyen, H.X. Nguyen, N.A. Mai, L.B. Dang, H.M. Pham, "A Modular Design Process for Developing Humanoid Mobile Robot Viebot," Advances in Science, Technology and Engineering Systems Journal (ASTES), **3**(4), 230–235, 2018.

[19] R.M. Murray, A Mathematical Introduction to Robotic Manipulation, CRC press, 1994.

[20] L. Sciavicco, B. Siciliano, Modelling and Control of Robot Manipulators, Springer, 2000.

[21] S. Shrivastava, "Matlab guide for forward kinematic calculation of 3 to 6 DoF SCARA robots," IJRET, **6**(8), 46–52, 2017, doi:10.15623/ijret.2017.0609009.

[22] M.F. Aly, A.T. Abbas, "Simulation of obstacles ' effect on industrial robots ' working space using genetic algorithm," King Saud University – Engineering Sciences, **26**, 132–143, 2014, doi:10.1016/j.jksues.2012.12.005.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 355-367 (2021)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Enhance Student Learning Experience in Cybersecurity Education by Designing Hands-on Labs on Stepping-stone Intrusion Detection

Jianhua Yang[1], Lixin Wang[*,1], Yien Wang[2]

[1]*TSYS School of Computer Science, Columbus State University, Columbus, GA 31907, USA*

[2]*College of Engineering, Auburn University, Auburn, AL 30060, USA*

ARTICLE INFO

Article history:
Received: 07 June, 2021
Accepted: 09 August, 2021
Online: 26 August, 2021

Keywords:
Stepping-stone
Intrusion Detection
Cybersecurity Curriculum
Ethical Hacking
Hands-on Experience

ABSTRACT

Stepping-stone intrusion has been widely used by professional hackers to launch their attacks. Unfortunately, this important and typical offensive skill has not been taught in most colleges and universities. In this paper, after surveying the most popular detection techniques in stepping-stone intrusion, we develop 10 hands-on labs to enhance student-learning experience in cybersecurity education. The goal is not only to teach students offensive skills and the techniques to detect and prevent stepping-stone intrusion, but also to train them to be successfully adaptive to the fast-changing dynamic cybersecurity world.

1. Introduction

1.1. Cybersecurity Significance

We live in a world where digital technologies are needed for various daily activities. The Internet has revolutionized data communications and significantly changed our daily lives. However, hackers can now easily launch cyberattacks using the Internet. As cyberattacks continue to grow, it is important to secure our critical infrastructures, organizations, business and networks.

1.2. The Importance of Stepping-stone Intrusion Detection

Intrusion techniques are widely used by intruders to invade a computing system. Intrusion detection systems (IDS) are installed on a lot of computer and network systems. Intruders tend to use several compromised hosts, called stepping-stones, to send attacking commands to a remote target host, in order to avoid being detected. Attacks that are launched through a chain of stepping-stone host are called stepping-stone intrusion. With a stepping-stone attack, intruders remotely login to such stepping-stones using tools such as SSH, rlogin, or telnet, and then send the attacking packets to the remote target host.

In this paper, after the survey of many known detection techniques for the stepping-stone intrusion, we propose ten hands-on labs which are developed based on the cutting-edge techniques in stepping-stone intrusion detection. The goal is to help students to learn the techniques of stepping-stone intrusion detection. We aim at educating learners to be qualified professionals in cybersecurity in order to defend various digital data and resources. It is also expected to enhance students' learning in cybersecurity education by conducting the hands-on labs designed.

2. Key Challenges

Before designing the hands-on labs on stepping-stone intrusion and its detection, we discuss how challenge the known detection approaches for stepping-stone intrusion are integrated into cybersecurity curricula. In order to educate learners to be qualified professionals in cybersecurity, it is necessary to teach offensive skills in college cybersecurity major curriculum.

Integrating stepping-stone intrusion and its detection techniques into cybersecurity curriculum can make us move forward a big step to achieve this goal. Although a great number of detection approaches for stepping-stone intrusion have been proposed since the emerging of the Internet, there are still a lot of challenges to integrate these detection approaches into cybersecurity curricula at the college level. The first challenge is why we need to teach college students ethical hacking skills. Would it be possible educate our students to become a hacker against us, not for us? The second challenge is that, since there are

[*]Corresponding Author: Lixin Wang, 4225 University Ave., Columbus, GA 31907, USA. Contact No: 001-706-507-8190. Wang_Lixin@ColumbusState.edu

too many algorithms for stepping-stone intrusion detection proposed in the literature, which approaches among them are suitable to our college students as learning materials? The third challenge is what hands-on labs can be developed and integrated into cybersecurity curriculum. We all know that the difficulty in teaching cybersecurity is not at the delivery of the theory and techniques; it is at the development of hands-on labs for students to practice hacking and defensive skills. Considering the limited budget in each four-year college, the cost is an important factor when designing these hands-on labs. However, we still want to motivate our students to learn cybersecurity skills via hands-on learning experience.

2.1. The Rationale to Teach Ethical Hacking Skills

Should we teach ethical hacking skills to cybersecurity major students? To the best of our knowledge, even though some four-year institutions have included ethical hacking skills as part of their cybersecurity curriculum, there are still some concerns and doubts from students' parents and local communities about the possibility that teaching ethical hacking skills would make their kids to conduct some malicious activities, and commit crimes. We must convince students' parents as well as the local communities with the following advice: 1) the word 'hacker' has long been understood negatively. Hacking actually involves computing skills to find vulnerabilities of a system, penetrate a system, and be able to remove evidence of accessing to a system [1]. Similar to the case that doctors who might criminally abuse their medical skills to hurt humans, a hacker who knows some special offensive hacking skills might also misuse their techniques. However, we should not define the term hacking by its misuse; 2) cybersecurity is a two-edged sword: offensive and defensive. To be effective at defence, students must fully understand the capabilities of hackers and the way how hackers perform cyberattacks; 3) it is widely believed that including both perspectives of "defender" and "attacker" and the related skills could make the cybersecurity curriculum more meaningful and practical [2]. On the other hand, teaching hacking skills can make cybersecurity professionals be equipped with offensive techniques, and well prepared to defend their computing and network system; 4) regardless of teaching hacking skills or not, hackers were out there, and will still be out there. Should hacking skills be integrated into cybersecurity curricula, it would be possible to promote conscious ethical practices and minimize the likelihood that students would misuse the skills.

2.2. Challenging to Integrate the Techniques to a 3-Credit Hours Course

What techniques should be selected to train our students with cybersecurity skills, as there are tons of approaches that have been proposed to detect stepping-stone intrusion since 1995? In a regular course with 48 academic credit hours, it is infeasible to cover all the techniques developed so far, but we do want to train our students not only to have an overall picture of the techniques on stepping-stone intrusion detection, but also to deeply understand some specific and typical intrusion detection approaches. The challenge is to develop contents modules and design hands-on lab exercises. In this paper, we only focus on the designing the hands-on labs on stepping-stone intrusion and its detection. Refer to our prior work [3] for the course modules we developed for integration of detection techniques for stepping-stone intrusion into cybersecurity curricula.

2.3. Challenge on Developing Hands-on Labs of Stepping-stone Intrusion and its Detection

The most difficult part of teaching cybersecurity courses is to design appropriate hands-on labs. We all know the importance of hands-on labs in cybersecurity education. Without the practicing of the techniques covered in cybersecurity class, it is hard to make our students to digest the cybersecurity skills. Conducting cybersecurity hands-on labs needs hardware and software that are more likely not free. Most colleges are equipped with good hardware, such as computers, routers, switches, and different type of servers, but lack of appropriate software. One reason is that some software helping students to practice cybersecurity skills are usually not free, and may be extremely expensive, such as Cyber-range, its price can be as high as more than one million dollars. Therefore, the challenge is how to design appropriate hands-on labs not only can help students to practice stepping-stone intrusion and its detection techniques, but also can reduce the cost to make labs affordable to most colleges.

3. Survey of the Techniques on Stepping-stone Intrusion and its Detection

Many methods have been proposed to detect stepping-stone intrusion. In [4], the authors proposed a thumbprint method to detect stepping-stone intrusion in 1995. This method was developed to compare the contents of TCP/IP packets from the incoming and outgoing sessions of a computer that is chosen to be the sensor for detection. In [5], the authors proposed a detection approach for stepping-stone intrusion by considering the time gaps between the packets captured from the outgoing connection and the incoming connection from a host. In [6], the authors proposed another method for stepping-stone intrusion detection. Their method did not follow the idea of using time-based thumbprints. Instead, the authors in [6] used the deviation between the incoming and outgoing sessions of a computer.

After 2000, a lot more methods were proposed for stepping-stone intrusion detection. One popular approach is to compare the number of packets from the incoming connection with that from the outgoing connection. For the details of this type of approach, please refer to the references [7-9]. A watermark correlation technique was proposed for stepping-stone intrusion detection [10-12]. The idea of using a watermark in stepping-stone intrusion detection is to insert a watermark in the incoming connection of a detection sensor, and then pay attention to the outgoing connections to see if the same watermark can be found in any of these outgoing connections. The rationale used in the papers [10-12] is to analyse and compare the incoming and outgoing connections of a sensor to see if there is any relayed pair. A sensor is defined as a computer host in which all the packets are captured and a detection program runs. If an incoming connection of a sensor is relayed with an outgoing connection, the sensor is considered as a stepping-stone host. However, a user might sometimes use a host as a stepping-stone legitimately due to some special applications. If so, the watermark approach discussed in [10-12] for stepping-stone intrusion detection may produce false positive errors, since this method simply compares an incoming connection with an outgoing one.

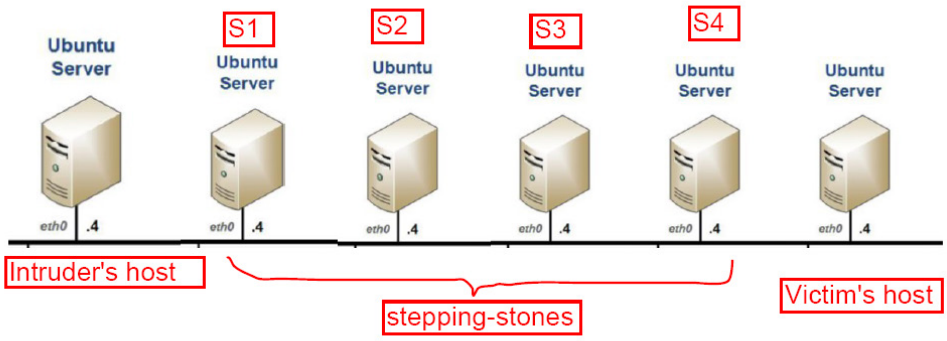


Figure 1: Four Stepping-stone Network Topology

A significant research conducted in [5] has shown that very few professional software employs three or more stepping-stones to access a remote server, although certain legal applications may utilize one or two stepping-stones to access a remote server. Therefore, in order to produce smaller false-positive errors to detect stepping-stone intrusion, an effective method is to estimate the length of a connection chain of stepping-stones. It is extremely challenging to estimate the length of an upper stream connection chain (from the attacker's host to the sensor in the connection chain). Thus, it is impossible to estimate the length of a whole connection chain. By far, most proposed approaches in the literature could only calculate the length of the downstream connection chain (from the sensor to the victim host). This approach to estimate the length of a downstream detection chain was investigated first in [13].

In [13], the authors studied the ratio between the Ack-RTT value and the Echo-RTT. Ack-RTT is defined as the gap between the time to send a packet out and the time to receive its corresponding acknowledgement packet. Echo-RTT is defined as the gap between the time to send a packet out and the time to receive its echo packet. In this way, the length of a downstream connection chain can be approximately estimated. However, this approach could incur false-negative errors.

In [14], the authors proposed a step-function approach motivated by the work that was done in [13] with the purpose of more accurately calculate the length of a downstream connection chain. In [15], the authors proposed another approach by mining network traffic to estimate the number of stepping-stones of a downstream connection chain in 2007. A couple of other methods were also developed in recent years for stepping-stone intrusion detection, including the method using the RTT-based random walk [16], and the method using the idea of RTT Cross-Matching [17].

The stepping-stone intrusion detection approaches have been investigated for about twenty-five years since 1995, unfortunately by far, these important methods have not yet been integrated into cybersecurity curricula at the college level in the U.S. It is vital to educate learners about the known detection approaches for stepping-stone intrusion as more and more professional attackers tend to launch their cyberattacks by using a chain of stepping-stones. Most universities/colleges' professors support to teach the skills and topics of ethical hacking and integrate them into the cybersecurity curricula due to two reasons. First, as far as we know, very few well-educated college students became malicious intruders; second, teaching offensive skills of ethical hacking for college students may produce more and more well-qualified professionals of cybersecurity workforce [18]. We propose ten hands-on labs that allow students to practice in various stepping-stone intrusion detection topics and help them better understand the topics included in the well-designed cybersecurity modules. These hands-on labs will also help enhance students' learning engagement significantly and greatly improve their hands-on experience in cybersecurity.

4. Hands-on Lab Development

Five modules for students to study stepping-stone intrusion and its detection techniques have been proposed and integrated into cybersecurity curriculum [3]. In these five modules, the most popular and the most recently developed techniques have been included. In order to help students to digest the detection and prevention techniques included in the five modules quickly and thoroughly, we design ten hands-on labs as the following,

1) setting up a stepping-stone intrusion connection chain;

2) capturing network traffic;

3) make C# code to capture network traffic;

4) content-based thumbprint detection;

5) time-based thumbprint detection;

6) step-function detection;

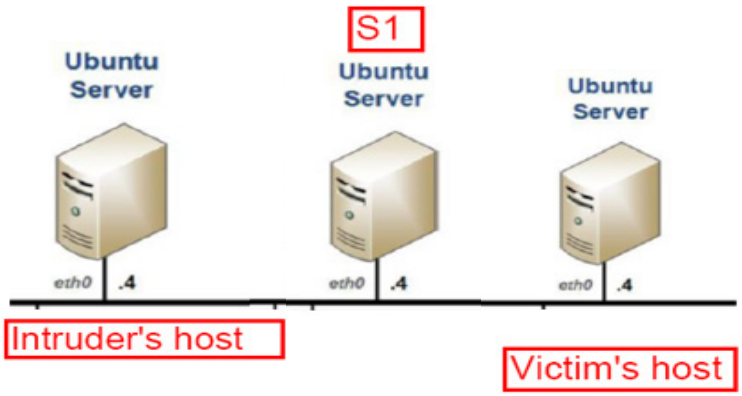
7) packet matching;

8) RTT-based random-walk detection;

9) estimating the length of a long connection chain;

10) intrusion detection using crossover packets.

We apply two rules including relevance and affordability to examine each hands-on lab developed. Relevance means if the lab is closely tied to the modules developed. Affordability means all the labs designed do not use expensive hardware and software. An ideal scenario is that students only need to use the Internet, and free download software to conduct the labs designed.



This designing rule can make it possible for most teaching-focus colleges/universities to offer the labs to cybersecurity majors. Depending on the curriculum design in different institutions, it is not necessary to adopt all the ten labs. However, Lab 1 and Lab 2 are not optional. All the computer hosts used in each lab must be connected in a local area network (LAN). Student must have login credential for each host. All the following labs share the same lab setup as below,

Hardware:

- Each computer must have minimally 4G memory and 500G hard drive capacity.
- Wired or Wireless computer network connection.

Software:

- Ubuntu server or any other type of Linux/Unix installed in each host.
- SSH/OpenSSH client side tool must be installed.
- Each host must have SSH server installed.
- Wireshark, or TcpDump

Login Credentials:

- User Name: Student (Assumed)
- Password: cpsc4166 (Assumed)

All the labs proposed in this paper need students to make a connection chain and to capture TCP/IP packets. A connection chain can be established using OpenSSH under Linux OS which can be a physically installed, or virtual one, such as an OS from VirtualBox, or VMware. It does not need too much memory and second storage. We tried computers with different memory sizes and storage capacity, and found that 4G memory and 500G storage are the minimized requirements. As for the software, TcpDump/Wireshark, SSH client and SSH server package are required minimally.

4.1. Setting up a Stepping-stone Intrusion Connection Chain

4.1.1 Lab objectives

1. Understand TCP/IP protocol; 2. Know how to establish a long interactive connection chain spanning multiple hosts; 3. Understand the concept of Stepping-stones; 4. Obtain the knowledge how an intruder lunches attacks over stepping-stones.

4.1.2 Network topology

It is the same topology as shown in Figure 1.

4.1.3 Lab instructions

1) Start up from any computer in the LAN, and login to a computer that is assumed the Intruder's host with the above credentials.
2) Please open a terminal at the Intruder's host.
3) Browse the current folder, and take a screenshot for the files in the folder.
4) Run SSH to connect to a local host S1: ssh Student@S1 (this can also be the IP address of S1 if host name S1 is not known) in the LAN.
5) As long as connecting to S1, you are prompted to input the password for the user.
6) If connected to S1 successfully, please browse the current folder, and take a screenshot including the folder's name, and all the files in the current folder. Run "ifconfig" to show the IP address and other network related information of S1. Take a screenshot of "ifconfig" results.
7) Compare the screenshot taken at the Intruder's host with the one taken at S1 to see if they are the same.
8) Repeat steps 4), 5), 6) 7) to connect to the computer hosts S2, S3, S4, and the last one respectively. The last host connected is called Victim's host.
9) So far you have locally connected to Victim's host via the hosts S1, S2, S3, and S4. Hosts S1, S2, S3, and S4 are used as stepping-stones.
10) If sniffing the packets at Victim's host, we can see all of the packets are from host S4 other than Intruder's host even though we know all the packets come from the Intruder's host originally. So in this way, intruders can protect themselves via the compromised hosts, such as the hosts S1, S2, S3 and S4.
11) Logout from Victim's host to S4 by typing "Exit" at Victim's

host.

12) Browse the current folder and compare with the screenshot taken at host S4 to see if it is disconnected from Victim's host.

13) Repeat steps 11) and 12) until come back to Intruder's host.

4.1.4 Critical Thinking Practice

1) Ethical Issue Discussion:
Please discuss if there are any ethical issues by making a connection chain across the Internet using legal credentials. How about is it by using illegal credentials?
2) Why do intruders make use of stepping-stones?
3) An interactive session can be encrypted by using SSH. Is it possible to get source IP and destination IP if a TCP/IP packet is captured from such a session? If yes, please tell how? If No, please tell why?
4) Compare to directly access a victim host, is it efficient to access the victim host via some compromised hosts?
5) In the lab, it has five connections in the long interactive session from Intruder's Host to Victim's Host. Each connection is encrypted and set up by using SSH/OpenSSH. Is the encryption key used for the connection from Intruder's Host to S1 the same as the encryption key used for the connection from S1 to S2? Why?

4.2. Capturing Network Traffic

4.2.1 Lab objectives

1. Understand the meaning of each field of a TCP/IP packet header; 2. Know how to store captured packets into different files; 3. Understand the features of TCP, UDP, IP, and ICMP packets; 4. Learn how to use Wireshark to capture network traffic.

4.2.2 Network topology

Refer to Figure 2.

4.2.3 Lab instructions

1) Select any three computer hosts in your local area network, and login to each host with the credentials given.

2) Run "ifconfig" to get the IP address at the three computers respectively and take a screenshot at each host.

3) Follow the instructions in Lab 1 to set up a connection chain as shown in Figure 2. This connection chain spans three computer hosts including Intruder's host, S1, and Victim's host.

4) Type some Linux/Unix commands at Intruder's host to make network traffic from Intruder's host to Victim's host via S1.

5) At S1, run Wireshark to capture TCP packets coming from Intruder's host and leaving to Victim's host only.

6) Store all the packet in Step 5) to a readable file (text file) including timestamp, source IP, destination IP, source Port number, destination Port number, Sequence number, Acknowledgement number, Flag, and Length.

7) At S1, run Wireshark to capture TCP packets coming from Victim's host and going to Intruder's host only. Repeat Step 6).

8) Repeat Steps 5), 6) and 7), but capture UDP packets.

9) Repeat Steps 5), 6) and 7), but capture ICMP packets.

4.2.4 Critical Thinking Practice

1) Ethical Issue Discussion:
2) Would it trigger any ethical issue to capture other users' network traffic under a host with legal login?
3) What is the difference between Display filter and Capture filter in Wireshark?
4) Give the display filter to find the packets of three-way handshake for a connection from host 192. 168.0.1.
5) What is a TCP Send packet?

4.3. Making a Code to Capture Network Traffic

4.3.1 Lab objectives

1. Understand LibPacp package for Linux server; 2. Learn the algorithms to capture computer network traffic; 3. Be able to make C code to capture TCP/IP Packets; 4. Obtain the knowledge to detect network adapters, and open an adapter; 5. Understand the techniques to set up and compile a packet-capturing filter.

4.3.2 Network topology

It has the same network topology as Figure 2 in Lab 4.2.

4.3.3 Mechanism on making the code to sniff network traffic

In order to make a code to capture network packets like what Wireshark does, Libpcap package must be installed in the Ubuntu server. If Windows server is used, please install WinPcap. The way to make a code to sniff computer network traffic is to call the functions built in Libpcap (packet capture) package. Libpcap provides an application-programming interface (API) for capturing network traffic.

We take an example, capturing raw IP packets, to examine the steps to sniff packets by making a program under Linx/Unix system. For the details of the code, please refer to the reference [19]. It has four steps to sniff computer network packets: 1) open a packet capture socket; 2) start packet capture loop; 3) parse and display packets; 4) Terminate capture program.

Open a packet capture socket: A socket is an endpoint for network communication that is identified in a program with a socket descriptor. Opening a packet capture socket involves a series of Libpcap calls that are encapsulated in open_pcap_socket() function. There are a couple of steps needed to open a packet capture socket. The first step is to select a network device using function pcap_lookupdev(). The second step is to open the network device selected for live capture using function pcap_open_live(). The third step is to call function pcap_lookupnet() to get the network address and subnet mask. The fourth step is to compile a packet capture filter by calling function pcap_compile(). The last step is to install the compiled packet filter program into the packet capture device. This causes

```c
pcap_t* open_pcap_socket(char* device, const char* bpfstr)
{
  char errbuf[PCAP_ERRBUF_SIZE];
  pcap_t* pd;
  uint32_t srcip, netmask;
  struct bpf_program bpf;

  // If no network interface (device) is specified, get the first one.
  if (!*device && !(device = pcap_lookupdev(errbuf)))
  {
    printf("pcap_lookupdev(): %s\n", errbuf);
    return NULL;
  }
  // Open the device for live capture, as opposed to reading a packet
  // capture file.
  if ((pd = pcap_open_live(device, BUFSIZ, 1, 0, errbuf)) == NULL)
  {
    printf("pcap_open_live(): %s\n", errbuf);
    return NULL;
  }
  // Get network device source IP address and netmask.
  if (pcap_lookupnet(device, &srcip, &amp;netmask, errbuf) < 0)
  {
    printf("pcap_lookupnet: %s\n", errbuf);
    return NULL;
  }
  // Convert the packet filter expression into a packet filter binary.
  if (pcap_compile(pd, &bpf, (char*)bpfstr, 0, netmask))
  {
    printf("pcap_compile(): %s\n", pcap_geterr(pd));
    return NULL;
  }
  // Assign the packet filter to the given libpcap socket.
  if (pcap_setfilter(pd, &bpf) < 0)
  {
    printf("pcap_setfilter(): %s\n", pcap_geterr(pd));
    return NULL;
  }
  return pd;
}
```

(a)

```c
void capture_loop(pcap_t* pd, int packets, pcap_handler func)
{
  int linktype;
  // Determine the datalink layer type.
  if ((linktype = pcap_datalink(pd)) < 0)
  {
    printf("pcap_datalink(): %s\n", pcap_geterr(pd));
    return;
  }
  // Set the datalink layer header size.
  switch (linktype)
  {
  case DLT_NULL:
    linkhdrlen = 4;
    break;
  case DLT_EN10MB:
    linkhdrlen = 14;
    break;
  case DLT_SLIP:
  case DLT_PPP:
    linkhdrlen = 24;
    break;
  default:
    printf("Unsupported datalink (%d)\n", linktype);
    return;
  }
  // Start capturing packets.
  if (pcap_loop(pd, packets, func, 0) < 0)
    printf("pcap_loop failed: %s\n", pcap_geterr(pd));
}
```

(b)

```c
void parse_packet(u_char *user, struct pcap_pkthdr *packethdr,
          u_char *packetptr)
{
  struct ip* iphdr;
  struct icmphdr* icmphdr;
  struct tcphdr* tcphdr;
  struct udphdr* udphdr;
  char iphdrInfo[256], srcip[256], dstip[256];
  unsigned short id, seq;
  // Skip the datalink layer header and get the IP header fields.
  packetptr += linkhdrlen;
  iphdr = (struct ip*)packetptr;
  strcpy(srcip, inet_ntoa(iphdr->ip_src));
  strcpy(dstip, inet_ntoa(iphdr->ip_dst));
  sprintf(iphdrInfo, "ID:%d TOS:0x%x, TTL:%d IpLen:%d DgLen:%d",
      ntohs(iphdr->ip_id), iphdr->ip_tos, iphdr->ip_ttl,
      4*iphdr->ip_hl, ntohs(iphdr->ip_len));
  packetptr += 4*iphdr->ip_hl;
  switch (iphdr->ip_p)
  {
  case IPPROTO_TCP:
    tcphdr = (struct tcphdr*)packetptr;
    printf("TCP  %s:%d -> %s:%d\n", srcip, ntohs(tcphdr->source),
        dstip, ntohs(tcphdr->dest));
    printf("%s\n", iphdrInfo);
    printf("%c%c%c%c%c%c Seq: 0x%x Ack: 0x%x Win: 0x%x TcpLen: %d\n",
        (tcphdr->urg ? 'U' : '*'),
        (tcphdr->ack ? 'A' : '*'),
        (tcphdr->psh ? 'P' : '*'),
        (tcphdr->rst ? 'R' : '*'),
        (tcphdr->syn ? 'S' : '*'),
        (tcphdr->fin ? 'F' : '*'),
        ntohl(tcphdr->seq), ntohl(tcphdr->ack_seq),
        ntohs(tcphdr->window), 4*tcphdr->doff);
    break;
```

(c)

Figure 3: Packet Capture Sample Code

Libpcap to start collecting the packets with selected filter. The sample code in Figure 3-(a) shows the four steps in opening a packet capture socket.

Start packet capture loop: Libpcap provides three functions to capture packets: pcap_next(), pcap_dispatch(), and pcap_loop(). Since function pcap_next() can only grab one packet at the time to be called. So the program must call this function in a loop to receive multiple packets. The other two functions pcap_loop and pcap_dispatch() can loop automatically to receive multiple packets. Datalink type can be determined by calling pcap_datalink(), and then start packet capture. The sample program shown in Figure 3-(b) uses pcap_loop() to sniff multiple packets. In this code, first to determine the datalink type by calling pcap_datalink(), and then start packet capture loop.

Parse and display packets: The general technique for parsing packets is to set a character pointer to the beginning of the packet buffer then advance this pointer to a particular protocol header by the size in bytes of the header that precede it in the packet. The header can then be mapped to an IP, TCP, UDP, and ICMP header structure by casting the character pointer to a protocol specific structure pointer. A parse_packet() function starts off by defining pointers to IP, TCP, UDP and ICMP header structures. The packet pointer is advanced past the datalink header by the number of bytes corresponding to the datalink type determined in capture_loop(). Casting the packet pointer to struct tcphdr and struct udphdr pointers gives us access to TCP and UDP header fields respectively. The struct icmphdr pointer enables us to display ICMP packet type and code along with the source and destination IP addresses. The sample code in Figure 3-(c) shows the steps to parse and display packets, such as TCP packets that are used to detect stepping-stone intrusion.

Terminate Capturing: The last step is to terminate the packet capture by interrupt signals SIGNIT, SIGTERM, and SIGQUIT through calling function bailout() which displays the packet count, closes the packet capture socket then exits the program.

4.3.5 Lab instructions

1) Start up running your code, and select the interface to sniff
2) Click “Start” button to start packet sniffing
3) Display the following information for each packet captured: source/destination IP address, source/destination port number, packet type, sequence number, acknowledge number, TCP flags, fragmentation information, checksum, receive window, TTL, upper layer protocol, timestamps in format of mm/dd/yy.
4) Click one TCP/IP packet captured to show the details in each of its header field. Take a screenshot for the header details.
5) Store captured packet in a .txt file that can be opened by WordPad, or any other text editor tool.

4.3.4 Critical Thinking Practice

1) Ethical Issue Discussion:
   Would it trigger any ethical issue to capture other users’ network traffic using self-made code under a host with legal login?
2) What is the difference between Winpcap and Libpcap?

3) What functions are called in order to open a packet capture socket?

4) What is the purpose to call pcap_compile()?
5) What is the function of pact_next()?
6) Which function is called to determine the datalink type of a packet?

4.4. Content-based Thumbprint Detection

4.4.1 Lab objectives

1. Understand TCP/IP protocols and network traffic behaviour; 2. Know how to establish an interactive TCP session; 3. Understand using Thumbprint to detect Stepping-stone intrusion; 4. To be familiar with TcpDump and Wireshark.

4.4.2 Network topology

The network topology used in this lab is the same as Figure 2 in Lab 4.2.

4.4.3 Lab instructions

1) Select any three computers in your local area network and name them to be Intruder's host, S1, and Victim's host.

2) Start up the computers in Linux and login to each host with given credentials. Open a terminal in each host.

3) Run "ifconfig" to get the IP address for each host, and take a screenshot from each host.

4) Run SSH from Intruder's host to connect to S1, then to Victim's host just as shown in Figure 2. An interactive session is set up spanning three hosts with S1 working as a Stepping-stone.

5) Students will monitor the traffic of the incoming connection from Intruder's host, and the traffic of the outgoing connection to Victim's host from S1. Here we use the number of TCP packets to represent the corresponding network traffic.

6) Run TcpDump at host S1 to monitor the TCP packets coming to/from Intruder's host but to S1 with destination/source port 22 and store all the packets in IncomingTCP.txt, and also monitor the TCP packets going to Victim's host or come back to S1 with destination/source port 22, and store all the collected packets to OutgoingTCP.txt.

7) In either IncomingTCP.txt or OutgoingTCP.txt, each packet is stored in one row including the following fields separated by ";": Packet Order number; Timestamp; Source IP; Destination IP; Source Port; Destination Port; Flag; Sequence Number; Acknowledge Number; Packet Length

8) Keep operating at Intruder's host for about 15 minutes to make network traffic to Victim's host via S1.

9) Count the number of packets in the two files respectively by counting the number of rows, or just simply check the last row "Packet Order number" field.

10) Compare the two number to see if they are close enough.

11) Identify the Send and Echo packets in the two files. Count the number of Send and Echo packets from IncomingTCP.txt, and denote them as In-S and In-E respectively. Similarly count the number of Send and Echo packets from OutgoingTCP.txt, and denote them as Out-S and Out-E respectively.

12) The rules to determine Send or Echo packet at S1 are as the following,

    a. Send packet is a packet in the incoming link that comes to S1 with Flag.P set up, but in the outgoing link that leaves S1 to Victim's host with Flag.P set up;

    b. Echo packet is a packet in the incoming link that leaves S1 to Intruder's host with Flag.P set up, but in the outgoing link that comes to S1 with Flag.P set up.

13) Compare if the following relation maintains,

    a. In-S is close to Out-S, and

    b. In-E is close to Out-E, and

    c. The sum of In-S and In-E is close to the sum of Out-S and Out-E

14) Please draw your conclusion based on the results from Steps 10) and 13).

4.4.4 Critical Thinking Practice

1) Ethical Issue Discussion:

   If a user has a legal login to a host, captures network packets, and obtains the contents of each packet, would the user's action result in an ethical issue?

2) What is the TcpDump command to sniff the packets in the incoming link?

3) What is the TcpDump command to sniff the packets in the outgoing link?

4) What conclusion you can make based on the information you have in step 10) of the Lab Instructions above? Why?

5) What conclusion you can make based on the information you have in step 13) of the Lab Instructions above? Why?

6) Write a TcpDump command to sniff the packets only acknowledge the requests from Intruders' Host at S1.

4.5. Time-based Thumbprint Detection

4.5.1 Lab objectives

1. Understand using time-based thumbprint to detect stepping-stone intrusion; 2. Learn how to generate time-based thumbprint; 3. Know how to compare time-based thumbprint; 4. Understand the efficiency of thumbprint comparison algorithm.

4.5.2 Network topology

The network topology used in this lab is the same as Figure 2 in Lab 4.2.

4.5.3 Lab instructions

1) Refer to Lab 1 to make an interactive TCP session with at least one host in between attacker and victim machines.
2) On either of the machine of your choice except the target, filter the network capture & save the incoming and outgoing packets including timestamp information for each packet through TcpDump.
3) Examine the packets for the incoming connection and look for the timestamp there and list those timestamps in a sequence.
4) Repeat Step 3 but for the outgoing connection
5) For the incoming connection sequence (list) of timestamps, find the difference in neighboring timestamps and list them in a sequence. This can give a sequence of time gaps for this connection. Find difference using the equation: $|p_i - p_{(i+1)}|$, here p_i is the timestamps of i^{th} packet captured.
6) Repeat Step 5 but for the outgoing connection.
7) Compare the two sequences to get a similarity. If the similarity is larger than a predefined threshold, the host is used as a stepping-stone. Otherwise, not.

4.5.4 Critical Thinking Practice

1) Ethical Issue Discussion:
   If a user has a legal login to a host, captures network packets, and but could not obtain the contents of each packet due to encryption, would the user's action result in an ethical issue?
2) Please describe what a time session-based thumbprint is in your own words.
3) Why would an individual want to perform this method to detect a stepping-stone over other methods?
4) Why do we compare the two sequences of time gaps in our own algorithm as oppose to the Longest Common Subsequence algorithm which can also help to measure similarity?
5) Do you have a better method of comparing the sequences' similarity?
6) Would a time session-based thumbprint be effective with an encrypted connection? If yes, explain why.

4.6. Step-function Detection

4.6.1 Lab objectives

1. Understand packet matching algorithm: First-Match; 2. Learn how to use matched Send and Echo packets to determine the number of compromised hosts; 3. Demonstrate Step-Function algorithm; 4. Illustrate the limits of Step-function detection.

4.6.2 Network topology

The network topology used in this lab is the same as the one shown in Figure 1 of Lab 4.1.

4.6.3 Lab instructions

1) Start up with any computers in the LAN, and login to the Intruder's host, Victim's host, S1, S2, S3, and S4 with the appropriate credentials to make a connection chain.
2) Open a terminal on Intruder's host and S1.
3) On desired sensor host (S1 for initial run), start TcpDump to dump captured packets to a file along with any further options
   a) ###.###.###.###.X is Sensor's IP Address and X is a port number
   b) sudo TcpDump 'tcp[tcpflags] & tcp-push != 0 and host ###.###.###.###.X' -n --number > capturedFile
4) On Intruder's host, Run SSH to connect to a remote host S1: ssh Student@S1 (this can also be the IP address of S1 if host name S1 is not known).
5) As long as S1 is reachable, you will be prompted to input the password for the user "Student".
6) On Intruder's host, repeat steps 4 and 5 replacing S1 with S2, S3, S4, and Victim's host, respectively, to login to further hosts as needed.
7) Interact with Victim's host: browse directories, manipulate files, check available interfaces, etc.
8) End current SSH session and stop TcpDump on the sensor host.
9) Repeat steps 3-8 for multiple setups; such as two/three stepping-stones chains with the sensor on different steps each time
10) You may want to use grep to create two files: one for Send packets and one for Echo. Consider that [^\2\]{2,} matches 22 for SSH
    a) (grep -E '>/b###.###.###.###.[^\2\]{2,}'/bcapturedFile) > downEchoFile
    b) (grep -E '###.###.###.###.[^\2\]{2,}/b>'/bcapturedFile) > downSendFile
11) Use First-Match Algorithm to match Send/Echo Packets:

a) Iterate through both lists, starting with the lowest sequence numbered Send Packet

b) If the current packet is a Send, add it to a list of unmatched Send packets

c) If it is an Echo and there is at least one unmatched Send Packet, Search the list of unmatched Send packets from the beginning. Find the first send packet with an appropriate acknowledgement number [Echo.Seq == Send.Ack].

d) Use the absolute difference between the correct Echo's and Send's timestamps to determine the round trip time (RTT) of the request [RTT = |Echo.Timestamp – Send.Timestamp|]

e) Save RTT to a list of RTTs

f) If it is an Echo and all preceding Send packets have been matched, the algorithm fails. Check if a packet was missed, then try to determine what may have occurred.

12) Sketch the graph of RTT vs. Number of matched Packets

a) RTT in whatever unit of time (typically ms or μs);

b) Number of matched packets indexed from 1 to the number of matches.

4.6.4 Critical Thinking Practice

1) Ethical Issue Discussion:

If a user has a legal login to a host, captures network packets, obtains the round-trip time between matched Send and Echo packets, but could not identify the contents of each packet due to encryption, would the user's action result in an ethical issue?

2) What is the purpose of tcp-push != 0 in the above capture?

3) Explain the difference in the grep statements listed above. Why does the first point to Send packets, while the second points to Echo packets?

4) Did you notice any effects to performance (positive/negative) as more links were introduced to the connection chain? Explain.

5) Would there be any difference to this analysis if the data were clear text, sent using Telnet, or encrypted like in SSH? Justify.

6) Can you determine the length of the entire connection chain with this method? If so, explain why. If not, which portion can you determine the length?

4.7. Packet Matching

4.7.1 Lab objectives

1. Understand the significance of packet matching; 2. Determine the differences in the different packet matching algorithms; 3. Learn how to apply packet matching to detect stepping-stone intrusion; 4. Distinguish the limits of different packet matching algorithms.

4.7.2 Network topology

The network topology used in this lab is the same as the one shown in Figure 2 of Lab 4.2.

4.7.3 Lab instructions

1) Start up any computers in the LAN, and login to the computer, which assumes to be called Intruder's host with the above credentials.

2) On desired sensor host (S1 for initial run), start TcpDump to dump captured packets to a file along with any further options

a) ###.###.###.###.X is Sensor IP Address and X is port number

b) sudo TcpDump 'tcp[tcpflags] & tcp-push != 0 and host ###.###.###.###.X' -n --number > capturedFile

3) Make an SSH connection chain from Intruder's host through any stepping-stone saying host S1 (sensor) to Victim's host.

4) Interact with Victim's host from Intruder's host via the connection chain: browse directories, manipulate files, check available interfaces, etc.

5) Terminate the SSH chain by using the 'exit' command on each of the stepping-stones and Victim's host from the shell of Intruder's host

6) You may want to use grep to create two files: one for Send packets and one for Echo

a) Upstream
   i. (grep '###.###.###.###.X/b>'/b</bcapturedFile) > upEchoFile
   ii. (grep '>/b###.###.###.###.X'/b</bcapturedFile) > upSendFile

b) Downstream – consider that [^\2\]{2,} matches 22 for SSH
   i. (grep -E '>/b###.###.###.###.[^\2\]{2,}'/bcapturedFile) > downEchoFile
   ii. (grep -E '###.###.###.###.[^\2\]{2,}/b>'/bcapturedFile) > downSendFile

7) Use First-Match Algorithm to match Send/Echo Packets:

a) Iterate through both lists, starting with the lowest sequence numbered Send Packet

b) If the current packet is a Send, add it to a list of unmatched Send packets

c) If it is an Echo and there is at least one unmatched Send Packet, Search the list of unmatched Send packets from the beginning. Find the first send packet with an appropriate acknowledgement number [Echo.Seq == Send.Ack].

d) Use the absolute difference between the correct Echo's and Send's timestamps to determine the round trip time (RTT) of the request [RTT = $|Echo.Timestamp - Send.TimeStamp|$]

i. Save RTT to a list of RTTs

e) If it is an Echo and all preceding Send packets have been matched, the algorithm fails. Check if a packet was missed, then try to determine what may have occurred.

8) Use the Conservative Algorithm to match Send/Echo Packets:

a) Iterate through both lists, starting with the lowest sequence numbered Send Packet

b) If the current packet is a Send:

i. If previous packet was Send and time gap was 1 second or more, clear the sendQ and make a note of match-flag = true

ii. Otherwise, add it to a list of unmatched Send packets

c) If it is an Echo:

i. If there is at least one unmatched Send Packet and match-flag = true, search the list of unmatched Send packets from the beginning. Find the first send packet with an appropriate acknowledgement/sequence number [Echo.Seq == Send.Ack && Echo.Ack > Send.Seq].

1. Use the absolute difference between the correct Echo's and Send's timestamps to determine the round trip time (RTT) of the request [RTT = $|Echo.Timestamp - Send.TimeStamp|$]

a. Save RTT to a list of RTTs

ii. Otherwise, set match-flag = false

9) Use the Greedy Heuristic Algorithm to match Send/Echo Packets:

a) Iterate through both lists, starting with the lowest sequence numbered Send Packet

b) If the current packet is a Send:

i. If previous packet was Send and time gap was 1 second or more, clear the sendQ

ii. Otherwise, add it to a list of unmatched Send packets

c) If it is an Echo:

i. If there is at least one unmatched Send Packet, search the list of unmatched Send packets from the beginning. Find the first send packet with an appropriate acknowledgement/sequence number [Echo.Seq == Send.Ack && Echo.Ack > Send.Seq].

1. Use the absolute difference between the correct Echo's and Send's timestamps to determine the round trip time (RTT) of the request [RTT = $|Echo.Timestamp - Send.TimeStamp|$]

a. Save RTT to a list of RTTs

ii. Otherwise, no match detected.

4.7.4 Critical Thinking Practice

1) Ethical Issue Discussion:
If a user has a legal login to a host, captures network packets, matches each Send packet with its corresponding Echo, but could not identify the contents of each packet due to encryption, would the user's action result in an ethical issue?

2) Which two TCP packet types can we exploit to properly match packets within a connection?

3) Explain why Echo.Seq == Send.Ack is used.
4) Explain why Echo.Ack > Send.Seq is used.
5) Why does Conservative Algorithm clear the Send Queue?
6) Looking at the results of running through the algorithms, what differences do you see between them? Explain why that might be.

4.8. RTT-based Random-walk Detection

4.8.1 Lab objectives

1. Understand random-walk model; 2. Learn how to apply random-walk model to detect stepping-stone intrusion; 3. Be familiar with the techniques to evade detection; 4. Demonstrate using RTT to resist intruders' evasion.

4.8.2 Network topology

The network topology used in this lab is the same as the one shown in Figure 2 of Lab 4.2.

4.8.3 Lab instructions

1) Refer to Lab 1 to make an interactive TCP session including at least one stepping –stone host that is used as a sensor.

2) On the sensor, filter the network capture & save the incoming and outgoing packets through TcpDump.

3) Examine the packets for the incoming connection, and match the Send & Echo packets using conservative packet matching algorithm from Lab 4.7, and obtain the number of RTTs from matched packets for this connection, N^{RTT}_{in}.

4) Repeat Step 3) for the packets collected from the outgoing connection, and obtain N^{RTT}_{out}.

5) Take the difference of N^{RTT}_{in} and N^{RTT}_{out}. $N^{RTT}_{in\text{-}out} = |N^{RTT}_{in} - N^{RTT}_{out}|$

6) Compare $N^{RTT}_{in\text{-}out}$ to a predefined upper bound. If it is less than the upper bound, then the incoming & outgoing connections are a relayed pair. The sensor is used as a stepping-stone. If not then, the machine is not used as a stepping-stone.

4.8.4 Critical Thinking Practice

1) Ethical Issue Discussion:
If a user has a legal login to a host, captures network packets, obtains the round-trip time between matched Send and Echo packets, but could not identify the contents of each packet due to encryption, would the user's action result in an ethical issue?
2) Please describe how a RTT-based Random-Walk Detection works in your own words.
3) Why would an individual want to perform this method to detect a stepping-stone over other methods?
4) Could an intruder manipulate this approach to give a false negative?
5) Would this method be effective with an encrypted connection? If yes, explain why.

6) Perform a network capture by following the above instructions with the predefined threshold, Γ, being equal 30. From the results, is the machine a stepping-stone?

4.9. Detection by Estimating the Length of a Long Connection Chain

4.9.1 Lab objectives

1. Understand the RTTs of the packets from the same connection chain can be mined to the same cluster; 2. Learn the number of compromised hosts is equal to the number of outstanding clusters; 3. Demonstrate the approach to estimate the length of a connection chain; 4. Obtain the knowledge on how clustering-partitioning algorithm can resist intruders' evasion.

4.9.2 Network topology

The network topology used in this lab is the same as the one shown in Figure 1 of Lab 4.1.

4.9.3 Lab instructions

1) Start up any computers in the LAN, and login to the computer that assumes to be called Intruder's host with the above credentials.
2) We will use at least 5 hosts in this connection chain. Decide which 5 hosts you want to use, and designate the 2nd host as a sensor host
3) On the sensor host, begin packet capture prior to making any of the connections.
4) Please open a terminal at Intruder's host.
5) Run SSH to connect to a remote host S1 (sensor host): ssh Student@S1 (this can also be the IP address of S1 if host name S1 is not known).
6) As long as connected to S1, you must be prompted to input the password for the user.
7) Repeat steps 4), 5), to connect to computer hosts S2, S3, S4, and the last one respectively. The last host you connect to remotely is called Victim's host.
8) So far you have remotely connected to Victim's host spanning hosts S1, S2, S3, and S4. Hosts S1, S2, S3, and S4 are used as stepping-stones in this lab.
9) Generate traffic to be captured by sensor. (ls, pwd, etc.)
10) After complete the packet capture, analyse the packets captured using clustering-partitioning algorithm. For the algorithm details, please refer to the reference [20].

4.9.4 Critical Thinking Practice

1) Ethical Issue Discussion:
   If a user has a legal login to a host, captures network packets, obtains the round-trip time between matched Send and Echo packets, and can estimate how many connections between the current host and the end of the connection chain. If the user could not identify the contents of each packet due to encryption, would the user's action result in an ethical issue?
2) Why is it important to begin packet capture before you initiate the connection chain? Please explain.
3) What results are we looking for after completing the clustering-partitioning algorithm? Why do these results indicate connections?
4) What is the maximum theoretical complexity of the partitioning clustering algorithm? Why is this algorithm likely never reach this complexity level? Please explain.
5) What percentage of the RTTs should be within a cluster to be considered a valid cluster?
6) If we collected 720 send packets and 810 echo packets, at most, how many comparisons would be necessary for partitioning-clustering algorithm?

4.10. Detection Using Crossover Packets

4.10.1 Lab objectives

1. Understand crossover packets; 2. Know the reason of generating crossover packets; 3. Obtain the relation between the length of a connection chain and the number of crossover packets; 4. Learn how to identify crossover packets.

4.10.2 Network topology

The network topology used in this lab is the same as the one shown in Figure 1 of Lab 4.1.

4.10.3 Lab instructions

We assume Intruder's Host is called iHost, and Victim's host is called vHost. After a connection chain is established, please type the following information at iHost to make some network traffic for each of the following: "This is s test from Hands-on lab 10. Please discard all the wrong messages!"

1) Make a connection chain from iHost to vHost via S1 only. Type the above information at iHost and capture Send and Echo packets at S1 from its outgoing connection. Store the packets to PacketFile1.
2) Make another connection chain from iHost to vHost, but via S1 and S2. Type the above information at iHost and capture Send and Echo packets at S1 from its outgoing connection. Store the packets to PacketFile2.
3) Make the third connection chain from iHost to vHost, but via S1, S2, and S3. Type the above information at iHost and capture Send and Echo packets at S1 from its outgoing connection. Store the packets to PacketFile3.
4) Make the fourth connection chain from iHost to vHost, but via S1, S2, S3, and S4. Type the above information at iHost and capture Send and Echo packets at S1 from its outgoing connection. Store the packets to PacketFile4.
5) Count the number Crossover packets in each file and compare them. Please conclude what you would find from the comparing the results.

4.10.4 Critical Thinking Practice

1) Ethical Issue Discussion:
   If a user has a legal login to a host, captures network packets, obtains the crossover packets, but could not identify the contents of each packet due to encryption, would the user's action result in an ethical issue?
2) Why is it unlikely that you will observe much, if any, Crossover in a LAN environment?
3) Does increasing the connection chain length increase or decrease the likelihood of observing packet Crossover? Why or why not?
4) Does packet Crossover help or hinder packet matching? Why?
5) Why are you more likely to observe packet Crossover in a WAN environment?
6) What information about a connection chain can you gather from detecting many packet Crossovers?

5. Discussion on the Labs Designed

In this session, we will discuss the innovation, contribution, and the effectiveness of the proposed work.

All the hands-on labs were designed based on some research papers. To the best of our knowledge, the is the first time that stepping-stone intrusion detection techniques are integrated into cybersecurity curriculum. The contribution is that college students can learn complex stepping-stone intrusion detection techniques and enhance their experience by conducting the hands-on labs. The labs designed are suitable for teaching-focus colleges who may have limited budget for their cybersecurity curriculum.

Each lab proposed has a critical thinking practice component including discussions about ethical issues, and the questions to train students to be qualified professionals of cybersecurity workforce. Most of the labs proposed were adopted in the course of "Intrusion Detection and Prevention" at Columbus State University, GA from 2018 to 2019. The instructors did class survey to ask the students if they agree with the labs adopted for the class. The survey results are shown in Table 1.

Table 1: Lab Survey Results

Item / Semester	Strongly Agree	Agree	Neutral	Disagree	Agree and Neutral Rate	Attending Rate
Spring 2018	5	4	3	1	92.3%	13 out of 15 = 86.7%
Spring 2019	11	9	6	0	100%	26 out of 28 = 92.9%
Spring 2020	9	5	2	2	88.88%	18 out of 19 = 94.7%
Spring 2021	11	11	4	2	92.9%	28 out of 29 = 96.6%
Average Rate					93.52%	92.73%

From the survey results, we can see that over four years, more than 90 percent of the students like the labs. Their comments and feedback are positive. There are also some negative comments and feedback. The following are some negative feedback extracted from the surveys:1) the time given to finish the labs are not enough; 2) most students prefer to use a physically installed Linux system to conduct the lab, other than a virtual Linux system because it is hard to copy the results out; 3) too many packets are required to capture which costs their too much time; 4) some students expect to have the first lab to refresh the Linux command, other than to make a connection chain.

6. Summary

In order to help college students to learn stepping-stone intrusion detection and prevention techniques and enhance their hands-on learning experience, we developed ten hands-on labs based on the significant results published in the area of stepping-stone intrusion detection since 1995. For making these hands-on labs be easily adopted by university professors in undergraduate cybersecurity courses, we used the following strategies while designing these hands-on labs: 1) save budgets for learners; 2) simplify the requirements for required hardware and software; 3) clear step-by-step instructions; 4) easy assessments by evaluators; 5) easy adoption by instructors.

Most of the hands-on labs we designed in this paper have been adopted in the undergraduate course of Intrusion Detection and Prevention at Columbus State University for four years. The average survey result shows that more than 90% of the students liked the labs and enjoyed the hand-on activities involved in the labs. The rate of disagreement/dislike is less than 10%. All the hands-on labs have been shared within the USA via the Clark system managed by Towson University, MD, USA. Records show that at least six colleges/universities downloaded the hands-on labs. We highly believe that our proposed hands-on labs in stepping-stone intrusion detection will help building the nation's cybersecurity workforce.

Cybersecurity is a rapidly changing and expending field. In order to make our students to be adaptable with fast changing cybersecurity techniques quickly after graduation, in the future, we will improve the proposed hands-on labs following NICE cybersecurity workforce framework initiated by NIST. In this framework, there are seven categories and each category contains one or more specialty areas. Each cybersecurity specialty area is composed of multiple work roles. Each work role includes Knowledge, Skills and Abilities (KSAs) and Tasks. The future hands-on labs will help our students to achieve three targets. First, they will obtain a body of information, which can be directly applied to the performance of a function. Second, they will enhance their skills needed for cybersecurity. Third, they will improve their competence to perform an observable behavior, which can result in an observable product.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work of Drs. Lixin Wang and Jianhua Yang is supported by National Security Agency (NSA) NCAE-C research grant H98230-20-1-0293 with Columbus State University, Columbus GA 31907, USA.

References

[1] P. Logan, A. Clarkson, "Teaching students to hack: curriculum issues in information security," Special Interest Group on Computer Science Education Symposium, St. Louis, MO USA, 2005.

[2] S. Bratus, A. Shubina, M.E. Lacasto, "Teaching the principles of the hacker curriculum to undergraduates," SIGCSE' 10, Milwaukee, Wisconsin USA, 2010.

[3] J. Yang, Y. Zhang, G. Zhao, "Integrate stepping-stone intrusion technique into cybersecurity curriculum," the Proceedings of 31st IEEE International Conference on Advanced Information Networking and Applications, Taipei, Taiwan, published in IEEE proceedings and Digital Library, 1-6, 2017, doi: 10.1109/WAINA.2017.29.

[4] S. Staniford-Chen, L.T. Heberlein, "Holding intruders accountable on the Internet," Proceedings of IEEE Symposium on Security and Privacy, Oakland, CA USA, 39-49, 1995, doi: 10.1109/SECPRI.1995.398921.

[5] Y. Zhang, V. Paxson, "Detecting stepping-stones," Proceedings of the 9th USENIX Security Symposium, Denver, CO USA, 67-81, 2000.

[6] K. Yoda, H. Etoh, "Finding connection chain for tracing intruders," Proceedings of 6th European Symposium on Research in Computer Security, Toulouse, France, Lecture Notes in Computer Science, 31-42, 2000.

[7] A. Blum, D. Song, S. Venkataraman, "Detection of interactive stepping-stones: algorithms and confidence bounds," Proceedings of International Symposium on Recent Advance in Intrusion Detection, Sophia Antipolis, France, 20-35, 2004.

[8] D.L. Donoho, "Detecting pairs of jittered interactive streams by exploiting maximum tolerable delay," Proceedings of 5th International Symposium on Recent Advances in Intrusion Detection, Zurich, Switzerland, 45-59, 2002.

[9] T. He, L. Tong, "Detecting encrypted stepping-stone connections," Proceedings of IEEE Transaction on signal processing, **55**(5), 1612-1623, 2007, doi: 10.1109/TSP.2006.890881.

[10] X. Wang, D.S. Reeves, S.F. Wu, J. Yuill, "Sleepy watermark tracing: an active network-based intrusion response framework," Proceedings of 16th International Conference on Information Security (IFIP/Sec'01), 369-384, 2001.

[11] X. Wang, D.S. Reeves, "Robust correlation of encrypted attack traffic through stepping-stones by manipulation of interpacket delays," Proceedings of ACM CCS '03, 2003.

[12] X. Wang, "The loop fallacy and serialization in tracing intrusion connections through stepping-stones," Proceedings of the 2004 ACM Symposium on Applied Computing, ACM Press, 2004.

[13] K.H. Yung, "Detecting long connecting chains of interactive terminal sessions," Proceedings of International Symposium on Recent Advance in Intrusion Detection (RAID), Zurich, Switzerland, 1-16, 2002.

[14] J. Yang, S.-H.S. Huang, "A real-time algorithm to detect long connection chains of interactive terminal sessions," Proceedings of 3rd ACM International Conference on Information Security (Infosecu'04), Shanghai, China, 198-203, 2004.

[15] J. Yang, S.-H.S. Huang, "Mining TCP/IP packets to detect stepping-stone intrusion," Journal of Computers and Security, Elsevier Ltd., **26**, 479-484, 2007, doi: 10.1016/j.cose.2007.07.001.

[16] J. Yang, Y. Zhang, "RTT-based random walk approach to detect stepping-stone intrusion," Proc. of 29th IEEE International Conference on Advanced Information Networking and Applications, Gwangju, South Korea, 558-563, 2015, doi: 10.1109/AINA.2015.236.

[17] J. Yang, "Resistance to chaff attack through TCP/IP packet cross-matching and RTT-based random walk," Proceedings of 30th IEEE International Conference on Advanced Information Networking and Applications, Crans-Montana, Switzerland, IEEE proceedings and Digital Library, 784-789, 2016, doi: 10.1109/AINA.2016.17.

[18] Z. Trabelsi, W. Ibrahim, "A hands-on approach for teaching denial of service attacks: a case study," Journal of information technology education: Innovations in Proactive, **12**, 299-319, 2013.

[19] J. Yang, L. Wang, B. Lockerbie, A. Lesh, "Manipulating network traffic to evade stepping-stone intrusion detection," Internet of Things, Elsevier, **3**(4), 34-45, 2018, doi: 10.1016/j.iot.2018.08.011.

[20] J. Yang, S.-H.S. Huang, M.D. Wan, "A clustering-partitioning algorithm to find TCP packet round-trip time for intrusion detection," Proceedings of 20th IEEE International Conference on Advanced Information Networking and Applications (AINA 2006), Vienna, Austria, **1**, 231-236, 2006, doi: 10.1109/AINA.2006.13.

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 368-375 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Personalized Serious Games for Improving Attention Skills among Palestinian Adolescents

Malak Amro[1] , Stephanny VicunaPolo[2], Rashid Jayousi[1], Radwan Qasrawi[1,*]

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

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

ARTICLE INFO

Article history:
Received: 15 July, 2021
Accepted: 09 August, 2021
Online: 26 August, 2021

Keywords:
Attention
Understanding
Cognitive skills
Serious Games
Game-based learning
Digital Games
Education outcomes

ABSTRACT

Serious games (SGs) are interactive and entertaining digital games with a special educational purpose. Studies have shown that SGs are effective in enhancing educational skills. Cognitive skills training through serious games have been used in improving students learning outcomes. In this article, we introduce the 'plants kingdom' serious game for improving adolescents' cognitive skills, mainly attention (Focus, selection, and sustained attention) and understanding skills. The game used the grade 8 Science book in designing the game content. The plant kingdom lesson was used for developing the game story and objects, its methods and tools were designed for the purpose of attention and understanding skills improvement. The game was evaluated on 43 students from public schools between the ages of 13-15 years, the study selected data from the students who had completed 5 playing sessions. The attention and understanding skills were assessed using the automatic recording and analysis of the game player's data. The variables utilized from the players' data included player ID, session number, gender, number of trials, level, drag and drop time, distance, reason for failure, position, speed, status, time, and playing tool. Results showed that the game improved the attention and understanding skills of students by 27% and 25 % respectively. The study showed the significant effect of serious games in enhancing students' cognition; thus, integrating serious games into the education system can potentially improve learning objectives and outcomes.

1. Introduction

In a rapidly changing digital environment, and the ever-growing number of persons interacting with digital technologies daily, Digital Games Based Learning (DGBL) have gained momentum in their application as learning aids for informal and formal education. DGBL provides users with the ability to improve decision-making, memory, mathematical, and spatial skills [1]. Digital Games that are used with the aim of educating, instructing, or training have come to be known as serious games (SG) [2]. Serious games were first defined by [1] as activities that "unite the seriousness of thought and problems that require it with the experimental and emotional freedom of active play"[1].

SGs differ from computer games in purpose, while the computer or digital games aim to entertain, serious games seek to achieve a certain educational, medical, or social outcome for the players. SGs have been used in a variety of fields, such as to aid children with Autism Spectrum Disorders (ASD) with Social Emotional Learning (SEL) [3], training the eyes of children with oculomotor dysfunction (OMD) [4], and cognitive training for chronic stroke survivors[5].

Several researchers have studied a further application of serious games whereby games are used to enhance educational skills [6], [7]. Serious games are directly correlated with an increase in the engagement, attraction, and interest of players in learning new skills[8]. Likewise, SGs may also enhance cognitive skills, such as selective attention, an important skill for students' academic performance [9]. Serious games show to increase players' motivation and adaptability to new educational content as the games can create scenarios that aren't easily accessible to them[10].

The ability to focus mental resources on the information most relevant at a given moment is referred to as attention. Students' attention skills are volatile and fluctuate subject to the environment they are in, which may hinder their learning skills or overall academic performance. Nonetheless, serious games allow for the creation of personalized training experiences that match

*Corresponding Author: Radwan Qasrawi, Email: radwan@staff.alquds.edu

the student's abilities and training needs and thus aid in the development of skills that enhance their cognitive functions, namely attention [11], [12].

2. Literature Review

Given that serious games aim to utilize new gaming technologies to enhance the skills of users in a particular field [12], this paper aims to study the application of serious games in adolescents' learning skills, particularly attention skills, to improve their educational outcomes. Several studies have explored the correlation of serious gaming and learning skills, and most recently, researchers have begun to focus on specific skills, such as attention and memory in a variety of sectors and applications [13], [14].

2.1. Serious Games and Learning Skills

In the article "Serious Play: Literacy, Learning and Digital Games" [15], the author discussed the Serious Play Project implemented by three Australian universities across two Australian states. The project aimed to learn about how digital games may be incorporated into education, namely the opportunities that games provide for creativity and innovation, and how learning through games challenges multimodal literacy learning. The project introduced games to classrooms between years 1 through 10 in a wide range of subjects, such as Information Technology, English, Literacy, and Social Studies among others. The findings show that serious games support learning skills in all subject areas, particularly in formal curriculum areas like Mathematics and Geography. The games enhanced students' interpersonal collaboration, negotiation, and autonomy.

In a previous study, a new model for the automatic collection of players' data and analysis of their skills was introduced by [8]. The model used game-learning analytics and robotic process automation for data collection and analysis, the model was tested on a sample of Palestinian public-school students between the ages of 13 and 14 (eighth grade). An automatic analysis model was created within a game focused on improving adolescent's attention and understanding of a given topic. The study found that serious games have the potential to be used as educational tools to ultimately aid in students' learning process and attention skills. Furthermore, the study found that the merging of game analytic tools and robotic process automation can be replicated in serious game development in order to measure players' enhanced skills. Thus, this approach to serious games could replace the traditional pre-post testing methodology currently used in such interventions.

The influence of serious games on formal education by examining their design and deployment was explored by [16]. Although the majority of studies on the topic of SGs have focused on their application in informal learning. In [16] researchers sought to develop a framework for introducing SGs into the formal pedagogical curriculum which pays special attention to the role of the educator. The review finds that serious games have the potential to significantly contribute to students' learning skills, yet, taking into account that game-based environments are rapidly evolving, educators and practitioners must be trained and prepared for the incorporation of new game-based teaching methodologies.

2.2. Serious Games and Attention (Cognitive skills)

In [17], researchers introduced a design for a computerized serious game to increase cognitive skills among the elderly "Smart Thinker." The game aimed to increase cognitive skills, such as memory and attention in elderly adults to fortify their cognitive performance. Their study consisted of 59 older adults playing Smart Thinker under the supervision of social workers. The findings show that the game, "Smart Thinker," had a significant impact in regards to the enhancement of cognitive and attention skills of the intervention group. Thus, serious games show to improve memory and attention skills among the elderly.

Similarly, in [5], the authors aimed to assess the value of "Neuro-World," a serious game, in cognitive training for stroke survivors who, as a result, present mild or moderate cognitive disabilities. The study utilized a hybrid model between ANOVA and Tukey's posthoc tests and found that all outcomes presented significant advancements except for language. Therefore, Neuro-World showed to enhance the cognitive function and decrease depression symptoms among the subjects in the intervention group [5].

In [11], researchers sought to study the assessment step of an SG for attention enhancement. A serious game for enhancing attention skills based on an Open Learned Model (OLM) was developed (Keep Attention), and a study was conducted to evaluate to what extent the OLM influences users' decision-making in attention training [11]. The study concluded that the Open Learner Model is effective in personalizing the user's experience, guaranteeing transparency, and helping users self-regulate their skills.

Furthermore, in [10], the authors took it one step further by approaching SGs for improving cognitive functions through the lens of Virtual Reality (VR). Thus, their study explored Virtual Reality Serious Games (VRSGs) to optimize the cognitive performance of users. To accomplish this, they utilize a combined 'Learning Mechanics –Game Mechanics' (LM-GM) model and further add VR characteristics. The study found that VRSGs improve the presence, immersion, and cognitive performance of users as the VR component reinforces players' embodied cognition through a purposeful and interactive design[10]., VR offers additional gaming experiences that don't translate to traditional computerized games.

In [18], the authors reviewed the use of educational serious games in all educational levels with a focus on knowledge, learning, memory, and attention. The review of the literature shows that a large number of researchers support the use of ICTs [19], in particular serious games, for enhancing attention skills and overall improving academic performance. Moreover, in [1], the authors showed that players demonstrate better selective attention over space, and can focus on one specific object at a given time with less effort than non-players as a result of their gaming activity. Thus, the study concludes that video games, particularly serious games, have the potential of significantly enhancing spatial skills, promote communication, improve memory and attention skills in students from all grade levels, but in particular in children with Attention Deficit Disorders (ADD).

The literature reviewed shows a consensus among researchers about the positive correlation of serious games with learning skills,

in particular skills that relate to cognition, such as memory and attention [20]. The impact of SGs in attention skills is increasingly applicable in school-aged children, particularly children with cognitive or intellectual disabilities.

3. Research Methodology

3.1 Serious Game Design

In this research, we developed an educational serious game for improving the basic cognitive processes of memory and attention, such as focus, orientation, recall, and selection, among school children. The game's design provides students with a series of training activities through game playing that would lead to the enhancement of their cognitive abilities, particularly attention skills. The game used a mixture of instructional and interactive video-game-play methods to achieve the proposed objectives.

The "Plant Kingdom" SG was developed based on Bloom's hierarchy of cognitive learning [21]. Therefore, the game seeks to aid students in navigating several levels to achieve increased attention skills. The research focused on improving the attention and understanding cognitive skills of adolescents, including logical thinking, intellectual, awareness, observation, knowledge, interaction, intuition, and decision-making. The game content was planned in a way to stimulate the player's recall, concentration, and memory skills through visual and audio activities, and motivation and rewards actions.

The grade eight science book was used in developing the serious game's playing materials and objects. The course learning outcomes were considered as a reference for the game objectives' design. The SG includes eight playing levels with ascending difficulties designed based on the desired skills with the help of gaming tools such as mazes, mix-match, basket-filling, and maps. The levels' difficulty was adjusted by altering game variables, including touch sensitivity, length and complexity of the game pattern, game tool, response time, success and failure, and reasons for failure.

Figure 1: The left side indicates lesson subtopics in Arabic (4 titles: plants characteristics, physical materials, structures of plants, and plants classification. The right side shows plant pictures as described in the students' textbook.

The serious game application utilized the learning contents of the grade 8 science course (specifically: the plant kingdom lesson in the students' science course curriculum) and furthered modeled them in the framework of skills development. To achieve this, objects' shapes, colors, motion, size, location, and other related human-computer interaction standards were considered in the game application development. In addition, the development team considered brain processing measurement features during game design, such as response, response time, and decision making. A group of experts (serious game experts, teachers, and education supervisors) worked on the creation of the game story, scenarios, objects, and characters.

Figure 2. Two game levels (simple and complex). Simple level (left-side) used a standard plan classification tree, and the complex level (right-side) used a recall method of the classification order of plants.

Figure 3. Two game levels with different playing tools. The Maze (left-side) drags plant properties from the upper list onto the corresponding baskets through the maze playing tool. The Tree (right-side) drags the plant properties to the correct location using finger touch movements

Figures 1, 2, and 3 show a sample of game-playing tools that have been used in game design to enhance the development of cognitive skills among adolescents.

3.2 Game Development

The "Plant Kingdom" game was developed following the 'normal serious game' design methodology that achieved the game objectives in enhancing adolescents' cognitive skills. Figure 4 illustrates the methodological steps undertaken during game design, from the definition of game objectives until the game's evaluation. Steps 1-5 include the development of the game's story and scenarios, which reflect the plant kingdom textbook material. The objects and pictures were selected to match those shown in the student textbook.

The scenarios followed the teaching material's learning outcomes and were interactively designed to enhance students' motivation within the learning environment. Step 6 concerns the game tools design, in which Mazes, drag and drop, and mix and were used for the study's purposes. These tools are commonly found in games and are considered efficient tools for enhancing players' concentration, planning, and visual-motor skills integration. Touch sensitivity and movement speed have also been considered in the game design. Sensitivity and speed complexity become increasingly difficult with higher game levels. The final step concerns game evaluation, whereby game player data were automatically collected and managed using robotic process automation and automatic analytical tools for evaluating the serious game objectives [8].

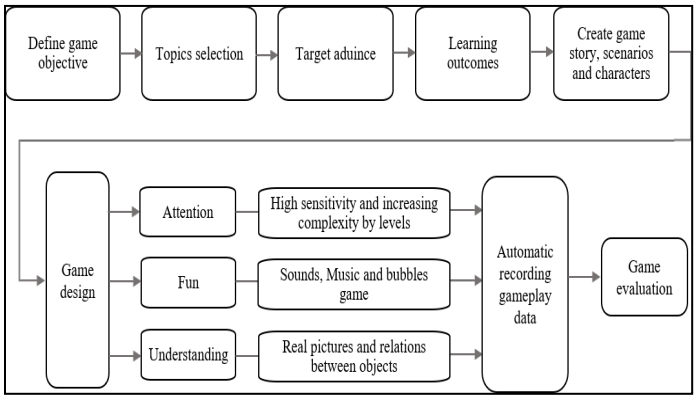


Figure 4: Methodological steps of game design

Utilizing the standard game design procedure, the game development stage for this study included the following aspects: storytelling, goal, feedback, time, rewards, instructional methods, and rules. Storytelling was employed in different ways, such as drag and drop and Mix and match. The maze itself was designed as an aid for players' attention as users were forced to concentrate on the element while moving it across the screen. The elements were shuffled after a certain period of time and players had to withdraw specific images before and during shuffling. Figures 1-3 illustrate the design of game levels using the maze, drag, and drop, and mix and match.

The software development incorporated the use of the Unity 3D mobile App development environment, in addition to free software development tools, such as PHP and MySQL to maintain user records and feedback database. Figures 5 and 6 show screenshots of the game generation stage using the Unity 3D video-game engine (version 2019.3.11), PHP MyAdmin 5.1.1, and MySQL development tools. Microsoft Visual Studio 2019 was used to edit the program code in C#.

In addition, a Web-based platform was created to manage users' game data, increase visibility, and deliver information about the SG. Programming and video-game experts were responsible for game design and implementation, as per international guidelines. The prototype was validated by end-users (stakeholders from across several domains) and tested in real-life scenarios with real data in the early development stages.

3.3 Game Testing and Validation

Prior to the release of the final prototype, the game was pilot-tested twice, firstly by 3 teachers and 4 eighth-grade students, and secondly by two groups. The groups comprised the original testing sample and an additional group made up of 2 teachers and 5 eighth-grade students. Test groups were asked to play the game 'as many times as desired' for approximately 30-40 minutes dedicated to each level in order to complete all 8 levels. Pre-post data were collected upon finalization of the pilot-test, analyzed, and considered in the next version of the game. Finally, the game's final version was launched and reviewed by experts and teachers before being given to the student sample.

3.4 Data Collection

The game's target audience was eighth-grade students between the ages of 13 and 14. The study's target population was selected following the recommendations of the Palestinian Ministry of Education (MoE) supervisors. The target age was chosen given that 8th grade proves critical as students transition to adolescence and are increasingly exposed to smart technologies.

A sample of 60 students weighted by gender was selected, however, only 43 students completed the study. The rest of the students failed to complete more than 3 sessions and were

Figure 5: Sample of game programing using Unity

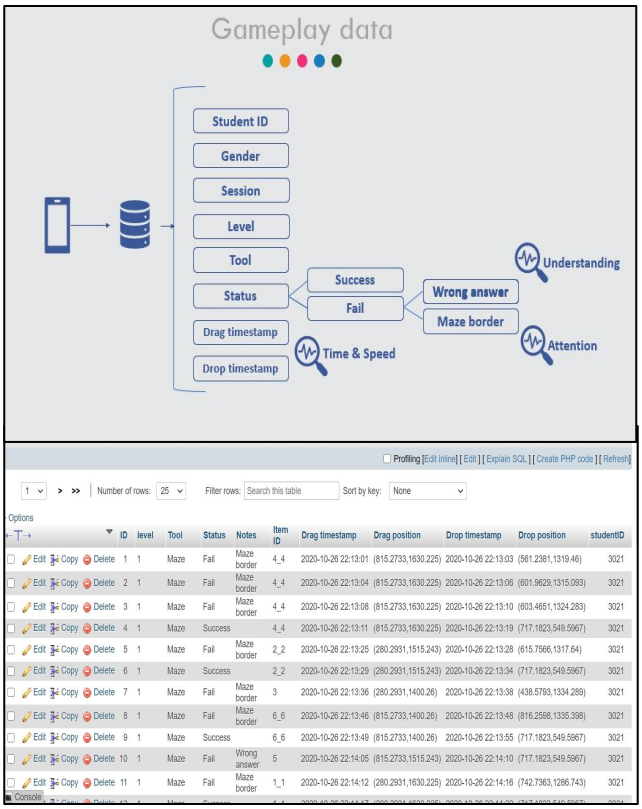


Figure 6: Sample of game programing using PHP and MySQL

therefore excluded from the analysis. The student's parents signed consent forms before gameplay. Participants were required to have access to a mobile phone with an active internet connection.

The game's application was accessible for download on the Google Play store. A research assistant guided the student in downloading the game into their smartphones. Students were asked to play the game in their homes due to COVID19 pandemic protection procedures. The students' school teachers and science teachers supervised and mentored students throughout game-playing. The player's game data was automatically stored in the game platform. The data included the variables: player ID, gender, number of trials, playing tool, session number, level number, location, drag and drop time, status (Success or Fail), the reason for failure (bumped into the maze border or wrong answer), distance, time, and speed. Moreover, attention and understanding variables were created in the following way:

a) The attention variable: Interpreted as the relationship between the amount of time it takes to complete a task, the player's status (Success or Failure), and the cause of failure.
b) The understanding variable: Interpreted as the ratio between correct and incorrect responses, the number of trials per level, the rise in correct answers, and the decline of incorrect answers.

3.5. Experiment Setting

The research took place under direct supervision and partnership with the Palestinian Ministry of Education (MoE). The MoE approved the testing of the game prototype and was responsible for the nomination of 2 science supervisors and 2 science teachers to oversee the study's implementation. The Ministry was also responsible for identifying and selecting participants. Following parents' approval upon communication with the supervisors, students attended an in-person orientation session to explain the research objectives and aid participants in downloading and installing the SG onto their mobile phones. Additionally, supervisors provided students with detailed instructions for gameplay as participants were to use the game at home. Figure 7 shows participants making use of the "Plant Kingdom" serious game.

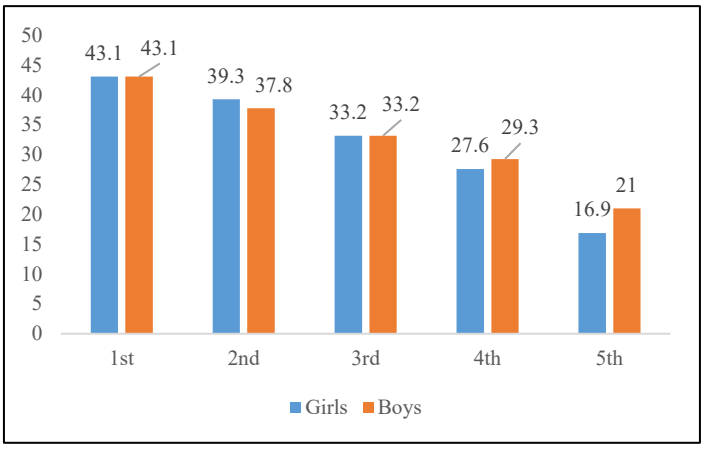


Figure 7: A sample of students playing the game.

The students were asked to play one level per day. To achieve the game objectives, the minimum required number of playing sessions was five. Out of a sample of 60 participants, 43 students successfully completed the 5 sessions required.

4. Results

4.1 Population Studied

The game was tested on 43 grade 8 students (31 girls and 13 boys) aged 13-15 years. The results are shown in Table 1 detail the percentage of participants' distribution by gender and playing sessions. Overall, the number of playing trials for the 43 players was 25,494. The results evidenced that the first session had a higher number of trials than the last session among girls (69.7% and 67.5% respectively). On the other hand, boys reported a higher number of trials in the last session than in the first session (32.5% and 30.3% respectively. However, girls presented a higher overall number of trials than boys (69% and 31% respectively);

this has been attributed to the fact that the girls to boys ratio in the sample is 31:13.

Table 1: Participants distribution of playing trials by gender and playing sessions.

		Gender		
		Girls	Boys	Total
		n (%)	n (%)	N
Session	1	4730 (69.7)	2059 (30.3)	6789
	2	3982 (69.4)	1756 (30.6)	5738
	3	3366 (69.2)	1498 (30.8)	4864
	4	3039 (68.1)	1422 (31.9)	4461
	5	2458 (67.5)	1184 (32.5)	3642
	Total	17575(69)	7919(31)	25494

4.2 Outcomes

Results in Table 2 show the average playing time (minutes) by gender and playing session. Results indicated that the average playing time decreased as the playing sessions increased for both girls and boys. The average playing time for Session 1 compared to Session 5 by gender (in minutes) was 10.7 to 5.9; 3.7 to 2.0 for both girls and boys respectively. Overall, the playing time deviation decreased by 2 minutes while the number of playing sessions increased.

Table 2: Average playing time (minutes) by gender and playing session.

		Player Gender		
		Girls	Boys	Total
Session	1	10.7	3.7	14.4
	2	8.9	3.1	12.0
	3	7.6	2.4	10.0
	4	6.7	2.5	9.2
	5	5.9	2.0	7.9

Figure 8 shows the percentage of player status (success or fail) by gender and playing sessions. The failures decreased as playing sessions increased among both girls and boys. 43.1% of both boys and girls failed in session 1, while only 16.9% of girls and 21% of boys failed in Session 5. The girls reported a failure enhancement of 26.2%, while boys reported an enhancement of 24.8%. The results indicated that girls reported an overall 83.1% success rate in the last session, while boys reported a success rate of 79.0% in the last session.

The status of participants was analyzed through the collection of individual data on the reason for failure at each level. The results shown in Table 3 indicate participants' reported reason for failure as either a wrong movement or simply a wrong answer.

Table 3 compares players by gender. In session 1 (pre-test), girls showed a higher percentage of wrong movements than boys (35.8% and 32.3% respectively). In the fifth session (post-test), girls had a greater decrease in the amount of border touch and wrong movements than boys (7% and 8.4% respectively).

Table 3: The game-play status by gender and playing session.

		Player Gender					
		Female		Male		Total	
		Wrong Touch	Wrong Answer	Wrong Touch	Wrong Answer	Wrong Touch	Wrong Answer
		n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Session	1	980 (35.8)	1057 (32.6)	455 (32.3)	433 (33.1)	1435 (34.6)	1490 (32.8)
	2	706 (25.8)	859 (26.5)	358 (25.4)	306 (23.4)	1064 (25.7)	1165 (25.6)
	3	494 (18.1)	623 (19.2)	256 (18.2)	242 (18.5)	750 (18.1)	865 (19.0)
	4	363 (13.3)	476 (14.7)	220 (15.6)	197 (15.0)	583 (14.1)	673 (14.8)
	5	192 (7.0)	224 (6.9)	118 (8.4)	131 (10.0)	310 (7.5)	355 (7.8)

Table 4: Average time difference of player fails per gender and playing session (minutes).

		Player Gender					
		Female		Male		Total	
		Wrong Touch	Wrong Answer	Wrong Touch	Wrong Answer	Wrong Touch	Wrong Answer
		X^2	X^2	X^2	X^2	X^2	X^2
Session	1	.15	.13	.12	.11	.14	.13
	2	.12	.13	.12	.09	.12	.12
	3	.13	.13	.10	.08	.12	.12
	4	.16	.13	.10	.10	.14	.12
	5	.18	.17	.10	.10	.15	.14

Table 5: Deviation percentage in attention and understanding using pre-posttest analysis by gender and player status.

	Girls	Boys	Total	P-Value
Attention	27.1	23.5	26	0.0001
Understanding	25.7	23.1	25	0.0001

The average playing time was reported for each level and playing session. The average time of game completion was 2.5 minutes, while the average time per session was 34 minutes. Table 4 shows the players' fail average time difference, the results show an increase in the average time difference between sessions one and five. The girls' fail time increased from 0.15 minutes to 0.18 minutes from sessions 1 to 5. Therefore, girls spent more time making use of their focus and attention skills while game-playing, which explains the decrease of fails from 35.8% to 7% as shown in Table 3. For boys, the time difference decreased from 0.12 minutes to 0.1 minutes.

The serious game effectiveness was assessed by evaluating the changes in-game players' results. The percentage of deviation in fails rate, average playing time, and fails between sessions one and five indicates the effectiveness of the proposed game in improving students' attention and understanding skills. Results in Table 5 show the deviation percentage for attention and understanding skills by gender. Attention skills improved by 26% (27.1% girls and 23.5% Boys) and understanding skills improved

by 25%. Girls reported a higher improvement than boys for both skills.

5. Discussions and Recommendations

Towards our goal of enhancing cognitive skills training among school children, we have created a serious game titled "plant-kingdom" and measured its impact on attention skills improvement through a random sample of grade 8 students. We used the game players' data for analyzing the game results and evaluating the outcomes. The average playing time per session leading to improvements in attention skills was 34 minutes a day, two times a week. The analysis of data considered every student who completed a minimum of five playing sessions. The results obtained were positive, we observed a statistically significant improvement in attention skills through the increase of average playing time, decrease in playing fails, and increase in sessions' success rate. Our results were found to be consistent with other similar studies [22, 23].

The study also yielded a statistically relevant reduction of 'fails' as playing sessions increased. Hence, the data analyzed indicated that the total number of trials decreased as the number of sessions increased. Likewise, the success rate increased while the failure rate and average playing time per level decreased. The above indicators reflect that students attained higher levels of attention and understanding of the science lesson as they continued to play the 'plant kingdom's serious game.

In the same manner, a comparison of players' outcomes indicates that the skills of sustained and selective attention grew among players as sessions increased. This is evidenced in the fact that as game-playing continued, students, girls to a larger extent, developed the ability to stay within the playing borders on the touch screen while dragging and dropping the game object.

Moreover, a notable advancement in understanding skills was observed during the answer activities. The students obtained a lesser amount of wrong answers as the game sessions progressed. Thus, the understanding of the game's content, and of the lesson plan by extension, has remarkably improved after game-play. Self-correction skills, understanding, and attention are all skills that the students had to improve on during game-play in order to decrease the 'fail' rate.

Given that the game was designed based on the students' 8th-grade Science material, the 'Plant-kingdom' serious game proved to be useful in the achievement of the course's educational outcomes. Therefore, our study's results evidence the effectiveness of serious games in improving students' cognitive skills. The skills enhancement reported in other studies [24], [25] is consistent with the results obtained. The reported improvement in players' outcomes reflects the value of serious games in educational contexts, particularly attention skills (focus, selection, and sustained attention). Furthermore, the current study, in resonance with similar previous studies [16], [26], [27], demonstrates that serious games are able to promote the enhancement of attention and understanding skills among adolescents.

Nonetheless, it is worthy to note that the study had several limitations. Firstly, with only 43 students studied, the research possessed a small sample size. Although the preliminary results presented here were enough to examine the effect of the Plant-Kingdom game in improving attention and understanding skills, we recommend the replication of this study with a larger and more significant population.

Secondly, the experiment was conducted under COVID-19 restrictions and Ministry of Health guidelines, therefore challenging the safe and correct implementation of the research. For this reason, a virtual platform was used for game supervision, thus limiting the game's evaluation. We recommend future researchers deliver their proposed intervention in person to facilitate the evaluation process, and supervise the correct use of the serious game, especially for children.

Finally, the game used the automatic recording and analysis of game players' data instead of the traditional pre-post testing measure. Thus, the number of variables used in the evaluation of game effectiveness was limited. We recommend that future studies consider the evaluation of a larger number of variables

6. Conclusions

Firstly, the preliminary results presented in this study could be used in larger-scale research to further investigate the impact of SGs in achieving learning outcomes and enhancing student's cognitive skills development. The expansion of this research could bring about innovative change to the field of education and technology.

The unification of serious games in formal education is just at its onset, but as evidenced in this research, could be greatly beneficial for students' learning outcomes. Thus, this study concludes that utilizing serious games as a learning tool informal education settings (i.e. schools, academies, universities), would enhance students' cognitive skills and improve their learning outcomes by extension, all the while providing a fun, innovative, and interactive learning methodology.

Furthermore, the serious game methodology in formal education could prove even more beneficial than evidenced in this paper for students diagnosed with Attention Deficit Disorders, Autism, or other learning disabilities.

7. Abbreviations

Digital Games Base Learning (DGBL), Serious Games (SG), Autism Spectrum Disorder (ASD), Social Emotional Learning (SEL), Oculomotor Dysfunction (OMD), Open Learner Model (OLM), Virtual Reality (VR), Virtual Reality Serious Games (VRSG), Learning Mechanics- Game Mechanics (LM-GM), Attention Deficit Disorders (ADD), Hypertext Pre-Processor (PHP), Ministry of Education (MoE).

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors wish to thank and acknowledge all participants, their families, and the schools that contributed to the success of this

study. The authors acknowledge the role of the l thanks to the Palestinian Ministry of Education, the INTETECH company, and Al Quds University for aiding in the implementation of this research project.

References

[1] G. Papanastasiou, A. Drigas, C. Skianis, M.D. Lytras, "Serious games in K-12 education: Benefits and impacts on students with attention, memory and developmental disabilities," Program, **51**(4), 424–440, 2017, doi:10.1108/PROG-02-2016-0020.

[2] P. Wouters, H. van Oostendorp, Overview of Instructional Techniques to Facilitate Learning and Motivation of Serious Games, 2017, doi:10.1007/978-3-319-39298-1_1.

[3] C. Wu, Q. Zheng, Simulation and Serious Games for Education, 2016.

[4] I. Heldal, C. Helgesen, Q. Ali, D. Patel, A.B. Geitung, H. Pettersen, "Supporting school aged children to train their vision by using serious games," Computers, **10**(4), 2021, doi:10.3390/computers10040053.

[5] H.T. Jung, J.F. Daneault, T. Nanglo, H. Lee, B. Kim, Y. Kim, S.I. Lee, "Effectiveness of a serious game for cognitive training in chronic stroke survivors with mild-to-moderate cognitive impairment: A pilot randomized controlled trial," Applied Sciences (Switzerland), **10**(19), 2020, doi:10.3390/APP10196703.

[6] J.I. Navarro, E. Marchena, C. Alcalde, G. Ruiz, I. Llorens, M. Aguilar, "Improving attention behaviour in primary and secondary school children with a Computer Assisted Instruction Procedure," International Journal of Psychology, **38**(6), 359–365, 2003, doi:10.1080/00207590244000042.

[7] D.K. Ramos, H.M. Melo, "Can digital games in school improve attention? A study of Brazilian elementary school students," Journal of Computers in Education, **6**(1), 5–19, 2019, doi:10.1007/s40692-018-0111-3.

[8] R. Qasrawi, M. Amro, R. Jayousi, "Automatic analytics model for learning skills analysis using game player data and robotic process automation in a serious game for education," Proceedings - 2020 International Conference on Promising Electronic Technologies, ICPET 2020, 94–98, 2020, doi:10.1109/ICPET51420.2020.00026.

[9] T. Moore, M. Zirnsak, "Neural Mechanisms of Selective Visual Attention," Annual Review of Psychology, **68**, 47–72, 2017, doi:10.1146/annurev-psych-122414-033400.

[10] M. Alcañiz, S. Göbel, M. Ma, M. Fradinho, J. Baalsrud, H. Tim, M. Eds, D. Hutchison, Serious Games, 2017, doi:10.1007/978-3-319-70111-0.

[11] N. Hocine, "Personalized Serious Games for Self-regulated Attention Training," ACM UMAP 2019 Adjunct - Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, (ExHUM), 251–255, 2019, doi:10.1145/3314183.3323458.

[12] J. Mishra, D. Bavelier, A. Gazzaley, "How to Assess Gaming-Induced Benefits on Attention and Working Memory," Games for Health Journal, **1**(3), 192–198, 2012, doi:10.1089/g4h.2011.0033.

[13] H.-L. Chan, W. V. Giannobile, R.M. Eber, J.P. Simmer, J.C. Hu, " Characterization of Periodontal Structures of Enamelin -Null Mice ," Journal of Periodontology, **85**(1), 195–203, 2014, doi:10.1902/jop.2013.120651.

[14] N. Akshoomoff, "Selective attention and active engagement in young children," Developmental Neuropsychology, **22**(3), 625–642, 2002, doi:10.1207/S15326942DN2203_4.

[15] سینا ا., No Title1386 ,قانون در طب.

[16] S. Arnab, R. Berta, J. Earp, F. de Sara, M. Popescu, M. Romero, I. Stanescu, M. Usart, "Framing the adoption of serious games in formal education," Electronic Journal of E-Learning, **10**(2), 159–171, 2012.

[17] H. Chi, E. Agama, Z.G. Prodanoff, "Developing serious games to promote cognitive abilities for the elderly," 2017 IEEE 5th International Conference on Serious Games and Applications for Health, SeGAH 2017, 2017, doi:10.1109/SeGAH.2017.7939279.

[18] G. Papanastasiou, A. Drigas, C. Skianis, M.D. Lytras, "Serious games in K-12 education," Program, **51**(4), 424–440, 2017, doi:10.1108/PROG-02-2016-0020.

[19] I.D. Danilov, "Geological and paleoclimatic evolution of the Arctic during Late Cenozoic time," The Arctic Seas, **10**, 759–760, 1989, doi:10.1007/978-1-4613-0677-1_27.

[20] V. Bolón-Canedo, A. Alonso-Betanzos, Intelligent Systems Reference Library 147 Recent Advances in Ensembles for Feature Selection.

[21] H.A. Efe, R. Efe, "Evaluating the effect of computer simulations on secondary biology instruction: An application of Bloom's taxonomy," Scientific Research and Essays, **6**(10), 2137–2146, 2011, doi:10.5897/sre10.1025.

[22] C. Alonso-Fernández, I. Martínez-Ortiz, R. Caballero, M. Freire, B. Fernández-Manjón, "Predicting students' knowledge after playing a serious game based on learning analytics data: A case study," Journal of Computer Assisted Learning, **36**(3), 350–358, 2020, doi:10.1111/jcal.12405.

[23] Y. Wang, P. Rajan, C.S. Sankar, P.K. Raju, "Let Them Play: The Impact of Mechanics and Dynamics of a Serious Game on Student Perceptions of Learning Engagement," IEEE Transactions on Learning Technologies, **10**(4), 514–525, 2016, doi:10.1109/tlt.2016.2639019.

[24] C. Boletsis, S. McCallum, "Smartkuber: A Serious Game for Cognitive Health Screening of Elderly Players," Games for Health Journal, **5**(4), 241–251, 2016, doi:10.1089/g4h.2015.0107.

[25] M.H. Chen, W.T. Tseng, T.Y. Hsiao, "The effectiveness of digital game-based vocabulary learning: A framework-based view of meta-analysis," British Journal of Educational Technology, **49**(1), 69–77, 2018, doi:10.1111/bjet.12526.

[26] P. Cardoso-Leite, D. Bavelier, "Video game play, attention, and learning: How to shape the development of attention and influence learning?," Current Opinion in Neurology, **27**(2), 185–191, 2014, doi:10.1097/WCO.0000000000000077.

[27] J.M. Halperin, D.J. Marks, A.C. V. Bedard, A. Chacko, J.T. Curchack, C.A. Yoon, D.M. Healey, "Training Executive, Attention, and Motor Skills: A Proof-of-Concept Study in Preschool Children With ADHD," Journal of Attention Disorders, **17**(8), 711–721, 2013, doi:10.1177/1087054711435681.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 376-384 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques

Zhiyuan Chen*, Howe Seng Goh, Kai Ling Sin, Kelly Lim, Nicole Ka Hei Chung, Xin Yu Liew

School of Computer Science, University of Nottingham Malaysia, Semenyih, 43500, Malaysia

ARTICLE INFO

Article history:
Received: 24 June, 2021
Accepted: 10 August, 2021
Online: 26 August, 2021

Keywords:
Agriculture Commodity Price Prediction
Machine Learning
Long Short-Term Memory Model
Mean Square Error
ARIMA
SVR
Prophet
XGBoost

ABSTRACT

The intention of this research is to study and design an automated agriculture commodity price prediction system with novel machine learning techniques. Due to the increasing large amounts historical data of agricultural commodity prices and the need of performing accurate prediction of price fluctuations, the solution has largely shifted from statistical methods to machine learning area. However, the selection of proper machine learning techniques for automated agriculture commodity price prediction still has limited consideration. On the other hand, when implementing machine learning techniques, finding a suitable model with optimal parameters for global solution, nonlinearity and avoiding curse of dimensionality are still biggest challenges. In this research, we address these problems by conducting a machine learning strategy study and propose a web-based automated system to predict agriculture commodity price. In the two series experiments, five popular machine learning algorithms, ARIMA, SVR, Prophet, XGBoost and LSTM have been compared with large historical datasets in Malaysia. The results validate the efficiency of the proposed Long Short-Term Memory Model (LSTM) to serve as the prediction engine for the proposed system. Particularly in the long-term experiment testing, the average performance of LSTM with MSE has improved 45.5% while ARIMA has dropped 74.1% and the average MSE of LSTM is 0.304 which outperformed all other four algorithms.

1. Introduction

The increasing availability of large amounts of agricultural commodity prices historical data and the need of performing accurate predicting of price fluctuations in agricultural economy demands the definition of robust and efficient techniques, which are able to infer from current observations. The traditional way to solve the prediction problem lies in linear statistical methods (such as ARIMA models), and more recently with the emergence of machine learning techniques, the solution has largely shifted from statistical methods to machine learning area. However, the selection of proper machine learning techniques for automated agriculture commodity price prediction still has limited consideration. On the other hand, when implementing machine learning techniques, finding optimal parameters of learning algorithm for global solution, nonlinearity and avoiding curse of dimensionality are still biggest challenges, therefore machine learning strategies studies are needed.

In practice, volatility in price of agricultural commodities is often unpredictable as they are affected by eventualities for example fluctuations of oil price, greenhouse effects and natural disasters such as flood or attacks by disease. Uncertainties of agricultural commodities price endanger the accessibility of food by consumers which leads to food insecurity and causes starvation and malnutrition. Instability of agricultural commodities price due to oversupply or lack of demand causes unnecessary food wastage. In recent years, the price of some agricultural commodities (such as palm oil and rubber) has been steadily decreased. This downward trend is making an impact to Malaysia's economy and can contribute to a slower economic growth in various investments in Malaysia. There are many exogenous factors that could cause such a trend and time-series data analysis is required to forecast this trend to improve Malaysia's agricultural plantation plan for better country development. Being able to anticipate the fluctuations and patterns in agricultural commodities price will enable the government to propose new policies that can help prevent the country into worse economy state. Further, the agricultural commodities providers are able to control their supply based on the time series analysis in order to prevent a bad plantation plan.

Most of the time series analysis on agricultural commodities are based on the US and China commodities market and there is no concrete research done about the Malaysia commodities

*Corresponding Author: Zhiyuan Chen,Email: zhiyuan.chen@nottingham.edu.my

market. Therefore, thorough agricultural commodities prices analysis should be performed based on the Malaysia prices.

In this research, we address these problems by conducting a machine learning strategy study and propose a web-based automated system to predict agriculture commodity price.

The main contribution of our work is summarized as follows:

- We formulate prediction of agriculture commodity price as a machine learning problem, which is more accurate, robust and efficient with the increasing availability of large amounts of agricultural commodity prices historical data than the traditional statistical methods.
- The proposed web-based automated agriculture commodity price prediction system with the best performance machine learning model has been presented as an advanced solution to fulfill the need of performing accurate predicting of price fluctuations in agricultural economy demands.
- Five popular machine learning algorithms, ARIMA, SVR, Prophet, XGBoost and LSTM have been compared with large historical datasets in Malaysia, which could serve as the foundation for other Malaysia commodities market research.

This paper is an extension of work originally presented in ITCC 2020 [1].

2. Background

In the earliest studies, the focus of analysis is on commodity futures markets. William G et al. discovered that futures prices were good price predictors in the corn, soybean and potato market established from empirical forecast assessment [2]. Many other subsequent studies on different agricultural futures markets in 1970's was the result of high dependency used by the researchers, who were interested in specific market conditions and traditional econometric models [3]. However, there are two obvious problems when implementing the traditional econometric models in the earliest days. First is the limited analysis due to the capacity of the model and the high computational cost [4], [5]. Second is those models are estimated through assumption that variables are independent, normal distribution, which is unrealistic in the real world [6].

Later on, several research studies have been proposed to implement agriculture price prediction scheme using different machine learning algorithms [7]-[10]. However, the performance of machine learning in agricultural commodity futures prediction is rarely explored. Future forecasting is usually done by analysing past price data of each commodity, climate, location, planting area, and several other conditions. Therefore, it is challenging because of its inherent complexity and dynamism. According to the lowest error percentage, many researchers have selected ANN and PLS as prediction algorithms.

Table 1: Summary of Related Work

Article	Agriculture Commodities	Technique
Tomek et al. [2]	Corn, Soybean, Potato	Empirical Forecast Assessment: Linear Regression
Kofi et al. [3]	Wheat, Potato	Traditional Econometric Model: Linear Regression
Sariannidis et al. [4]	Rice	Statistical Models: Autoregressive Conditional Heteroskedasticity (ARC), Generalized Autoregressive Conditional Heteroskedasticity (GARC)
Zulauf et al. [5]	Corn, Soybean	Econometric models: Price-level and Percent-change Models
Onour et al. [6]	Wheat, Rice, Beef, Groundnut, Sugar, coffee	Statistical Models: ARC/GARC models, Stochastic Volatility (SV) models.
Xiong et al. [7]	Cotton, Corn	Vector Error Correction model (VECM) – multi-output support vector regression (MSVR)
Peng et al. [8]	Cabbage, Bok choy, Watermelon, Cauliflower	Autoregressive Integrated Moving Average (ARIMA), Partial Least Square (PLS), and Artificial Neural Network (ANN)
Kumar et al. [9]	Sugarcane	K-Nearest Neighbor, Support Vector Machine (SVM), Least Squared Support Vector Machine
Manjula et al. [10]	--	Optimal Neural Network classifier (ONN)
Cao et al. [11]	--	SVM, RBF neural network
Connor et al. [12]	--	Recurrent Neural Networks
Dasgupta et al. [13]	--	Gaussian Dynamic Boltzmann Machine (DyBM) model with a recurrent neural network (RNN)
Tang et al. [14]	--	Feed Forward, Backpropagation Neural Network models, Standard Box-Jenkins model
Namaki et al. [15]	--	Deep Neural Network
Siami-Namini et al. [16]	--	Long Short-Term Memory (LSTM), Autoregressive Integrated Moving Average (ARIMA)
Hochreiter et al. [17]	--	Long Short-Term Memory (LSTM)
Chen [1]	Chicken, Chili, Tomato	ARIMA, SVR, Prophet, XGBoost and LSTM

In [9], the author proposed a system to apply prediction by analyzing past soil and rainfall datasets. In another research paper, Askunuri Manjula [10] has done crop prediction using multiple features, such as weather forecasting, pesticides and fertilizers and past revenue. Both of these two research works have been done the prediction with implementing pre-processing and feature reduction functions.

A few research studies have focused on implementing neural networks for prediction [11]-[13]. However most of these works are mainly about interval prediction and the point forecasting of agricultural commodity prediction has been taken less notice.

Further on, because of gradient vanishing, these existing methods fail to capture very long-term information [11]. Next, the dynamic dependencies among multiple variables are not being taken into consideration [12]. Besides that, these studies fall short in distinguishing a mixture of short-term and long-term repeating patterns explicitly [13].

During the recent years, research works of the promising deep neural networks algorithms in time series forecasting can be classified into three categories [14]. First category is to identify statistically significant events; the second is to find and predict inherent structure and the third is to do accurate prediction on numerical value [15]. Looking into time series prediction through deep neural networks, the most popular approach is Long Short Term Memory (LSTM). This approach has been highly discussed due to its promising result and the capability of not only modelling nonlinear patterns, realizing complex causal relationships, as well as the learning rate on huge historical datasets. The LSTM model is said to be more accurate than the prediction of the ARIMA model by 85% on average from the results obtained in the research by Namini, S. S, Tavakoli, N. and Namin, A. S [16]. Furthermore, LSTM were introduced and aimed for a better performance by tackling the vanishing gradient issue that recurrent networks would suffer when dealing with long data sequences [17].

Other advance technologies, such as Box–Jenkins approach [18] and irrigation monitoring systems based on IoT technologies, GPS and LoRaWAN [19] have been presented by researchers recently. However due to the application in different domain or system complexity, we will not consider these techniques in this paper but it might be useful for our future work.

A summary of all these related works can be found in Table 1. Other background information and technical review of ARIMA, SVR, Prophet, XGBoost and LSTM have been discussed in our conference paper [1].

3. Proposed System

The main objective of the proposed automated agriculture commodity price prediction system is to assist government or farmers for a better agricultural plantation plan. To achieve the objective, this research utilizes machine learning techniques to provide an agriculture price forecasting feature into the web system.

The project studies the needs of the farmers in doing agricultural activities, and these studies have been adapted into the system and be delivered in a simpler, more comprehensive way to suffice the farmers' knowledge in doing agricultural activities.

3.1. System Requirement Specification

3.1.1. Functional Requirements

The system should have a web app which consists of:

- A sign up and login page
- A forecast page to show all graph and data
- A commodity information page
- A user profile page
- An enquiry page

The system should forecast the future prices of agriculture commodities

- The forecast should be shown in a graph form
- User should be able to choose duration of prediction
- User should be able to rescale the x-axis of the graph
- The forecast should be based on previously available data
- User shall be able to access different type of model in price forecast (Univariate / multivariate)
- Graph should be updated upon type of commodity, duration of view, forecast period, and type of model

Visualization of forecast result should be clear for users

- Past prices and forecast results shall be separated with different color in one graph
- When the cursor hovers above the graph line, price information of the x-axis value shall be shown to users

Users shall able to select interested commodity

- Commodity chosen can only be pre-existing in the database
- System should update graph to show new data point based on new data added
- User shall be able to access different model type for each commodity (univariate, multivariate)

System should store and retrieve data in a database

- Developers should be able to update the database with new data
- Developers should be able to edit previous data stored in the database

Commodities data shall be downloadable

- User shall be able to download complete past prices of selected commodities in .csv file

A user profile shall be created automatically upon registration

- User shall be able to edit the user profile and save it

System shall be able to authenticate existing user

- User account shall be saved in database during registration

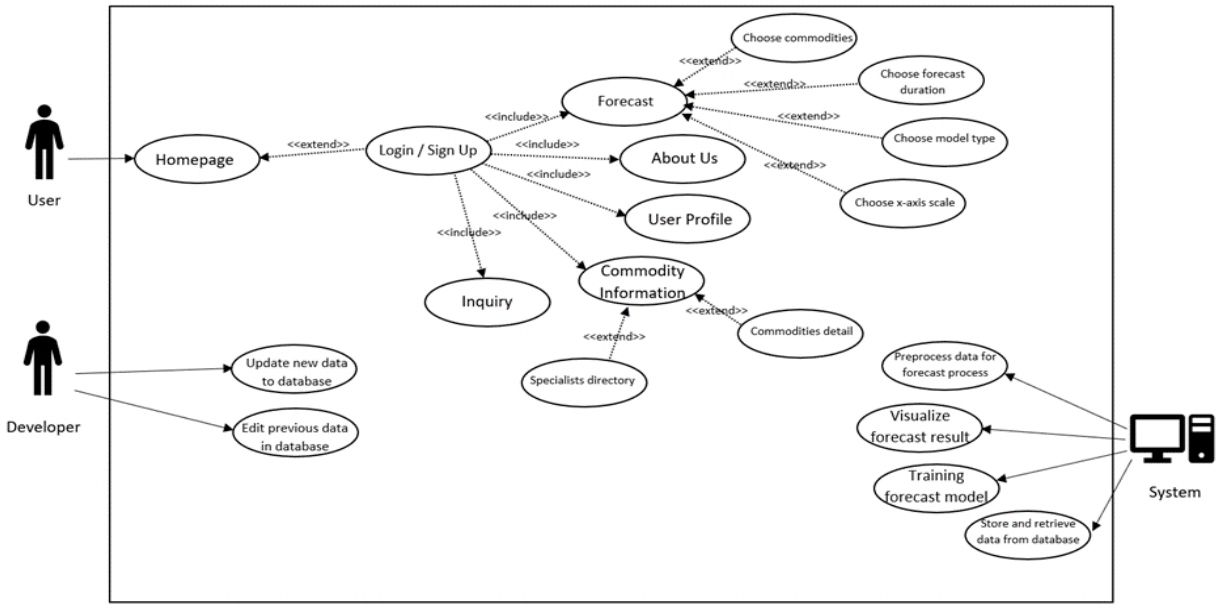
3.1.2. Non-Functional Requirements

Reliability

- The system shall be able to access the system anywhere with an internet connection
- The system shall be able to handle user queries not exceed 30 seconds when Tensorflow backend is running

Usability

- The system shall be simple to access for a first-time user
- The redirecting page amount for any feature of the system shall not exceed 5 pages



Performance

- The forecasting result shall be available for at least 7 commodities
- For each commodity, univariate model and multivariate model for forecast process shall be able to be accessed
- The meantime of download a file in csv format from a software shall not exceed 10 seconds

Security

- The system would require user account to access to all features
- User shall register their account with their email address one time only, system will reject account registration of existing account

Commodity products

- Commodities listed in the software shall be in English
- All commodities prices in the software shall be in Ringgit Malaysia, and local pricing for every commodities

Interface

- The system shall be portable in devices including personal computers, iPad and any mobile devices with Google Chrome software installed.
- The system layout shall be responsive to both full-size window and minimized window

3.2. Use Case Diagram

The use case diagram is shown in Figure 1. It explains the interactions that occur between the users and the system itself.

3.3. System Overview

Figure 2 describes the system Flow. The web system follows the Model-View-Controller (MVC) architecture. When a user enters a URL in their browser, the browser sends an http request to the web server, then the web server forwards the request to the application server and the URL setting contained in the urls.py file selects the view according to the url specified in the request, the view communicates with the database via models.py, renders the html or other format using templates (loads the static files) and returns the http response to the web server and finally, the web server provides the desired page to the browser.

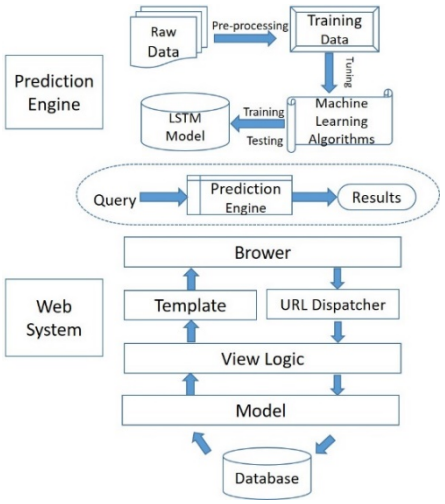


Figure 2: System Flow of the Automated Agriculture Commodity Price Prediction System

There are four processes in the prediction engine design of the proposed system: data pre-processing process, tuning process, training and testing process and decision-making process. Process 1 involves cleaning and reformatting the dataset; while process 2 focuses on defining optimal parameter value of each machine learning techniques. All machine learning models will be trained and tested in process 3. Five machine learning methods were compared in this research, which are ARIMA [20], SVR [21],

Prophet [22], XGBoost [23] and LSTM [17]. Technical details of these methods have been presented in our conference paper [1]. Finally, process 4 selects the best model out of all models to serve as the prediction engine. An algorithm snippet of the prediction engine is shown in Figure 3.

```
# Part 3 - Make the predictions and visualise the results
data_test = data.tail(52)
data_test = data_test[:]['Harga Ladang']
X_test = []
X_test.append(data_train_scaled[(len(data_train_scaled)-52):])


X_test = np.array(X_test)
X_test = X_test.reshape(X_test.shape[0], X_test.shape[1], 1)

predicted_price = regressor.predict(X_test)
predicted_price = sc.inverse_transform(predicted_price)
predicted_price = predicted_price.reshape(-1 ,1)
```

3.5. Implementation Practice

The main implementation of the application for machine learning algorithms to predict is Python programming language. It is an interpreted, high-level, general-purpose programming language used widely for machine learning. For the frontend design implementation of the application, Hypertext Markup Language (HTML), Cascading Style Sheets (CSS) and JavaScript have been used. These are high-level languages widely used to build a design and features of web applications.

3.5.1. Collaborative Software and Version Control

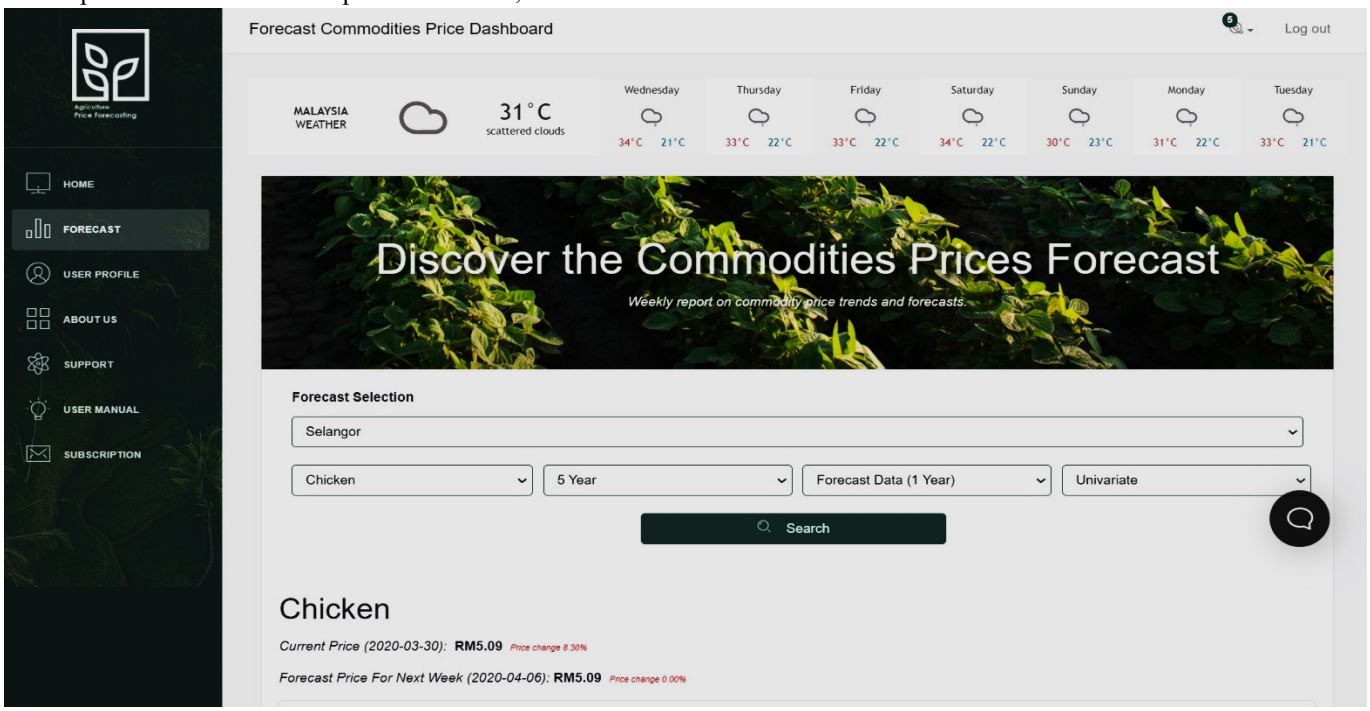
Django is the main framework used to build our web application, which is Python-based software that follows the model-template-view architectural pattern. Besides, a free cloud service which supports free GPU named Google Colab is in service to assist while implementing Long short-term memory (LSTM) algorithms in python programming language and it also supports the development for deep learning applications using popular libraries such as TensorFlow, Keras, PyTorch and so on.

Github, an open source version control website, is used as a centralized repository to keep, share, update codes and most importantly is to do version control. This platform is used as our service-oriented architecture for the application, also to manage the codebased, deliver the latest code and manage configuration of our web application.

3.5.2. Graphical User Interface (GUI)

During the initial implementation to achieve our prototype user interface (UI), we have used Hypertext Markup Language (HTML) to build the main structure of the homepage, login page and the forecasting dashboard. Thus, we modify using Cascading Style Sheet (CSS) and JavaScript to implement our proposed design accordingly.

In the latest design, there are 8 pages that build up the website, particularly, register, login, dashboard, user profile, support, weather forecast, about us and subscription page. Firstly, the user will be required to have an account in order to login the page, this will be handled in the register page for the user to create an account. Secondly, after getting an account, the user will be redirected to the login page with a simple input form for the user to log in with their account. All of the pages will require the user to login in order to view it. After logging in, there will be a side navigation bar containing each Uniform Resource Locator (URL) for relevant pages. In the forecast page as shown in Figure 4, users will also be able to select relevant commodities, data types, and duration of forecast they would prefer to see in the graph. The user profile will handle all the information that belongs to the user and



users will be able to edit and update their profiles. In the support page, users may submit their enquiries via the form provided. The about us page is a brief introduction regarding the website and its purpose and usage. In the subscription page, the user will be able to subscribe to different plans to unlock different features. Lastly, each page contains the identical top navigation bar for them to log out anytime.

4. Experiments and Results

The forecasting feature of the prediction system is the main focus of the project. In the decision of technique implemented to serve as the prediction engine, a series of experiments have been designed to discover the most optimal algorithm which should perform with highest accuracy and best capability of handling increasing data.

Based on our literature review analysis, five algorithms namely, ARIMA, SVR, Prophet, XGBoost and Long short-term memory have been selected as our potential prediction engine and their performance will be compared to learn the best model of interest from large datasets.

4.1. Datasets

Our experimental agriculture datasets are extracted from the report made by FAMA official government website (https://sdvi.fama.gov.my/analisahargamingguan/DownloadPassword.asp). This website consists of weekly data analysis reports from 2007-2020. The price datasets of different categories (December 2008 to March 2020) are selected and migrated into a raw dataset. Two series of experiments have been conducted, one is to consider only time-series dataset with 11 years' datasets (December 2008 to March 2019) and the other is with multivariable (Temperature, Humidity, Precipitation and Crude Oil Price) with the whole dataset.

Table 2 shows the statistics of the price with three selected commodities from two categories, chicken for poultry, chili and tomato for fruits. All three commodities datasets with 2% missing values.

Table 2: Price Statistics of the Raw Data

Commodities	Mean	Minimum	Maximum	StdDev
Chicken	4.84	3.50	6.25	0.52
Chili	5.92	2.90	12	1.55
Tomato	2.19	0.50	6.35	0.83

4.2. Experimental Setting

4.2.1. Configuration of ARIMA

ARIMA is a combination of two models that is Auto regressive (AR) and moving average (MA). ARIMA consists of three parameters (p,d,q), where: p is the number of autoregressive terms, d is the number of nonseasonal differences needed for stationarity, and q is the number of lagged forecast errors in the prediction equation [20]. A series of actions have been performed to preprocess the data. Firstly, an Augmented Dickey Fuller Test (ADCF) is performed to determine whether the time series is stationary and perform data transformation to assure the stationarity of the data [22]. And then a differencing method is used to transform a non-stationary time series into a stationary one, a lag of 1 has been given as the result shows that the mean and variance are more constant over time in comparison to before transformation, thus resulting in ARIMA (d,1).

AR demonstrates the correlation between the previous time period with the current. Thus, in order to predict the value of P, a partial auto-correlation function (PACF) graph will plot. This plot provides a summary of the relationship between an observation in a time series with observations at prior time steps with the relationships of intervening observations removed [25].

Inevitably, there is always noise or irregularity attached in a time series. Hence, MA has been implemented and the main purpose is to figure and average out the noises that could potentially affect the model. This has been achieved by plotting an Auto Correlation Function (ACF) autocorrelation plot. An ACF plot is an (complete) auto-correlation function which gives values of auto-correlation of any series with its lagged values, it describes how well the present value of the series is related with its past values [24][26]. From our experimental results, it is observed that the diagram shows that the graph cuts off and drops to zero for the first time on x-axis in 1 on ACF plot and 1.5 on ACF plot, thus, giving q=1 and p=1.5.

4.2.2. Configuration of SVR

To capture the non-linear pattern of the data prediction, the implemented SVR model has used Radial Basis Function as the kernel function. As a popular function used in most kernelling algorithm, it was claimed of having well performance under general smoothness assumptions [27] and this suits the situation in our study. All parameter settings remain default, however, some parameter values have been tuned in building the model, such as:

- The kernel coefficient (gamma) setting value is set as 1/total number of features to deal with various amounts of features.
- Regulation parameter(C) is set by fine-tuning to find the most suitable value for the model.
- The epsilon setting is used with default value.

To resolve the missing values, two different approaches have been implemented to the model:

- The missing value in the initial raw data - fill the missing value with valid, but not influence value (e.g. -99999) to avoid clearance of missing data affecting the prediction process.
- The missing value in the post-process raw data - clear the missing data to maintain the same amount of row in the data frame.

Training of the model has utilized the python sklearn function called train_test_split. The dataset is randomly split into the corresponding train set and test set, in this case, the ratio of training set and test set is 9:1.

4.2.3. Configuration of Prophet

The default setting of Prophet:

- growth ('linear')
- n.changepoints (n)
- changepoints.range (0.8)

- changepoint_prior_scale (m)

The number of change points n has been set as one per month. The changepoint_prior_scale is to decide how flexible the changepoints are allowed. To avoid the overfitting problem, m has to set to [10,30]. Prophet is relatively robust to missing data, so no resampling methods have been implemented. Cross-validation method has been implemented to split the date into training and testing sets.

4.2.4. Configuration of XGBoost

The default setting of XGBoost:

- objective='reg:linear',
- colsample_bytree=0.8,
- learning_rate=0.1,
- max_depth=8,
- n_estimators=1000,
- silent=1,
- subsample=0.8,
- scale_pos_weight=1,
- seed=27
- early_stopping_rounds=50

4.2.5. Configuration of LSTM

The default setting of LSTM:

- Number of epochs: 100
- Batch size: 10
- Layers: 4 LSTM layer (including input layer) each of size 50 unit and subsequent by a dropout layer of +/-0.2 based on the dataset after each layer. 1 Output dense layer of size 52.
- Optimizer and loss function: adam, mse.
- Activation: hyperbolic tangent
- Data transform: data are scaled by a minmaxscaler

4.3. Results

The final mean squared error score (MSE) for the prediction in the first series of experiments has been shown in Figure 5. The mean squared error score gives a better justification of the optimal algorithm that could be chosen as the prediction engine. From the results, ARIMA has shown as the most stable model amongst the three commodities, as it has the lowest mean square error value (0.251) on average, particularly with Chili, it gets as low as 0.027 for MSE.

From the literature review, LSTM is more superior than other machine learning models such as SVR and ARIMA in terms of accuracy and difficulties in handling the data. However, in the first series of experiments, due to the small sample size, it was not able to perform outstanding predictions to fit the system. We can conclude ARIMA has outstanding ability in handling small amounts of data in performing predictions.

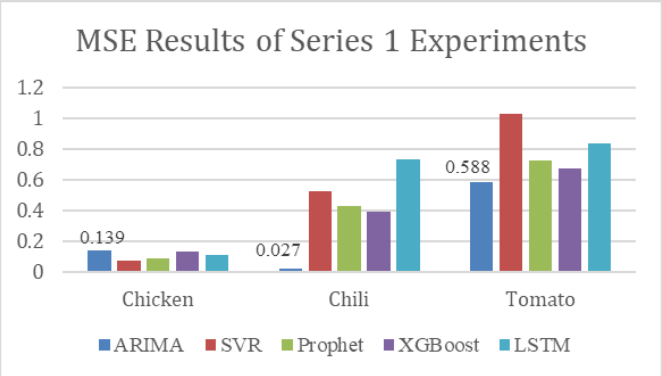


Figure 5: MSE Results of Series 1 Experiments

In order to test the longer term of agriculture commodity price prediction, which is more approximate to the real world situation, a second series of experiments have been conducted. In the second attempt of the experiment, the data has been enlarged to March 2020, and the feature space has been increased from one dimension to multi-dimensions with Temperature, Humidity, Precipitation and Crude Oil Price features being added to the model training process. Figure 6 shows the latest results of the second series of experiments.

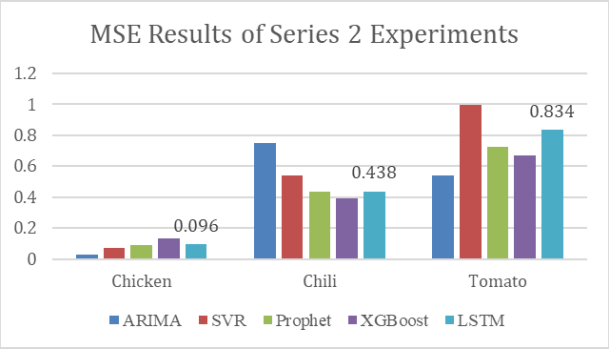
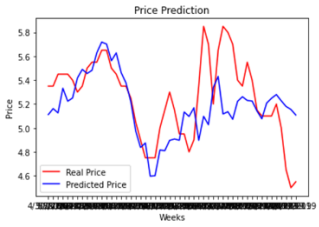


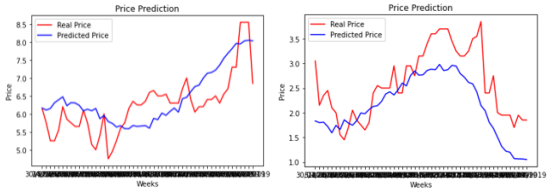
Figure 6: MSE Results of Series 2 Experiments

From the results, the average performance of LSTM with MSE has improved 45.5% while ARIMA has dropped 74.1%. The average MSE of LSTM is 0.304 which outperformed all other four algorithms. The improvement of the LSTM model shows its potential in handling increasing data and higher complexity of data compared to other models and in a long-term overview. Considering the fact that the collected data will continuously increase in terms of size and complexity, hence, LSTM is a better choice comparing to ARIMA (which shows the best result in the first series of experiment). Meanwhile LSTM's outperform capability of handling bigger dataset, as well as in the terms of system maintainability and scalability have fully justified the suitability to serve as the prediction engine for the proposed automated agriculture commodity price prediction system.

Figure 7 (a) to (c) are the graphs that illustrate the original prices, as well as the prediction prices using LSTM algorithm for 3 experimental commodities (Chicken, Chilli and tomato). The graph can help to visualise the outcome of predicted price in order to make a visual presentation to users.



(a) Chicken Price Prediction



(b) Chili Price Prediction (c) Tomato Price Prediction

Figure 7: LSTM Prediction Results

5. Conclusion

In this research, we study the challenges of prediction on price fluctuations in agricultural economy with increasing large amounts of agricultural commodity prices historical data. An accurate, robust and efficient agriculture commodity price prediction system with novel machine learning techniques have been proposed, which is able to automatically infer from current observations. The traditional way to solve the prediction problem lies in linear statistical methods (such as ARIMA models), and in this paper we present the novel and complete solution in machine learning area.

Firstly, the prediction of agriculture commodity price has been formulated as a machine learning problem. It is more accurate, robust and efficient for the increasing availability of large amounts of agricultural commodity prices historical data than the traditional statistical methods. Secondly the proposed web-based automated agriculture commodity price prediction system with the best performance machine learning model engine has been presented as an advanced solution for the need of performing accurate predicting of price fluctuations in agricultural economy demands. Thirdly five popular machine learning algorithms, ARIMA, SVR, Prophet, XGBoost and LSTM have been compared with large historical agriculture commodity price datasets in Malaysia, which could serve as the foundation for other Malaysia commodities market research.

In future, correlation analysis between different supporting factors and the historical price will be investigated in depth and the optimization of the proposed prediction engine model will be further improved. The proposed system will also enhance its trading aspect in near future, by reporting more in-depth location-specific price analysis to realize trading among farmers.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is part of a research funded by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme (FRGS/1/2018/ICT02/UNIM/02/1).

References

[1] Z.Y. Chen, K.L.Sin, "Long Short-Term Memory Model Based Agriculture Commodity Price Prediction Application," in Proceedings of the 2020 2nd International Conference on Information Technology and Computer Communications, Association for Computing Machinery, New York, NY, USA: 43–49, 2020, doi: https://doi.org/10.1145/3417473.3417481.

[2] W.G. Tomek, R.W. Gray, "Temporal Relationships Among Prices on Commodity Futures Markets: Their Allocative and Stabilizing Roles," American Journal of Agricultural Economics, **52**(3), 372–380, 1970, doi:https://doi.org/10.2307/1237388.

[3] T.A. Kofi, "A Framework for Comparing the Efficiency of Futures Markets," American Journal of Agricultural Economics, **55**(4_Part_1), 584–594, 1973, doi:https://doi.org/10.2307/1238343.

[4] N. Sariannidis, E. Zafeiriou, "The spillover effect of financial factors on the inferior rice market," Journal of Food, Agriculture & Environment, **9**(1), 336–341, 2011, doi:https://doi.org/10.1234/4.2011.1962.

[5] C.R. Zulauf, S.H. Irwin, J.E. Ropp, A.J. Sberna, "A reappraisal of the forecasting performance of corn and soybean new crop futures," Journal of Futures Markets, **19**(5), 603618, 1999, doi:https://doi.org/10.1002/(SICI)1096-9934(199908)19:5<603::AID-FUT6>3.0.CO;2-U.

[6] I. Onour, B.S. Sergi, "Modeling and Forecasting Volatility in the Global Food Commodity Prices (Modelování a Prognózování Volatility Globálních cen Potravinářských Komodit)," Agricultural Economics-Czech, **57**(3), 132–139, 2011, doi:https://doi.org/10.17221/28/2010-AGRICECON.

[7] T. Xiong, C. Li, Y. Bao, Z. Hu, L. Zhang, "A combination method for interval forecasting of agricultural commodity futures prices," Knowledge-Based Systems,77,92–102, 2015, doi:https://doi.org/10.1016/j.knosys.2015.01.002.

[8] Y.-H. Peng, C.-S. Hsu, P.-C. Huang, "Developing crop price forecasting service using open data from Taiwan markets," in 2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI), 172–175, 2015, doi: https://doi.org/10.1109/TAAI.2015.7407108.

[9] A. Kumar, N. Kumar, V. Vats, "Efficient crop yield prediction using machine learning algorithms," International Research Journal of Engineering and Technology,5(06),3151–3159,2018. Online Access.

[10] A. Manjula, G. Narsimha, "Crop Yield prediction with aid of optimal neural network in spatial data mining: New approaches," International Journal of Information and Computation Technology, **1**(6), 25–33, 2016. Online Access.

[11] L.J. Cao, F.E.H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," IEEE Transactions on Neural Networks, **14**(6), 1506–1518, 2003, doi:https://doi.org/10.1109/TNN.2003.820556.

[12] J. Connor, L.E. Atlas, D.R. Martin, "Recurrent networks and NARMA modeling," in Advances in neural information processing systems, 301–308, 1992, doi:https://dl.acm.org/doi/10.5555/2986916.2986953.

[13] S. Dasgupta, T. Osogami, "Nonlinear Dynamic Boltzmann Machines for Time-Series Prediction," Proceedings of the AAAI Conference on Artificial Intelligence,**31**(1),2017,doi:https://dl.acm.org/doi/10.5555/3298483.3298506.

[14] Z. Tang, C. de Almeida, P.A. Fishwick, "Time series forecasting using neural networks vs. Box- Jenkins methodology," SIMULATION, **57**(5), 303–310, 1991, doi:https://doi.org/10.1177/003754979105700508.

[15] M.H. Namaki, P. Lin, Y. Wu, "Event pattern discovery by keywords in graph streams," in 2017 IEEE International Conference on Big Data (Big Data), 982–987, 2017, doi: https://doi.org/10.1109/BigData.2017.8258019.

[16] S. Siami-Namini, N. Tavakoli, A. Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 1394–1401, 2018, doi: https://doi.org/10.1109/ICMLA.2018.00227.

[17] S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory," Neural Computation,**9**(8),1735-1780,1997, doi:https://doi.org/10.1162/neco.1997.9.8.1735.

[18] A.M. Ashik, K.S.Kannan, "Time Series Model for Stock Price Forecasting in India", in Logistics, Supply Chain and Financial Predictive Analytics. Asset Analytics (Performance and Safety Management). Springer, Singapore, 2018, https://doi.org/10.1007/978-981-13-0872-7_17.

[19] D. M. Matilla, Á. L. Murciego, D. M. Jiménez Bravo, A. Sales Mendes and V. R. Quietinho Leithardt, "Low cost center pivot irrigation monitoring systems based on IoT and LoRaWAN technologies," 2020 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor), 2020, 262-267, doi: 10.1109/MetroAgriFor50201.2020.9277548.

[20] G.E.P. Box, G. Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day, Inc., USA, 1990, doi:https://dl.acm.org/doi/10.5555/574978.
[21] M. Awad, R. Khanna, Support vector regression, Springer: 67–80, 2015, doi:https://doi.org/10.1007/978-1-4302-5990-9_4.
[22] S.J. Taylor, B. Letham, "Forecasting at Scale," The American Statistician, 72(1), 37–45, 2018, doi:https://doi.org/10.1080/00031305.2017.1380080.
[23] T. Chen, C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA: 785–794, 2016, doi: https://doi.org/10.1145/2939672.2939785.
[24] Y. W. Cheung, K. S. Lai, "Lag order and critical values of the augmented Dickey–Fuller test", Journal of Business & Economic Statistics, **13**(3), 277-280, 1995,doi: https://doi.org/10.1080/07350015.1995.10524601.
[25] R. H. Shumway, D. S. Stoffer, Time series analysis and its applications: With R , Principles of fuel cells, Springer, 2006.
[26] G. E. P. Box, W. G. Hunter, J. S. Hunter, Statistics for experimenters: an introduction to design, data analysis, and model building, John Wiley and Sons, 1978.
[27] A. J. Smola, B. Schölkopf, K. R. Müller, "The connection between regularization operators and support vector kernels", Neural networks, **11**(4), 637-649, 1998, doi: 10.1016/S0893-6080(98)00032-X.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 385-394 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

A Scheduling Algorithm with RTiK+ for MIL-STD-1553B Based on Windows for Real-Time Operation System

Jong-Jin Kim, Sang-Gil Lee, Cheol-Hoon Lee*

Department of Computer Engineering, Chungnam National University, Daejeon, 305764, Korea

ARTICLE INFO

Article history:
Received: 01 July, 2021
Accepted: 09 August, 2021
Online: 26 August, 2021

Keywords:
RTX
RTiK+
APIC
Tablet PC
Inspection Equipment
Real-Time Operating System
Windows
MIL-STD-1553B
RM Scheduling Algorithm

ABSTRACT

In devices using Windows operating system based on x86 system, the real-time performance is not guaranteed by Windows. It is because Windows is not a real-time operating system. Users who develop applications in such a Windows environment generally use commercial solutions such as the RTX or the INtime to provide real-time performance to the system. However, when using functions and API for simple real-time processing, an issue of high development cost occurs in terms of cost-effectiveness.

In this paper, the RTiK+ was implemented in the type of a device driver by controlling the MSR_FSB_FREQ register to generate a timer interrupt independent of Windows in the Windows 8 and providing a real-time functionality to the user mode by re-setting the local APIC count register. And the operating frequency of the CPU is changed to minimize power consumption for battery life in a mobile device such as a Tablet PC.

In particular, the weapon system uses highly reliable MIL-STD-1553B communication and performs BC and RT functions of MIL-STD-1553B to transmit and/or receive data in communication between component and component. It is significantly importance to guarantee integrity of data without loss data during communication. For this purpose, it is proposed to implant the Scheduling algorithm with the RTiK+ for MIL-STD-1553B communication for Windows 8 to support a real-time processing in the Windows operating system on the embedded system, and to use the periods of 2ms (max), 5ms and 10ms provided by the RTiK+ for real-time processing when performing the BC and RT functions of MIL-STD-1553B communication. In this paper, the method of the scheduling algorithm with RTiK+ for MIL-STD-1553B to provide real-time processing is proposed for the Windows based on x86 system.

1. Introduction

1.1. Research Background

Recently, with the development of hardware and embedded systems, various mobile devices such as Tablet PC and smart phone used in daily life are being used instead of dedicated equipment for special purposes according to the user environment. Therefore, it is necessary to support an accurate real-time processing function in order to transmit/receive accurate commands or data to a mobile device without data loss of information from the target equipment. And this paper is an extension of work originally presented in Real-time Processing Method for Windows OS Using MSR_FSB_FREQ Control [1-5].

In general, a real-time system should be predictable in any situation, which means that time deterministic should be guaranteed. In this paper, to ensure time determinism, the MSR_FSB_FREQ register, which determines the operating frequency of the CPU to provide real-time processing performance in the Windows 8 environment, is controlled and the local APIC timer count register value is reset to operate a window independent timer. A study was conducted to guarantee the periodicity set by the user. Also, when checking the function and performance of a guided weapon system using a window-based inspection equipment in a guided weapon that uses fast MIL-STD-1553B communication such as 2ms between components, The RTiK+ is first installed to provide real-time performance to the Windows 8, and then it use the period provided by RTiK+ for scheduling algorithm to meet deadline in MIL-STD-1553B communication in guided weapons system, and also a study for miss rates to

*Corresponding Author: Prof. Cheol-Hoon Lee, Department of Computer Engineering, Chungnam National University, Korea, Email: clee@cnu.ac.kr

guarantee data integrity without loss is conducted through experimental tests to be proofed.

1.2. Research Purpose

In the case of a tablet PC, The Windows as operating software is mainly used as an operating system to provide compatibility with libraries that have been developed and operated, dependency on the device's operating system, and extensibility for various applications. To support a real-time processing function in the tablet PC, a commercial solution that is installed in addition to the Windows operating system installed in the tablet PC and provides a real-time functionality must be used. Such products include IntervalZero's RTX, which is currently widely used. The purchase price is very high, and when installed in equipment for mass production, royalty and maintenance costs are high, which causes a high development cost burden on developers. To improve this matter, The RTiK+ was designed to provide a real-time performance in Windows 8 based on x86 system.

In addition, the Windows operating system used for inspection equipment prioritizes compatibility, stability and reliability due to the characteristics of the weapon system, so older versions such as Windows 7 and Windows 8 are used rather than the latest versions such as Windows 10 [6]. For example, the flight body inspection equipment that recently developed an unmanned aerial vehicle developed in 2017 uses the Windows 7 released in 2009 [7].

In this paper, the RTiK+ is designed as a method by controlling the MSR_FSB_FREQ register to generate a window-independent timer interrupt in Windows 8 and providing a real-time period to the user mode by resetting the counter register of the local APIC, and the CPU operating frequency. A study how to provide a real-time performance by accessing the MSR_PKG_CST_CONFIG_ CONTROL register that determines the C-States so that the frequency controlling the C-States does not change.

A MIL-STD-1553B communication with high reliability is used mainly in the precision guided weapon system [8], and the proposed RTiK+ to provide a real-time processing function in the Windows operating system of the equipment is implanted in Windows 8, and it supports the BC and RT function of the MIL-STD-1553B to provide a real-time processing. The scheduling algorithm with RTiK+ supports a real-time period to the BC and RT function when performing communication.

Finally, an evaluation for experiment results is conducted to proof the miss rate that complies with the deadline in the set period in the MIL-STD-1553B communication which RTiK+ for a real-time processing is applied. The RTiK+ to provide a real-time processing is implanted in the Windows 8 based on x86 system, and the scheduling algorithm is applied to perform the MIL-STD-1553B communication. The integrity is verified using the RDTSC for measurement of the period of a real-time function and PASS 3200 for data monitoring of the MIL-STD-1553B, and the miss rate of the scheduling algorithm for the MIL-STD-1553B is analyzed. And to see if the RTiK+ can be used as a replacement item for the RTX, a real-time performance is measured and analyzed using the same experimental configuration for measuring performance of the RTX, showing that the RTiK+ can be used as a substitute for a third party.

The paper consists of chapters 1 to 6. Chapter 2 studies RTX, window scheduling, RM algorithm, and MIL-STD-1553B. Chapter 3 describes the real-time processing in the tablet PC proposed in this paper, and the real-time performance implementation using the MIL-STD-1553B communication scheduling algorithm. Chapter 4 describes the experimental environment, composition, and results. Chapter 5 analyzes the algorithm and results, Chapter 6 concludes by describing the summary of this paper and future research.

2. Related Research

2.1. RTX

IntervalZero released the RTX operating in kernel mode (ring 0) based on x86 computers in 1995. The RTX is software that provides a real-time performance to support a real-time function in Windows XP or Windows 7 but there is no solution for Windows 8 currently. The RTX provides a development environment that is easy to access and familiar to developers who use Windows based on a real-time operating systems, which are the advantages of Windows [9].

2.2. Local APIC

An APIC is an interrupt controller provided by the x86 architecture. It implements the function of delivering a hardware interrupt to the interrupt descriptor table with the address of the interrupt handler. A local APIC is an interrupt controller provided by the x86 architecture as an extension of the programmable interrupt controller [10-13].

2.3. Scheduling of Process and Thread in Windows

The Windows operating systems based on x86 system have processes and threads to process specific tasks. In general, the threads are divided into user-level threads and kernel-level threads. A kernel-level thread is used in Windows, and in Windows, a thread is assigned in a processor by a kernel scheduler and implements a scheduling. It is a scheduling algorithm that is scheduled according to the priority of all threads created in the operating system regardless of the priority of the process. The Windows scheduler uses a round-robin scheduling method. A process has six priorities, and a thread has seven priorities, and consists of a total of thirty-one priorities from 0 to 31 [14].

2.4. RM Scheduling Algorithm

The RM Scheduling, The RM means Rate-Monotonic, for a real-time system is a scheduling algorithm that gives the highest priority to the shortest period. The operating system to which this algorithm is applied is generally called preemptive and has a decisive characteristic for response time [15]. When there are n processes, the upper limit of total utilization on the system can calculate using the equation (1) [16]. If the total system utilization is less than or equal to the utilization U_{RM} calculated by the equation (1), it proves that scheduling is possible with the RM scheduling algorithm.

$$U_{RM}(n) = n\left(2^{\frac{1}{n}} - 1\right) \quad (1)$$

The RM scheduling algorithm is a method of fixed priority scheduling in which a fixed priority is given to each task, and when a task having a higher priority arrives, the currently executing task is preempted. The RM algorithm determines the priority based on the period of the task, and it is an algorithm that gives a high priority to a task with a short execution period [17-19].

2.5. C-States

It can command the CPU to enter a low-power mode to save battery when the CPU is idle. Each CPU has multiple power modes, collectively known as the C-States. In general, in the case of a mobile device such as tablet PC in the x86 architecture, the low-power technique provides five states of the CPU called the C-States as shown in Table 1 [20].

Table 1: C-States mode in Intel's CPU

Mode	Status of CPU
C0	Processor state: Full On
C1	Processor state: Auto-Halt
C1E	Processor state: Auto-Halt
C3	Processor state: Deep Sleep
C6	Processor state: Deep Power Down

2.6. MIL-STD-1553B

The MIL-STD-1553B, which is mainly applied to aircraft weapon systems and guided weapon systems, has a maximum transmission speed of 1Mbps, and it is a half-duplex and a serial communication method.

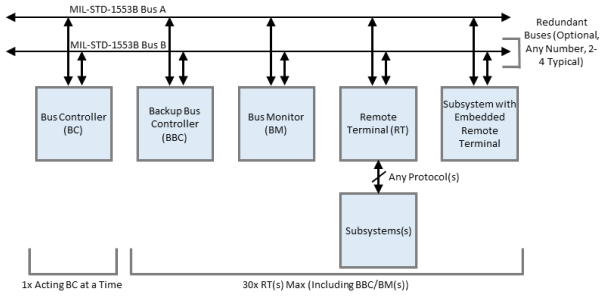


Figure 1: Interface Diagram of MIL-STD-1553B bus

Also, there is a restriction that the length of data that can transmit or receive data at a time is limited to a maximum of 32 words by a protocol. As shown in Figure 1, a bus controller (BC), a remote terminal (RT) and a monitoring (MT) are interfaced to the MIL-STD-1553B bus. The main function of the BC is to control the transmission and reception of all RTs on the bus. The RT has a function of transmitting or receiving data by responding to commands from the BC, and the MT has a function of monitoring all data moving on the bus [21, 22].

3. Methodology

3.1. Overview

In Windows 8, the initial count register (FFF0 0380h), as a local APIC timer register, is initialized to 0 after the operating system is booted. Therefore, since it is impossible to utilize a timer interrupt independent of the Window, there is a matter in providing a real-time processing function corresponding to the computer's system clock. In addition, The C-States are used to increase battery operation time by reducing power consumption of mobile device. The C-States function has the problem that the system clock is changed frequently flexibly, and an accurate period calculation for guarantying a real-time is not allowed [23, 24].

In this paper, in order to solve this issue, The RTiK+ is designed as shown in Figure 2, which uses a timer interrupt independent of the Windows to provide a real-time processing functions in the Windows 8 to support real-time performance for threads in the user mode. And the scheduling algorithm to support a real-time performance for MIL-STD-1553B communication is implemented based on RTiK+.

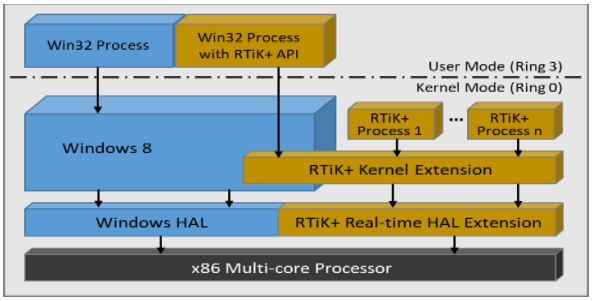


Figure 2: Architecture of RTiK+ for Real-Time Operating System

3.2. Local APIC Timer Control

As shown in Figure 2, The RTiK+ supplies a timer interrupt independent of the window through Kernel Resources access. And as shown in Figure 3, The RTiK+ is implemented to provide a real-time functionality in the Windows using the local APIC and kernel resources of the application processor [25, 26].

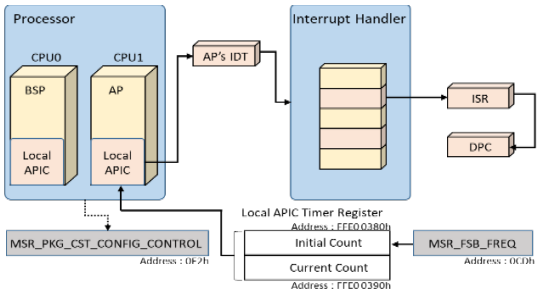


Figure 3: A Real-Time Processing Operating Process of the RTiK+

Before the timer interrupt is occurred, the initial counter register (FEE0 0380h) of the local APIC of the application processor sets the initial counter register value according to the value set in the MSR_FSB_FREQ register (0CDh) to generate a timer interrupt independent of the Windows. For such interrupt processing, a real-time task is created by registering an interrupt object in the interrupt descriptor table (IDT). In addition, The RTiK+ guarantees the periodicity of the real-time thread by using the MSR_PKG_CST_CONFIG_CONTROL register and minimizes the period error tolerance of the set task by controlling the C-States operation mode of the CPU to ensure the periodicity of the interrupt for thread processing [27-29].

3.3. Time Interrupt Control through FSB

As shown in Figure 4, x86-based system has two chipsets called the Northbridge and the Southbridge, the Northbridge is

between memory controller hub and the memory/graphic card slots, and the Southbridge is between I/O controller hub and PCI slots.

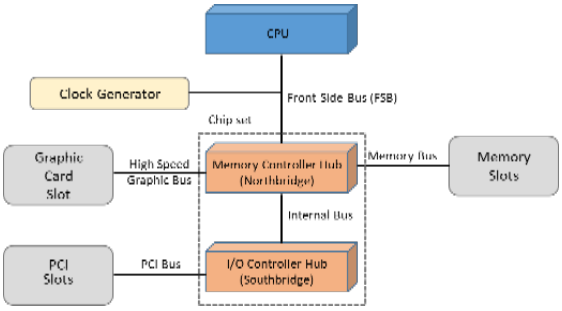


Figure 4: A Block Diagram of Chipset based on x86 Architecture [30]

A FSB also provides the ability to use a clock multiplier to determine the operating speed of the CPU and it provides synchronization by affecting the memory clock speed. The chipset block diagram of the x86 architecture is as shown in Figure 4, and the frequency generated by the clock generator is connected to the FSB and provided by the CPU and set as the initial counter register (FFF0 0360h) value of the timer register of the local APIC and then involves in determining the generation period of the timer interrupt. And in the case of the Windows 8, the operating frequency of the FSB according to the CPU workloads is provided through the lower three bits of D2:D0 of the MSR_FSB_FREQ register (0CDh), which is a hardware resource.

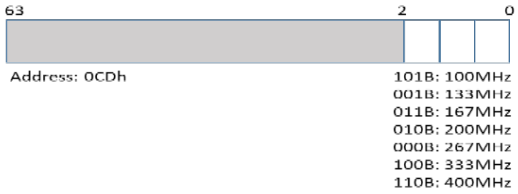


Figure 5: A MSR_FSB_FREQ Register

Therefore, as shown in Figure 5, to set the timer interrupt generation period of the local APIC to a fixed value, check the operating frequency of the currently executing FSB in the MSR_FSB_FREQ register, and set the local APIC timer to the value of the initial counter to generate an independent timer interrupt [31].

3.4. Time Deterministic through controlling C-States

Unlike the existing Windows 7, The Windows 8 provides a power saving function that reduces power consumption in real time by turning off power to some devices by adjusting the CPU clock low through access to the C-States for low-power control of the CPU. The following equation (2) is applied as the equation used to determine the CPU clock in Windows 8 [24, 32].

$$\text{CPU clock} = \text{FSB frequency} \times \text{Clock multiplier} \quad (2)$$

The FSB can control the power consumption by controlling the value of the clock multiplier by the Windows according to the hardware specifications. However, the method of adjusting the value of the clock multiplier to lower the clock of the CPU has a matter in that it is difficult to guarantee the time determinism of the timer interrupt of the local APIC. In addition, a problem occurs in that the aforementioned time determinism cannot be guaranteed because a delay time occurs due to the time when the CPU is activated from the idle state or sleep state.

In this paper, The C-States are controlled through the MSR_PKG_CST_CONFIG_CONTROL register (0E2h) provided by x86 system to guarantee the periodicity of the RTiK+. And each speed of the core clock according to the setting of the C-States is seemed as shown in Table 2. When the C-States are enabled, the clock is automatically adjusted based on the utilization rate the CPU is processing tasks. Also, when the C-States are disabled, the clock is kept at its maximum value regardless of the CPU utilization [33].

Table 2: A CPU Clock Status depending on the C-States

Type of CPU		C-States Enable	C-States Disable
Core 1	Standard	3199.86 MHz	3199.86 MHz
	Current	3199.86 MHz	3199.86 MHz
Core 2	Standard	1599.93 MHz	3199.86 MHz
	Current	1599.93 MHz	3199.86 MHz

3.5. Process in User Mode of the RTiK+

A study on the real-time processing function of the user mode provided in the existing Windows such as the Windows XP and the Windows 7 before the Windows 8 has a matter that the periodicity, data accuracy, and data integrity are not guaranteed because event signals are lost in an environment of multi-core processor. To solve this problem, the RTiK+ provides a real-time processing function in the user mode as shown in Figure 6.

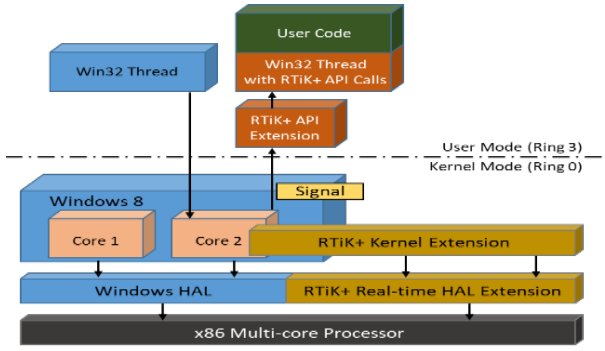


Figure 6: An Operation Process of the RTiK+ for a Real-Time Processing

The algorithm coded in the RTiK+ to set the priority of the process to the highest to guarantee data accuracy and data integrity is shown in Figure 7 [34]. It is implemented to call the Process_Affinity() function inside the real-time thread, and the real-time thread inside Process_Affinity(), by calling the SetProcessAffinityMask() function, an API provided by the Windows, the CPU is specially designated so that the real-time thread operates only on the core where RTiK+ operates.

Algorithm 1:

Result:

```
void Process_Affinity()
{
   SetPriorityClass(GetCurrentProcess(), REALTIME_PRIORITY_CLASS);
   SetProcessAffinityMask(GetCurrentProcess(), 0x02);
}
```

Figure 7: An Algorithm to set for Priority of Process

In addition, by calling the SetPriorityClass() function to set the priority of the real-time thread and process of RTiK+, the current real-time thread is designated as the highest level as a REALTIME_PRIORITY_CLASS, so that the CPU is not preempted by other Window's threads [35].

When the RTiK+ process is created, the process must be created by setting it to the REALTIME_PRIORITY_CLASS priority level, which is the highest priority. Figure 8 is the algorithm coded to provide a real-time processing function in the user mode of the RTiK+.

Algorithm 2:

Result:

```
void main()
{
    hThread = (HANDLE)_beginThreadex(NULL,0,ThreadFunction,NULL,0,(unsigned*)&dwThreadID);
    SetThreadPriority(hThread,THREAD_PRIORITY_TIME_CRITICAL);
}
```

Figure 8: An Algorithm to set for Priority of Thread

3.6. Priority-based Scheduling Algorithm

In the priority-based scheduling algorithm, a scheduling is implemented online, and when an event such as the creation or termination of a task occurs, the scheduling is performed first in the high-priority task. For this reason, the scheduling is performed while tasks are being performed, and there is an advantage of being able to respond more flexibly according to the state of the system, but on the other hand, it sometimes causes overhead in the operating system [36].

In this paper, the RM scheduling algorithm, a fixed-priority scheduling algorithm proposed by Liu and Layland, is applied to the RTiK+. The RM is an algorithm of giving a high priority to a task with a short period when all tasks are periodic tasks, the deadline coincides with the period, and the tasks are independent from each other. In this paper, the total utilization rate of the RM scheduling algorithm is U_{RM} (n) = n (2^(1/n)-1). If tasks are maximum 100, U_{RM} (100) = 0.69555. And no matter how many tasks are, the total utilization U_{RM} is within 0.69314.

3.7. A Process in User Mode of RTiK+

3.7.1. The BC Algorithm for a Real-Time Performance

In the MIL-STD-1553B, the BC function controls the direction of data when transmitting data through the bus, and transmits data to the RT by sending transmitter (Tx) and receiver (Rx) commands from the BC of the MIL-STD-1553B communication. In order to guarantee a real-time processing function of MIL-STD-1553B communication, it is implemented to operate the Tx and Rx commands in the real-time thread provided by the RTiK+. Figure 9 is the BC algorithm of MIL-STD-1553B communication that implements the Tx and Rx commands in the periodic thread of the RTiK+.

In this algorithm, Tx command releases the Rx link with the m1553_delete_link() function, connects the Tx link with the m1553_add_link_to_chain() function, sends the Tx command from the BC with the m1553_delete_link() function, and checks the data transmitted from the RT with m1553_read_link_data() function. And the Rx command releases the Tx Link with the m1553_delete_link() function, connects the RX Link with the m1553_load_chain() function, and transmits data to the RT with the m1553_load_chain() function.

Algorithm 3:

Result:

```
m1553_delete_link(device_number, RxLink.link_id, RxLink.chain_id);
status = m1553_add_link_to_chain( device_number, &TxLink );
if ( status == FAILURE )
{
    printf( "FAILURE\n" );
    printf( "%s\n", sbs_read_error() );
}

status = m1553_load_chain( device_number, 1 );
if ( status == FAILURE )
{
    printf( "FAILURE\n" );
    printf( "%s\n", sbs_read_error() );
}

uSleep(600);
m11553_read_link_data(device_number, TxLink.link_id, TxLink.buff_id[0], r_buffer);

m1553_delete_link(device_number, TxLink.link_id, TxLink.chain_id);
status = m1553_add_link_to_chain( device_number, &RxLink );
if ( status == FAILURE )
{
    printf( "FAILURE\n" );
    printf( "%s\n", sbs_read_error() );
}

m11553_write_link_data(device_number, RxLink.link_id, RxLink.buff_id[0], r_buffer);
status = m1553_load_chain( device_number, 1 );
if ( status == FAILURE )
{
    printf( "FAILURE\n" );
    printf( "%s\n", sbs_read_error() );
}
uSleep(600);
```

Figure 9: BC Algorithm for MIL-STD-1553B based on Real-Time Processing

3.7.2. The RT Algorithm for a Real-Time Performance

In the MIL-STD-1553B, the RT function transmits or receives data through the bus according to the BC commands. In order to check a real-time performance of the RTiK+, it is designed to store data in the buffer at the same period as the period set in the BC of the inspection equipment by the RTiK+ to send 32 words of the value increased by one (1) from the previous data. However, the matter of data loss due to the gap time for enabling the RTiK+ implanted in the BC and the RT function also occurs in the RT same as the BC. In order to solve this matter, a double buffer was used in the same way as the BC. Figure 10 shows the RT algorithm of the MIL-STD-1553B communication using double buffer.

Algorithm 4:

Result:

```
before_actptr_TX = actprt_TX;
actprt_TX = LL_get_ram( device_number, btrptr_TX + rt1sa2t.sa + 32);

If (before_actprt_TX != actptr_TX)
{
    if ( rt1sa2t.buff_id[0] == actprt_TX )
    {
        for (i=0; i<32, i++)
        {
            w_buffer[i] = (SBS16)( cnt );
        }

        m1553_write_sa_buffer( device_number, 32, rt1sa2t.buff_id[1], w_buffer);
        cnt++;
    }
    else if (rt1sa2t.buff_id[1] == actprt_TX
    {
        for (i=0, i<32, i++)
        {
            w_buffer[i] = (SBS16)( cnt );
        }

        m1553_write_sa_buffer( device_nmber, 32, rt1sa2t.buff_id[0], w_buffer);
        cnt++;
}
```

Figure 10: RT Algorithm for MIL-STD-1553B based on Real-Time Processing

3.7.3. Scheduling Algorithm of MIL-STD-1553B for a Real-Time Performance

When performing the MIL-STD-1553B communication, it takes time to activate the RTiK+ each time to utilize the RTiK+. A period error occurs in communication. Also, when a data transmission occurs in the BC due to the application of polling method applied in the paper and a data is received at the RT. In a state in which all data transmitted from the BC are not correctly received, the next data to be transmitted updates the buffer, resulting in a matter that correct data cannot be received at the RT. Therefore, in order to improve this problem, it is necessary to set the period of the RTiK+ for MIL-STD-1553B communication by considering the time it takes to transmit a data with the MIL-STD-1553B's BC command and the time it takes to receive data with the RT command. And a scheduling with RTiK+ for MIL-STD-1553B is required. Also, in the MIL-STD-1553 communication proposed in this paper, there is a problem in that it does not guarantee the periodic operation of data transmission from the BC to the RT because the Rx command is immediately sent after the Tx command is finished. In addition, since the RTiK+ operates at the period initially set, the period of the Tx command and the Rx command cannot be set respectively, so there can be a difference in the operating period of the RT and data loss can occur in this case. To solve this matter, when the Rx function is finished, the Rx Link is released, and the Tx Link is connected immediately to execute Tx function. The scheduling algorithm using the RTiK+ to guarantee a real-time processing performance when transmitting and receiving data in the MIL-STD-1553B communication was proposed.

This scheduling algorithm applied when performing the MIL-STD-1553B communication between the host PC and the target PC is assumed as follows.

- The RTiK+ uses the RM algorithm and preempts the CPU.
- The MIL-STD-1553B communication occurs periodically.
- Task period (Period, pi) and deadline (Deadline, Di) are the same. (pi = Di)
- The task start time (Release Time, Øi) is 0.
- The RTiK+ is implanted to both the host PC and target PC, and is activated only when the RTiK+ is utilized.
- The BC and the RT function are for MIL-STD-1553B, and interrupt cannot be used in an asynchronous method.
- The Tx and Rx command occur alternately. The data received by the Tx command is read from the buffer and transferred to the Rx command as it is.
- Transmits up to 32 words of data as in the Tx and Rx commands, and the transmission time is up to 0.80ms for all tasks [37].
- The Tx and Rx command cannot occur within a half of a period, therefore disconnect and create each time Tx to Rx transitions. Thus Tx(ei) > Rx(ei) or Tx(ei) < Rx(ei).
- The BC function and the RT function each use a double buffer.

Figure 11 shows the scheduling algorithm for the MIL-STD-1553B communication. The MIL-STD-1553B occurs when the Tx and Rx commands are alternated between the host PC and the target PC, and each task Ti (T0, T1, T2, T3, ..., Tn) has a maximum of 32 words of data. When transmitted, the execution time ei of each task is 0.80ms. At this time, since the deadline is the same as the period, it is scheduled so that transmission or reception of data within the period is normally completed without overhead. Also, when the MIL-STD-1553B communication is performed, RTiK+(pi), Tx(pi), Rx(pi), Buffer1(pi) and Buffer2(pi) are operated with an independent period each to match the period provided by the RTiK+.

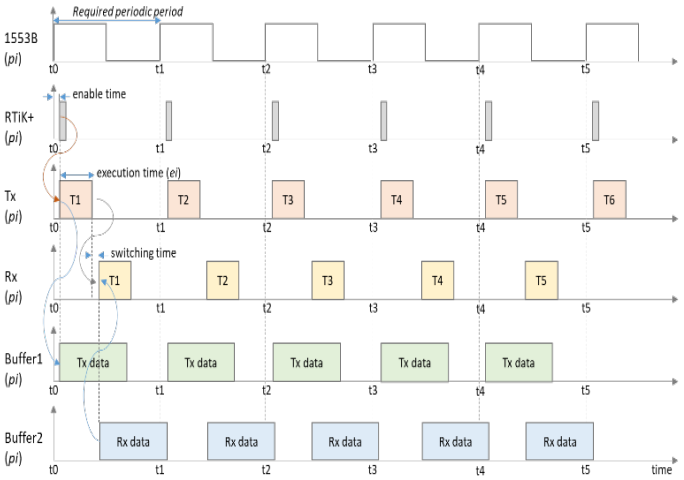


Figure 11: The Scheduling Algorithm with RTiK+ for MIL-STD-1553B

And if the task for the MIL-STD-1553B communication scheduling algorithm is as shown in Table 3, there are three Ti = (Øi, pi, ei, Di) tasks in the Tx and Rx commands linked to the RTiK+ pi period, verified if scheduling is possible through simulation that when the same 32 words of data transmission is occurred in the MIL-STD-1553B communication. That is, it was assumes that when this simulation is satisfied, all other communication conditions are also satisfied. At this time, when the Tx and Rx commands are executed, each task needs a maximum execution time of 0.80ms to transmit 32 words.

Table 3: A related Tasks of Tx/Rx for the MIL-STD-1553B Scheduling

Tasks	Release time (Øi)	Period (pi)	Execution time (ei)	Deadline (Di)
RTiK+ pi	0ms	2ms	0.01ms	2ms
Tx pi	0ms	2ms	0.80ms	2ms
Rx pi	0ms	2ms	0.80ms	2ms
T1	0ms	2ms	0.80ms	2ms
T2	0ms	2ms	0.80ms	2ms
T3	0ms	2ms	0.80ms	2ms

Using this method of the scheduling algorithm, each T1, T2, T3, …, Tn while Tx and Rx commands are being executed. It shows that the BC function of the MIL-STD-1553B can be implemented without data loss within the period set in the RTiK+ by the tasks. If the equation (3) is applied to calculate the system utilization rate for Tx(pi) and Rx(pi), the system utilization rate is 0.80/2 + 0.08/2 = 0.80, and the total utilization rate U_{RM} is 0.82 by applying the equation (1). It can be seen theoretically that the MIL-STD-1553B scheduling algorithm using the RM algorithm

operates normally with the system utilization rate (0.80) < U_{RM} (0.82).

Therefore, in the scheduling algorithm proposed in this paper, even if a Tx command executes in the MIL-STD-1553B communication and an Rx command executes immediately, the MIL-STD-1553B communication in real time meets the deadline without data loss within the period set in the RTiK+.

4. Results

4.1. Experimental Environment

To verify the real-time performance of the scheduling algorithm with the RTiK+ for MIL-STD-1553B communication, the experimental specifications for the host PC and the target PC is shown in Table 4.

Table 4: An Experimental Environment between host PC and target PC

Item	Host PC	Target PC
CPU	Intel Pentium Dual-Core 2117U @1.80 GHz	Intel Core i5-2500 @3.30 GHz
OS	Windows 8	Windows 7

4.2. Experimental Method

A verification is a measurement method that calculates the period by storing data using the RDTSC command to get a timestamp value to proof a real-time processing performance of the RTiK+. The algorithm coded in RDTSC for this experiment is shown in Figure 12.

The current timestamp stored in 64-bit values of EAX and EDX is read, store the timestamp value in the timearray[i].QuadPart array at i index, and calculate the difference between adjacent index values in the array. The f file is designed to be saved. In this way, a timer interrupt set with a period of 2ms occurs in the RTiK+, and whenever a real-time thread is executed in the user mode, the value of the timestamp is stored in the array.

Algorithm 5:

Result:

```
Hthread()
{
   _asm
   {
      RDTSC
      MOV lowval, EAX
      MOV highval, EDX
   }
   clockcal2_LowPart = lowval;
   clockcal2_HighPart = highval;
   timearray[i].QuadPart = clockval2.QuadPart;
   if(i==32000)
   {
      for(j=1, j<32000, j++)
      {
         fprint(f, "%.10f\n", (double)(timearry[i].QuadPart-timearray[j-1].QuadPart))/CPU);
      }
      i=0;
      break;
   }
}
```

Figure 12: An Algorithm with RDTSC for Period Measurement of the RTiK+

4.3. Measurement and Results

4.3.1. A Measurement for MIL-STD-1553B Communication

To verify a real-time performance support of the RTiK+ in the MIL-STD-1553B communication, the experimental environment was configured as shown in Figure 13. The RTiK+ is implanted to the host PC equipped with the MIL-STD-1553B communication, and the Period of the MIL-STD-1553B communication is set to 2ms with the RTiK+. The experimental data were measured with the RDTSC command as previously explained.

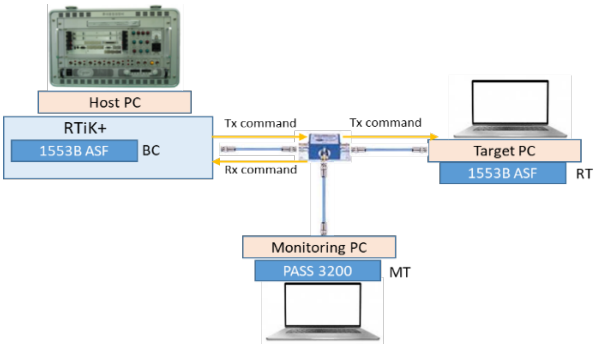


Figure 13: A Configuration to measure Period of MIL-STD-1553B

And to check the integrity of the data that there is no loss of transmitted and received data when performing the MIL-STD-1553B communication, The SBS Technologies' PASS 3200, a special-purpose MIL-STD-1553B monitoring equipment, is as shown in Figure 13, it connects between the host PC operating with the MIL-STD-1553B BC function and the target PC operating with the MIL-STD-1553B RT function to operate the MIL-STD-1553B MT function.

4.3.2. Result of Measuring Period during MIL-STD-1553B

In this paper, the maximum experiment period of 2ms was measured by RDTSC command during the MIL-STD-1553B communication. In addition, in order to compare a real-time performance for the MIL-STD-1553B between RTiK+ and RTX, the Period of the MIL-STD-1553B communication was measured as shown in Table 5 by applying the same experimental environment and experimental method.

Table 5: Comparison with Period of MIL-STD-1553B between RTiK+ and RTX

Period	Period of RTiK+			Period of RTX		
	Min.	Max.	Tolerance	Min.	Max.	Tolerance
2ms	1.993	2.031	1.57%	1.901	2.015	4.91%
5ms	4.955	5.003	0.90%	4.998	5.252	5.04%
10ms	9.993	10.005	0.06%	9.996	10.003	0.04%

4.3.3. A Result of Monitoring Data during MIL-STD-1553B

During the MIL-STD-1553B communication, in order to check the accuracy and integrity of the transmitted data, all data passing through the MIL-STD-1553B communication bus is stored with the PASS 3200 using the MIL-STD-1553B MT function as Figure 13. As shown in Figure 14, the first data transmitted by performing Tx command from the MIL-STD-1553B communication BC function to RT function is 32 words, all of which are 0xC912, and 32 words received by Rx command from RT function are all 0xC912. It seems that there is no loss or wrong data.

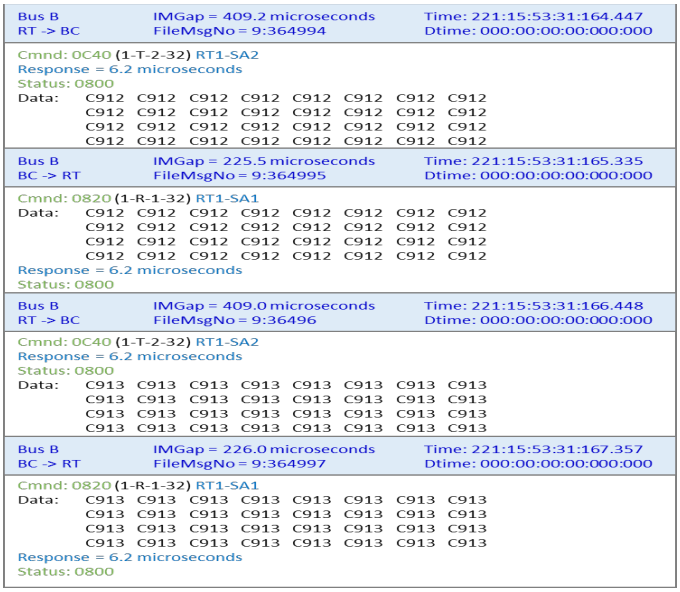


Figure 14: A Result of the MIL-STD-1553B monitoring by PASS 3200

5. Discussion

5.1. Process in User Mode of RTiK+

If the scheduling algorithm for the MIL-STD-1553B is analyzed based on the experimental results for period of the MIL-STD-1553B communication, Figure 15 and 16 is shown. As can be seen from this result, there was no case of overload the deadline in the 2ms period of the MIL-STD-1553B communication to which the scheduling algorithm with RTiK+ was applied, and it was proofed that the MIL-STD-1553B communication operates normally using the 2ms period of the RTiK+. As shown in Table 5, when the tolerance of the RTiK+ used in the MIL-STD-1553B communication is 1.57%, it is ±0.03ms from 2ms, but when the period for each task of Tx(pi) and Rx(pi) transmits 32 words, the maximum transmission time is 0.80ms, which is less than the half period of 2ms. The system utilization rate is 0.32 from the equation (3), and U_{RM} is calculated as 0.82 from the equation (1). It means that the scheduling algorithm for the MIL-STD-1553B operates normally without any communication errors.

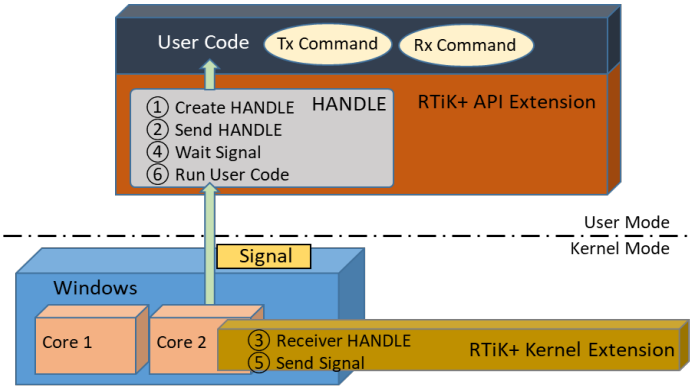


Figure 15: An Operation Diagram of Tx and Rx command for MIL-STD-1553B Scheduling Algorithm

In addition, this is the result of proving that the miss rate is 0 in the 2ms period provided by the RTiK+ for real-time processing function in the MIL-STD-1553B communication.

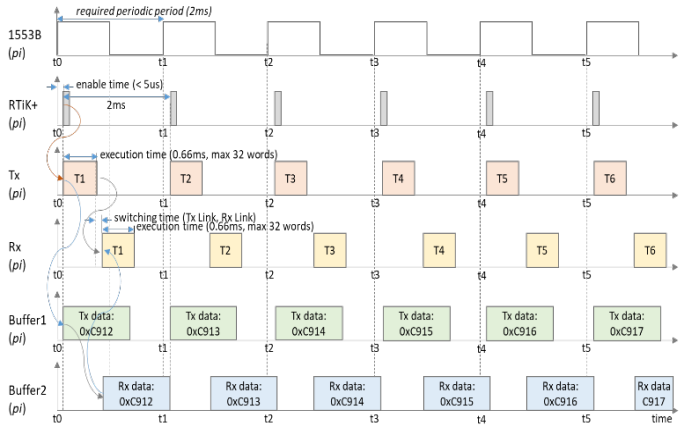


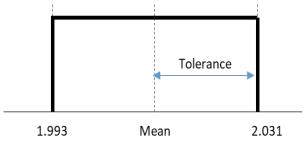
Figure 16: An Analysis Diagram for MIL-STD-1553B Scheduling Algorithm

5.2. Evaluation for Real-Time Performance of RTiK+

In order to evaluate a real-time performance of the MIL-STD-1553B communication with the result of Table 5, the 2ms period is schematically shown in Figure 17. In Figure 17(b), it can be seen that all the periods occurring in the scheduling algorithm with a period of 2ms of the RTiK+ for the MIL-STD-1553B are between 1.993ms and 2.031ms.

	Period of Scheduling algorithm with RTiK+ for MIL-STD-1553B		
	Min	Max	Tolerance
2ms	1.993	2.031	1.57%

(a) Result of Period of 2ms for MIL-STD-1553B



(b) Graph of Period of 2ms for MIL-STD-1553B

Figure 17: A Performance of 2ms Period for MIL-STD-1553B

For a performance measurement of a real-time system, 32 words of data are transmitted according to the period provided by the scheduling algorithm with the RTiK+ in Windows for the MIL-STD-1553B communication, and the miss rate is applied to verify the jobs performed after the deadline. The figure 18 is shown for miss rate during the MIL-STD-1553B communication to decide success or failure in real-time system.

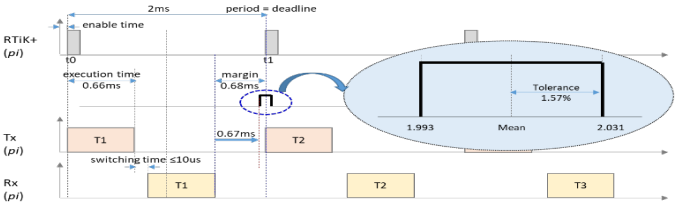


Figure 18: A Miss Rate of the MIL-STD-1553B Communication

The system based on x86 with Windows 8 applied in this paper is a hard-real-time system that must meet a deadline. Therefore, if all deadlines are met, the MIL-STD-1553B communication is a success, otherwise it is a failure.

$$\text{Margin of MIL-STD-1553B} = (pi\text{-}ei)\text{-Tolerance (\%)} \quad (5)$$

Substituting the measured result into Equation (5), actually the margin of the MIL-STD-1553B is 0.39ms that (2ms-1.60ms) - (2ms-1.9938ms). And the miss rate of the MIL-STD-1553B is defined as in equation (6).

Miss rate of MIL-STD-1553B = Number of Error / Number of Communications (6)

The MIL-STD-1553B using PASS 3200 analyzed all data as shown in Figure 14 in the period of 2ms, there was no communication error and all data were normal during the MIL-STD-1553B.

Table 6: Performance of Scheduling Algorithm by RTiK+ for MIL-STD-1553B

Period	Min	Max	Miss Rate
2ms	1.997ms	2.006ms	0%
5ms	4.961ms	5.038ms	0%
10ms	9.995ms	10.007ms	0%

If the MIL-STD-1553B communication period is 1ms and the algorithm proposed in this paper is applied, the maximum transmitted data is Tx(ei) = Rx(ei)) = 0.80ms. This is 0.60ms from the 1ms period of MIL-STD-1553B communication, so even if a double buffer is used, overhead may occur. RTiK+ itself can provide a period within 3% error when there are no workloads on the CPU in a 1ms period.

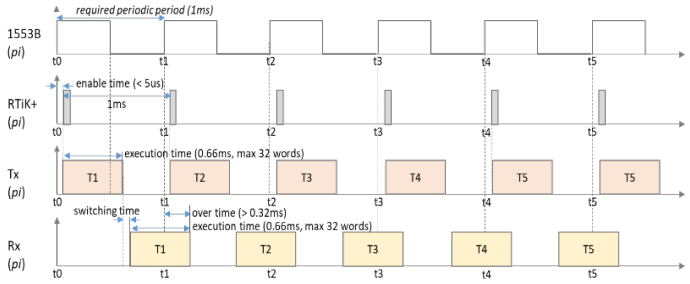


Figure 19: An Analysis Result of 1ms Period for Scheduling Algorithm

It is predicted that MIL-STD-1553B communication will not operate normally due to miss rate because of failure to comply with the deadline due to the lengthy duration as shown in Figure 19.

6. Conclusion

In recent year, with the development of high-speed communication technology and computer technology, the scope of a real-time information sharing has been expanded, and the performance of electronic equipment has been improved remarkably. The accuracy of data collection and real-time data communication are required in industry and defense system.

The RTiK+ presented in this paper controls the MSR_FSB_FREQ register to support a real-time performance on Windows in an environment where x86-based Windows 8 is installed, and calculates the CPU clock tick value required for the operation of the local APIC timer to be a Windows independent timer interrupt was made to occur, and in this way, the matter that local APIC is initialized after booting in Windows 8 is solved.

The experimental verification was checked by measuring the period by calculating a timestamp in the PC's internal memory using RDTSC, and measuring a real-time processing performance of the RTiK+.

And to support a real-time performance in the MIL-STD-1553B communication, The RTiK+ was implanted in the Windows based on x86 system, and a real-time processing functions were added to the BC and RT functions. Here, the real-time performance was provided when BC and RT function of the MIL-STD-1553B communication are implemented, and by applying the scheduling algorithm with RTiK+ for a high-speed MIL-STD-1553B communication such as 2ms of period is implemented to prevent data loss due to error of the period. To proof this scheduling algorithm experimentally, The RTiK+ was implanted in the Window 8 and the MIL-STD-1553B communication of the period of 2ms, 5ms, and 10ms was set to verify whether a real-time processing function of the RTiK+ normally while MIL-STD-1553B communication is performing.

In addition, in order to compare a real-time performance between the RTiK+ and the RTX as a third party for the MIL-STD-1553B communication, the experiment result for the period of the MIL-STD-1553B by applying the experimental environment and experimental method is from 1.993ms to 2.031ms in the period of 2ms. The RTiK+ had a tolerance of 1.5%, and the results of this experiment proved that RTiK+ can be applied to systems by replacing third party such as RTX. Finally, a real-time processing performance of the scheduling algorithm with RTiK+ proposed in this paper for MIL-STD-1553B implemented has a miss rate of 0 during the MIL-STD-1553B communication with a period of 2ms for system that requires high reliability. It also proved that the real-time performance and data integrity for scheduling algorithm with RTiK+ are guaranteed.

In the future research, the methods proposed in this paper is needed in the RS-232C, RS-422 and Ethernet used mainly in industry and defense system, and the research for the Windows 10 based on x86 system is needed as well.

Acknowledgment

I would like to thank my dissertation advisor Professor Lee Cheol-Hoon from CNU for his supervise and expert advice, as well as Mr. Muammar Abdulla Abushehab for extraordinary support.

References

[1] C.M. Krishna, Gang G. Shin, Real-Time Systems, McGrow-Hill, 1997.
[2] C. Koo et al, "Distributed Simulator design by using of SimNetwork to overcome speed limit on GenSim", Recent Advances in Space Technologies (RAST), Proceedings of 5th International Conference on RAST 2011, 430-435, 2011, doi: 10.1109/RAST.2011.5966872.
[3] H. Kim et al., "The Design and Performance Verification of Real-Time Inspection Equipment Software based on WindowsOperating System", Journal of the Korea Contents Association, **17**(10), 1-8, 2018, doi: 10.5392/JKCA.2017.17.10.001.
[4] S. Koh et al., "The method of development for enhancing reliability of missile assembly test set", The Journel of The Korea Academia-Industrial Cooperation Society, **19**(8), 37-43, 2018, doi: 10.5762/KAIS.2018.19.8.37.
[5] J. Kim et al., "Real-time Processing Method for Windows OS Using MSR_FSB_FREQ Control", Journal of Korea Multimedia Society, **24**(1), 95-105, 2021, doi: 10.9717/KMMS.2020.24.1.095.
[6] J. Lee et al., "Timer Implementation and Performance Measurement for Providing Real-time Performance to Windows 10", Journal of the Korea Contents Association, **20**(10), 14-24, 2020, doi: 10.5392/JKCA.2020.20.10.014.
[7] S. Kwon , "Design and Implementation of Air Vehicle Test Equipment for Unmanned Aerial Vehicle", The Journal of Advanced Navigation Technology", **24**(5), 251-260, 2020, doi: 10.12673/JANT.2020.24.4.251.
[8] J. Jung et al., "Analysis for Next-generation High-Speed MIL-STD-1553B Bus Technology", Journal of Aerospace Ssytem Engineering, **11**(6), 76-83, 2017, doi: 10.20910/JASE.2017.11.6.76.
[9] http://www.intervalzero.com
[10] Intel, Intel® 64 and IA-32 Architectures Software Developer's Manual, **2**

(2A, 2B, 2C & 2D), Volume 3A and Volume 3B, Intel, 2016.
[11] Intel, Intel® 64 Architectures x2APIC Specifiation, Intel, 2010.
[12] Intel, Intel Core i7-600, i5-500, i5-400 and i3-300 Mobile Processor Series 2010.
[13] A. Jo, "Integrated Middleware for Real-Time Device Driver on Windows", Journal of Korea Contents Association, **13**(3), 22-31, 2013, doi: 10.5392/JKCA.2013.13.03.022.
[14] H. Lihua, Analysis of Fuel Cell Generation System Application, Ph. D Thesis, Chongqing University, 2005.
[15] https://en.wikipedia.org/Rate-monotonic_scheduling
[16] J. Liu, "Real-Time Systems", Prentice Hall, 2000.
[17] M. Gong et al., "Dynamic Voltage Scaling for Scheduling Mixed Real-Time Tasks", Proceeding ICESS 2007 Third International Conference, 488-497, doi: 10.1007/978-3-540-72685-2.
[18] M. Lee, C. Lee, "Low Power EccEDF Algorithm for Real-Time Operating Systems", The Journal of the Korea Contents Association, **15**(1), 31-43, 2015, uci: 1410-ECN-0101-2016-004-001099668.
[19] M. Jung et al., "Optimal RM scheduling for simply periodic tasks on uniform multiprocessors", Proceedings of the 2009 International Conference on Hybrid Information Technology, 383-389, 2009, doi: 10.1145/1644993.1645064.
[20] M.E. Russinovich, "Windows Internals, Part 2 (6th Edition)", Microsoft, 2012.
[21] https://en.wikipedia.org/wiki/MIL-STD-1553
[22] ILC DDC, MIL-STD-1553 Designer's Guide (5th Edition), ILC Data Device Corporation, 1998.
[23] J. Kim et al., "Real-Time Supports on Tablet PC Platforms", Proceeding IEEE 2020 15th International Conference for Internet Technology and Secured Transactions(ICITST), 111-117, 2021, doi: 10.23919/ICITST51030.2020.935.1322.
[24] J. Kim, "Real-time Processing Method for Windows OS Using MSR_FSB_FREQ Control", Journal of Korea Multimedia Society, **24**(1), 95-105, 2021.
[25] D.A. Solomon, "Inside Microsoft Windows 2000, Third Edition", Microsoft, 2000.
[26] http://msdn.microsoft.com/en-us/library/ms810029.aspx
[27] D.A. Godse, A.P. Godse, Microprocessors, Technical Publications Pune, 432-472, 2007.
[28] O. Bailey, Embedded systems: desktop integration, Wordware Publishing Inc, 34-56, 2005.
[29] Intel, "MultiProcessor Specification (Version 1.4)", Intel, 1997.
[30] J. Choi, "A Study on Developing the Missile System Test Set Software with DDS", Proceeding of Korea Institute of Communication Sciences Conference, **2018**(11), 614-615, 2018, uci: 1410-ECN-0101-2019-567-000045708.
[31] https://en.wikipedia.org/wiki/Front-side_bus
[32] https://en.wikipedia.org/wiki/CPU_multiplier
[33] https://docs.microsoft.com/en-us/windows-hardware/drivers/ddi/wdm/nf-wdm-mmmapiospace
[34] https://docs.microsoft.com/en-us/windows/win32/procthread/scheduling-priorities
[35] http://msdn.microsoft.com/en-us/library/ms685100 (VS.85).aspx
[36] C. Lee, Real-Time Systems, CNU, 2012.
[37] K. Kim et al., "An Indirect Data Transfer Technique based on MIL-STD-1553B for the Software Upgrade of Embedded Equipments on a Missile Assembly with Booster", Journal of the Korea Institute of Military Science and Technology, **19**(1), 43-49, 2016, doi: 10.9766/KIMST.2016.19.1.043.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 395-400 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Devices and Methods for Microclimate Research in Closed Areas – Underground Mining

Mila Ilieva-Obretenova*

Electrical Engineering Department, The University of Mining and Geology, "St. Ivan Rilski", Studentski Grad, "prof. Boyan Kamenov" Street, Sofia 1700, Bulgaria

ARTICLE INFO

Article history:
Received: 29 April, 2021
Accepted: 14 August, 2021
Online: 28 August, 2021

Keywords:
Sensors
Measurement Devices
Microcontroller
Data transfer
Underground Mining

ABSTRACT

Technical safety and health are especially important for mining-extracting industry. Even though the respective lows and good engineering practices exist, technologies develop and could address even better security for humans and equipment. The research question is to survey microclimate sensors in underground mining and to find whether they are ready for automation. The article is inspired from the research in "Computer System for Microclimate Management in Closed Areas of the Post-Mining Galleries and Greenhouses". The author offers a short review of state and trends in development of sensors and devices for monitoring and reporting of environmental parameters in underground mining. All environmental parameters: air temperature; temperature of surrounding building constructions, heated surfaces of technological machines and equipment; heat flow, heat irradiation; relative humidity; velocity of air flow; noise; illumination; gas concentration; dust level; ionizing radiations; radon concentration in air are represented with relevant measurement devices and measurement units. The next step is representing of fast-developing sensors using scientific references. Author performs quality assessment of their suitability for automated data transfer and management. The assessment criteria are: Analogue measurement devices, Digital measurement devices, Availability of microcontroller. Findings are proposed for discussion: recent used devices in underground mining are not suitable for automatization, because they miss a controller. The availability of controller also presumes availability of management services. On the base of articles about management of sensors and controllers future work is proposed: 1. Integration of existing elements sensor and controller and defining of its management; 2. Moving of new elements and synthesis of algorithms for its management. This will lead to more precise assessment of industrial risk and improvement of safety activities.

1. Introduction

Mining-extraction sector is especially important for the economics in Bulgaria. Therefore, the technical safety and professional health must embrace different activities on protection of employees and assets by reducing to minimum of dangers, hazards, failures, and oversights. Even though the respective lows and good engineering practices exist technologies develop and could address even better security of humans and equipment.

The article is inspired from research in [1] which considers a computer system for microclimate management in closed areas of post-mining galleries and greenhouses. Other authors describe this problem too. In [2], the authors represent Pictograms for admissible norms of harmful substances, but they miss devices and measurement methods. In [3], the organizers display a standard for microclimate design, but they miss details for modern devices for monitoring and reporting. In [4], the authors consider safety technics in underground mining, but without details in trends for devices and sensors. In [5], the author describes methods and devices of illumination technology in underground mining and tunnels, but without including other parameter of environment and the appropriate measurement devices. In [6] and [7], the authors represent rules for radioactivity (radon) measurement, while [8] portrays harmless work with cyanides, and [9] displays design of industrial ventilation, but they all miss automated data transfer. In [10], the authors describe a wireless permanent monitoring of coal

*Corresponding Author: Mila Ilieva-Obretenova, Bulgaria, Sofia 1505, Tcherkovna Str. 21, App. 8, milailieva@abv.bg, mila.ilieva@mgu.bg

strata, which could be applied to wireless permanent monitoring of microclimate. In [11], the author explains levels of IoT environment in underground mining as the emphasis is on the security of data transfer, but different sensor types and their management are not detailed.

The purpose of the article is to make a short review of state and trends of sensors and devices for monitoring and reporting of environmental parameters in underground mining and to accomplish a quality assessment of their suitability for automated data transfer and management.

2. Methodology

The survey of sensors for underground mining is accomplished in University of Mining and Geology, a leading university on mining science in Bulgaria. Author performs the research after publishing the guide on safety technics in underground mining sponsored [4]. All parameters of the closed area, basic sensors, and measurement units are represented. In the next step the fast-developing sensors with their improved versions are depicted. The author uses specialized references. Assessment criteria for measurement devices are defined on the base of Computer System for Microclimate Management in Closed Areas of the Post-Mining Galleries and Greenhouses [1]. Sensors have analogue or digital output in this system. This leads to the important need of placement of an analog-to-digital converter (ADC) and building of digital interface (SPI, I2C, CAN, Ethernet, etc.) [12], [13]. Most of the built-in ADCs have a resolution of 10 or 12 bits and contain 8-16 input channels which is sufficient for these applications [14]. The core of the system is a microcontroller that samples data from the sensors, processes the information and sends it to a computer through interfaces. Since this system will sample data during long periods of time (minutes, hours, days or even months), a backup power supply is needed along with the main one [15]. The switching between these two power supplies must happen automatically; therefore, an electronic switch is needed too. The results from the measurements should be transferred to the computer, where the end-user can view them. This is done with the interface RS232 [16]. As more than one system is used, the interface could be Wi-Fi to speed up the process of data gathering [17]. The network of measurement elements consists of a base microcomputer station and several sensors, distributed in the space, and linked to the station by cables or wireless Ethernet. The measurement of all climatic elements is made within 1 sec in total with an interval between the separate measurements - 3 minutes. It is necessary to perform measurements with low speed, as most of the values are not changing rapidly and because of the inertia of the sensors. Figure 1 shows the existing Computer System for Microclimate Management in Closed Areas of the Post-Mining Galleries and Greenhouses [1]. Cortex-M_0 architecture is used. It has a 32-bit embedded microprocessor, especially designed for low-powered optimized microcontrollers.

Therefore, specialists could assess the new sensors by the following three criteria:

- Whether are they analogue?
- Whether are they digital?
- Whether are they connected to microcontroller for automated data transfer?

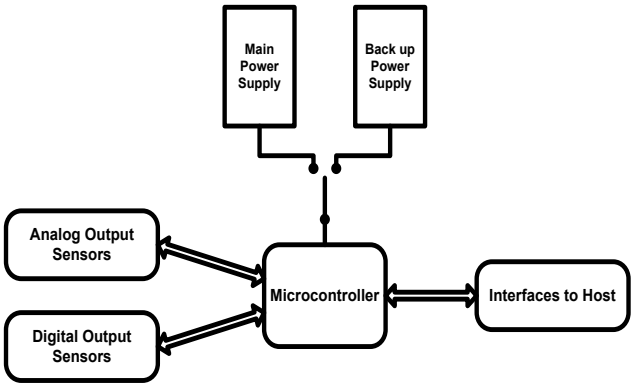


Figure 1: Computer system for microclimate management in closed areas of the post-mining galleries and greenhouses

Next steps are:

- Discussion of findings and their consequences and
- Proposal of future work about devices` investigation.

3. Results

The first result from the survey represents the systemized environmental parameters in underground mining, conventional devices for their measurement and the corresponding measurement units. It is shown in Table 1.

Table 1: Environmental parameters in underground mining, conventional devices for their measurement and the corresponding measurement units

Parameter	Device and measurement unit
1. Air Temperature	Quicksilver and Thermistor Thermometers, [ºC]
2.Temperature of surrounding building constructions, heated surfaces of technological machines and equipment.	Contact Thermometers, [ºC]
3.Heat flow, Heat irradiation	Actinometer, [W/m^2]
4.Relative Humidity	Psychrometer of Assmann, [%]
5.Velocity of air flow	Anemometer, Kata Thermometer, [m/s]
6.Noise	Noise meter (Sound meter) for exposition; integrating noise meter; stationary installations; installations, fitted to human (dosimeters), [dB]
7.Illumination	Luxmeter (Illuminance Meter), [lx] Luminance Meter [cd/m^2] Spectroradiometer, [nm]
8.Gas concentration	Gas analyzer, [%]
9.Dust level	Konimeter, [particles number/cm^3]
10.Ionizing radiations	Geiger-Muller counter, [Sv] Scintillation counter, [Sv] Pocket chambers and Pocket Dosimeters, [Sv] Film badge Dosimeter, [Sv] Thermoluminescent Dosimeters (TLDs), [Sv]

	Electronic Personal Dosimeter (EPD), [Sv]
11.Radon Concentration in Air	Hermetically closed chamber, [%]

The second result from the survey shows the fast-developing sensors according to specialized references. These are sensors for air temperature, heat flow, relative humidity, illumination, and gas concentration.

3.1. Sensors for Air Temperature

Classical measurement devices for air temperature are quicksilver and thermistor [18] thermometers, but they miss automated data transfer.

In [19], the authors represent smart temperature sensors and systems of temperature sensors. For temperature measurements in smart sensors and microelectromechanical systems the most frequently used elements are transistors, thermocouples, and thermopiles (several connected thermocouples), because these elements could be accomplished with integrated circuits technology. For the temperature range from -55ºC to 150ºC bipolar transistors are especially useful as temperature sensors because they could be produced by practically all standard technologies for integrated circuits together with other circuits. With similar circuits could be made a reference for a voltage gap. This allows performing of radiometric research, where temperature sensing voltages are compared with predefined voltage reference. By CMOS analogue circuits extremely high precision could be achieved by using error-reducing technics such as dynamic synchronization of elements and chopping (signal cutting). The smart temperature sensor is represented with output signal with modulated duty cycle. High precision is reached with remarkable low energy consumption. The output of sensor is digital, but without automated data transfer.

In [20], the authors portray a temperature sensor with bipolar transistor used as a diode. Thereby bigger temperature range, lower power consumption and smaller chip area are reached. The sensor works in time mode: The time is measured for a capacitor charging to activating voltage of the transistor. A comparator is used instead of conventional sigma-delta analog-to-digital converter. Therefore, output of the sensor is digitized, but without automated data transfer.

3.2. Sensors for Heat Flow

In underground mining a heat flow is measured, coming from surrounding building constructions, heated surfaces of technological machines and equipment. Classical measurement instruments for heat flow are with thermistors (thermopiles) [21]. The sensor consists of a set of thermistors and an array of converters produces voltage, proportional of the heat flow and allows direct measurement of heat flow. The availability of thermistors array allows to be made a comparison between heat flow assessment, obtained from converters, and calculated heat flow, using temperature differences and Fourier Low. The element is extremely sensitive with accuracy lower than 0.1 heat flow units (HFU) and could be used for long term readings. The output of element is digital, but without controller for data transfer.

The recent heat flux meters (HFM) suffer from low performance by weak and constant flows, non-uniformity of the heat flux in the measurement section and impossibility to control the heat flux at different temperatures.

In [22], the authors show a calibration system of a HFM in different operating conditions (low and middle heat fluxes) to design a new prototype with heat flux uniformity and to improve performance in a specific sub-region of the measuring section where the HFM is applied. Furthermore, a metrological characterization of the system allows a combined standard error to be better than 6% at low heat flows. The element is analogue, without controller for data transfer.

In [23] and [24], the authors report heat flow sensors integrated in textile. They determine the amount of heat exchanged between human body and environment. The aim is to improve comfort, efficiency, and sometimes safety of the wearing person. Passive heat flow sensors are used. Such sensors are still under developing and ideas for their appropriate calibration are proposed. Moreover, the existing heat flow sensors are impermeable and do not allow evaporation. This shows less precise results when phenomenon evaporation exists. A flexible heat flow sensor is developed, which is sensitive to water vapor. This sensor takes into consideration phenomenon evaporation and allows better energy measurement during heat exchange. The sensor is analogue without controller for data transfer.

3.3. Sensors for Relative Humidity

Classical devices for relative humidity measurements are psychrometers. Recent research [25] displays indoor air quality and the connected health effects. Complaints about sensory irritation in eyes and upper airways, also perception for dry air are among the most important symptoms in closed areas. In this line sensors for dynamics of particles, bacteria, and viruses in closed areas are necessary for improvement of labour quality and with a sharp focus on absolute humidity impact. The sensors could be analogue or digital and with microcontroller.

In [26], the authors consider a sensor for air humidity measurement using sensitive capacity-dependent crystal. Probe sensitivity is represented too. Moreover, new idea for excitation of entire humidity sensor with stochastic test signals is described, and the humidity measuring method is given. It includes the influence of test signals on weighting function uncertainty and on the A/D-D/A conversion. Error by humidity measurement is lower than 0.2% (t = 15ºC - 25ºC and humidity = 50% - 98%). There is a conversion ADC, DAC in sensor, but it misses a controller.

Research in [27] describes a program, which could simulate simultaneously heat and humidity transfer from walls and plates in closed area and its impact on indoor temperature and humidity. The aim is a sensor development. Recommendation is to be digital and connected to microcontroller.

3.4. Sensors for Illumination

Illumination on basic workplaces and galleries, where the miners move, is obligatory. This is not easy, because the atmosphere in coal mining is often accompanied with methane availability and is highly explosive. Humidity is high, but environment reflectance is extremely low. Therefore, special

explosion-proof illuminaters are used and their application is by hardest conditions – low reflection factor, methane availability, high humidity, low and narrow galleries etc.

What are the conditions in ore mining and coal extracting mining? Basically, there are not methane in the first and atmosphere is not explosive. In underground ore mining and these for extracting of non-metalliferous illumination is performed with illuminaters proof from dust and water. In such mining there is no danger from explosion of highly explosive compounds, but they are threatened from fire [5].

For illuminance tracking are applied illuminance meter (luxmeter) and luminance meter.

For measurement in mining galleries are suitable new generation luxmeters "ТКА-Lux" ("ТКА-Люкс") and "ТКА-РКМ-31" ("ТКА-ПКМ-31"). They have metrological characteristics on the level of the best world manufacturers. The illuminance measurement range is 10 – 200 000 lx, error – 6%.

They have the following performance:

- Viewing angle: 1.0 – 1.5º,
- Range of measurement: 10.0 – 2000 cd/m^2,
- Distance to measured object ~ 7m.

Luxmeters with dark scale are applied in coal mining, which is dangerous with gas and dust.

In underground mining exists blindness danger too. Therefore, it is necessary to measure luminance. It is measured with luminance meter. Basic measurement unit is candela [cd]. Recent luminance meter is TKA [5] and [28].

Illuminance assessment in underground mining must obligatory include:

- Illuminance assessment,
- Unevenness assessment,
- Blindness assessment,
- Discomfort assessment.

Special luxmeters are used for illuminance measurement on given surface (workplace). The most widespread luxmeters are photoelectric (objective) luxmeters, consisting of selenium element in series with galvanometer and micro amperemeter. The principle operation of photoelectric luxmeter is conducting a photocurrent between galvanometer`s electrodes by illumination of photoelement. The photocurrent is proportional of illuminance and this dependance could be assumed as linear, when the inner resistance is significant bigger than outer. This ratio is obtained by measurement of small illuminances and application of measurement devices with small inner resistance. The boundary electromotive force of selenium element (0.3 – 0.5 V) is obtained by illuminance 1000 lx and by bigger illuminance it does not change. Therefore, the photoelement illuminance does not have to exceed 1000 lx. For measurement of bigger illuminance is applied an absorbing filter, which decreases X-fold illuminance on photoelement. Selenium element is characterized by its spectral sensitivity too. It is graduated with the help of filament lamp. When an illuminance of other light sources (another light spectrum) is measured it is necessary to use correcting light filters or correcting coefficients, which by naturally illuminance is 0.8, and for luminescent lamps are from 0.88 to 1.2 according to their type [4].

In [29], the author publishes a "Luxmeter, showing light intensity, as it is received from a sensor".

The listed sensors are analogue and without controller for data transfer.

In [30], the authors show a spectroradiometer as a part of program-instrumental complex for management of illuminance installation in closed areas. Problems of spectroradiometer development are considered. Basic principles for building of automated illuminating system are formulated, based on criteria for quality and quantity assessment of illuminance effect. Scheme of automated illumination complex is considered. New development is represented – spectroradiometer, covering optical irradiation measurement and being a block of system for spectral and energetic management of LED illuminating installations. Attention is paid on the lack of metrological insurance of control of photometric characteristics of energy conservation technology, based on LEDs. It is foreseen that the sensor is digital and has a microcontroller.

3.5. Sensors for Gas Concentration

For measurement of gas concentration portable analyzers are widely used [4]: gas indicating tubes and gas analyzers. From gas analyzers most popular are: Gas Sense – 1000.L.EX; portable device for leakage discovery of explosive gases – Methane (CH_4), Propane (C_3H_8), Butane (C_4H_{10}), Liquified Petroleum Gas (LPG), Hydrogen (H_2), Ethanol (C_2H_5OH); Dräger Pac 5500 – measurement of Carbon Monoxide (CO), Hydrogen Sulfide (H_2S) or Oxygen (O_2); MSA EX-METER II – for measurement in explosive atmosphere; MSA SOLARIS 4 – Four channel gas analyzer, analogue, without microcontroller.

The trend is semiconductor sensors to be doped with metal oxides (e.g., tin) and their application for discovery and protection of environment from different gases like Carbon Monoxide (CO), Carbon Dioxide (CO_2), Hydrogen sulfide (H_2S), Ammonia (NH_3), swamp gas methane, LPG, and many others pollutant gases [31]. Different characteristics of gas sensor in composite form by different temperatures are also displayed. Evidence is produced that sensors in composite form are more resilient than these with a single material. A research about sensor`s output – analogue or digital – is not represented. About its ability for data transfer too.

The author in [32] considers manufacturer methods of zinc oxide nanorods and their application for gas discovery with low concentration due to their various conductance range, reaction of oxidative and reductive gases and extremely high sensitivity and selectivity. A research about sensor output - analogue or digital – is not given. For data transfer ability too.

The survey in [33] describes an electronic gas sensor from nano graphene, which could discover extremely low quantity gas (harmful chemical agents). Nano graphene does not have grainy structure and offers long term stability. Moreover, material has high quality crystal lattice, thus charge carriers move with high

velocity and generate low noise. All this supposes design of digital output of the sensor. Due to their specific properties carbon nanomaterials have potential for manufacturing of automated sensors.

The authors in [34] portray high selective gas sensors for smart monitoring and control of air quality, search and save of people etc. Substance choice in complex gas composite is the main challenge. The article represents a guide for material engineering for absorbing, size selecting and catalytical filters. Accent is laid on materials design with purposeful gas separation, portable elements integration and performance. All this shows that gas sensors could be designed with digital output and microcontroller for data transfer.

4. Discussion

The performed short review leads to the findings that the new measurement devices in underground mining are not suitable for automation. Small part of them has a digital output, and most of them are without microcontroller for data transfer. To accelerate and facilitate automation sensors must be *delivered with controller.* The availability of controller also presumes availability of management services, united in Program Library. The resent achievements are represented with articles with low level detail [35] or high-level concepts [36] and [37]. Other authors go even further away proposing sensors for inspecting of underground mining to be moveable and to be carried from a drone, but without giving detail about their management [38]-[41]. Therefore, Articles for sensors with microcontrollers and management programs must describe comprehensive:

- Integration of sensors and controller;
- Algorithms for management of integrated sensor-controller;
- Function for bearing of sensors for timely reading of fast changing parameters at different places in space aiming sensor security;
- Algorithms for management of sensors bearing, as some problems are considered: signal propagation; life, size, and safety of battery; security of bearing system etc.

5. Conclusion

Main conclusions from this research are as follows:

1. Criteria for assessment of measurement devices in underground mining are synthesized. Methods on which devices perform their functions are specified.
2. Devices and measurement units which cover all environmental parameters in underground mining are listed, namely: air temperature; temperature of surrounding building constructions, heated surfaces of technological machines and equipment; heat flow, heat irradiation; relative humidity; velocity of air flow; noise; illumination; gas concentration; dust level; ionizing radiations; radon concentration in air.
3. Representations of fast-developing devices are accomplished: for air temperature; for heat flow; for relative humidity; for illumination and for gas concentration. Quality method is used. Assessments organized according to synthesized criteria are defined.
4. On the base of discussion about availability of controller to sensor the future work is proposed in following aspects:

a. Integration of existing elements sensor and controller and defining of its management;

b. Moving of new elements and synthesis of algorithms for its management.

Formulated aim is complex, and this article concentrate on devices` assessment. Representation level and assessments have the necessary depth to illustrate the needed innovations. Basic conclusion from this article is that the developed assessments are eligible base for development of sensors and their management in closed areas, structured as underground mine workings. This will lead to more precise assessment of industrial risk and improvement of safety activities.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgment

This article is a part of project “Survey and Management of Sensors for Microclimate in Mining Galleries” in University of Mining and Geology “St. Ivan Rilski”, Bulgaria.

References

[1] M. Ilieva-Obretenova, "Computer System for Microclimate Management in Closed Areas of the Post-Mining Galleries and Greenhouses." in 7th International Conference on Energy Efficiency and Agricultural Engineering (EE&AE), Ruse, 2020, 1-4, doi: 10.1109/EEAE49144.2020.9278785.

[2] Ministry of labour and social policy. Healthy Workplaces Campaign, Pictograms 2018-2019, Sofia. (In Bulgarian)

[3] https://russian.worldbuild365.com/news/nbs9ontsc/hvac/mikroklimat-po-gostu-chto-dolzhen-uchitivat-kazhdyy-proektirovschhik. Accessed June 2021.

[4] B. Vladkova, Техническа безопасност, University of Mining and Geology “St. Ivan Rilski” – Sofia, 2020. (In Bulgarian)

[5] G. Ganchev, Книга за осветлението. Минното осветление е тъмна работа, University of Mining and Geology - Sofia, 2018. (in Bulgarian)

[6] D. Pressyanov, et al., "Passive radon monitors with part-time sensitivity to radon". Radiation measurements, **118** (2018) 72-76, doi: 10.1016/j.radmeas.2018.08.014

[7] H. Mischo, Radioactivity Practical Course. Institute of Mining and Special Civil Engineering. Chair of Underground Mining Methods, Technical University Bergakademie Freiberg, 2018.

[8] B. Vladkova, Цианидите в добивната промишленост – технологични особености и безопасност при работа с тях. University of Mining and Geology, Sofia, 2016. (In Bulgarian)

[9] Z. Dinchev, Основи на проектирането на промишлена вентилация. University of Mining and Geology, Sofia, 2016. (in Bulgarian)

[10] "Wireless Round-the-Clock Observation of Coal Stacks, No Matter where Measurement Points May Move." 2016, https://www.yokogawa.com/es/library/resources/applicatoin-notes/wireless-round-the-clock-observation-of-coal-stacks-no-matter-where-measurement-points-may-move/.

[11] Y. Anastasova, "Internet of Things in the mining industry – security technologies in their application." Sustainable Extraction and Processing of Row Materials Journal, **1**, 2020, 7-10, doi: 10.5281/zenodo.4275895

[12] I. Nikolov, "Development of a warehouse microclimate managing system." Innovation and entrepreneurship, **6**(3), 2018, 166-174

[13] S. Velinova, "Possibilities for using of mining galleries for growing plants on artificial lighting." International Scientific Conference UNITECH 2017, 148-153

[14] N. Kolev, Evaluation of main elements of the soil energetic balance by electronic means. Doctor of Science thesis, Sofia, 1996.

[15] N. Kolev, et al., "Organisation of long-term microcomputer systems evaluation of field water balance elements." Proceedings of the International

Conference Energy efficiency and agricultural engineering, Russe, 2006, 347-353
[16] Z. Nenova, G. Georgiev, St. Ivanov. "Computer based system for monitoring of the air parametars in the closed premises." Russe University`s scientific publications, **54**(3), 186-190, 2015 (in Bulgarian)
[17] A. Levi, R. Ivanov, I. Bogdanov, N. Kolev. "Microcomputer systems network for plant growing management - approaches and structure." Electrotechnica & Electronica (E+E), 2013, **48**(7-8), 12-16, doi: 10.1007/s10973-019-08321-6.
[18] N. Nakayama, K. Misumi H., Koike T. TEMPERATURE SENSOR AND TEMPERATURE SENSOR SYSTEM, United States Patent, Patent No US 8,419,275 B2, Apr. 16, 2013
[19] C.M. Gerard, Smart temperature sensors and temperature sensor systems, Smart Sensors and MEMs (Second Edition), Woodhead Publishing, 2018, doi: 10.1016/B978-0-08-102055-5.00003-6.
[20] D. Zhu, C. Koh. TEMPERATURE SENSOR. United States Patent, Patent No US 10,378,969 B2, Aug. 13, 2019
[21] C.R. Carrigan, TRIAXIAL THERMOPILE ARRAY GEO-HEAT-FLOW SENSOR, PATENT-US-A7516399, 1992, DE92 004606
[22] G. Cortellessa, et al. "Experimental and numerical analysis in heat flow sensors calibration." J Therm Anal Calorim **138**, 2901 – 2912, 2019. doi: 10.1007/s10973-019-08321-6.
[23] T.C. Codau, et al., "Embedded textile heat flow sensor characterization and application." Sensors and Actuators A: Physical, **235**, 2015, 131 – 139, doi: 10.1016/j.sna.2015.10.004.
[24] E. Onofrei, et al., "Textile sensor for heat flow measurement." Textile Research Journal, **87**(2), 165-174, doi: 10.1177/0040517515627167.
[25] P. Wolkoff, "Indoor air humidity, air quality, and health – An overview." International Journal of Hygiene and Environmental Health, **22**1(3), 2018, 376 – 390, doi: 10.1016/j.ijheh.2018.01.015.
[26] V.D. Matko, "Sensor for high-air-humidity measurement." in IEEE Transactions on Instrumentation and Measurement, **45**(2), 561 – 563, April 1996, **doi:** 10.1109/19.492787
[27] M. Transfer, "Dynamik Modelling of Indoor Air Humidity. (PDF) Dynamic Modelling of Indoor Air Humidity (researchgate.net) Accessed April 2021
[28] G. Ganchev, et al., Оценка на заслепяване и намаляване на усещането за заслепяване. University of Mining and Geology - Sofia. 2017. (in Bulgarian)
[29] C Eng Faruk Bin Poyen. Luxmeter, showing light intensity, as it is received from a sensor. January 2017, (PDF) Luxmeter displaying light intensity as received by sensor (researchgate.net), Accessed April 2021
[30] S. Baev, "Спектрорадиометър в състав програмно-апаратного комплекса управления облучательными установками в закрытом грунте." SVETOTEHNIKA, SPECIAL ISSUE, 2019, 55-58, 2019.
[31] R. Nikam, et al., "Tin Doped Gas Sensors in Semiconductor Metal Oxide Form and Their Scientific Applications: A Review." J. Biol. Chem. Chron. 2019, **5**(3), 53-56, 2019
[32] Y. Wang, "ZnO Nanorods for Gas Sensors. Nanorods and Nanocomposites. Morteza Sasani Ghamsari and Soumen Dhara, IntechOpen, 2020 DOI: 10.5772/intechopen.85612.
[33] M. Miandehy, Nano graphen gas sensor (performance and applications). January 2021
[34] V. Broek, Jan; Weber, Ines C.; Güntner, Andreas T.; Pratsinis, Sotiris E. "Highly selective gas sensing enabled by filters." Materials Horizons, **8**(3), 661-684, 2019, doi: 10.1039/D0MH01453B.
[35] J. Miller, Proof-of-concept: The day after. Mentor®. A Siemens Business. 2018. www.mentor.com
[36] W. Ruh, "Drilling Deep into Digital Industrial Transformation will determine who survives and thrives." – Internet of Things Magazine, September 2018.
[37] C. Perera, et al., "Designing the Sensing as a Service Ecosystem for the Internet of Things." – Internet of Things Magazine, December 2018, 1, 2.
[38] R. Gonzalez, et al., "Navigation technics for mobil robots in greenhouses." Applied Engineering in Agriculture. 2009, Vol. 25, No. 2, pp. 153-165
[39] roboticsbiz.com. Drones in underground mines – Applications and benefits, Drones in underground mines - Applications and benefits (roboticsbiz.com) September 28 2020, Accessed April 2021
[40] J. Shahmoradi, et al., "A Comprehensive Review of Applications of Drone Technology in the Mining Industry." Drones. 2020. 4. DOI: 10.3390/drones4030034.
[41] University of Utah. A Safer Battery to Power Drones in Underground Mines. A Safer Battery to Power Drones in Underground Mines | Technology Org, April 9, 2021

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 401-405 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Quantum Secure Lightweight Cryptography with Quantum Permutation Pad

Randy Kuang*, Dafu Lou, Alex He, Alexandre Conlon

Quantropi Inc., Ottawa, K1Z 8P9, Canada

ARTICLE INFO

Article history:
Received: 18 June, 2021
Accepted: 17 August, 2021
Online: 28 August, 2021

Keywords:
AES
Quantum Permutation Gates
Quantum Permutation Pad
Permutation Matrix
Quantum Algorithm
QPP
Shannon entropy
Lightweight Cryptography
Streaming Cipher
Block Cipher

ABSTRACT

Quantum logic gates represent certain quantum operations to perform quantum computations. Of those quantum gates, there is a category of classical behavior gates called quantum permutation gates. As a quantum algorithm, quantum permutation pad or QPP consists of multiple quantum permutation gates to be implemented both in a quantum computing system as a quantum circuit operating on n-qubits' states for transformations and in a classical computing system represented by a pad of n-bit permutation matrices. Since first time proposed in 2020, QPP has been recently applied to create a quantum safe lightweight block cipher by replacing SubBytes and AddRoundKey with QPP in AES called AES-QPP. In AES-QPP, QPP consists of 16 selected 8-bit permutation matrices based on the shared classical key materials. For quantum safe, the key length can be any size from 256 bits to 4 KB. That means, this QPP holds up to 4 KB of Shannon information entropy. Its code size is less than 2 KB with 4 KB of RAM memory. In this paper, we propose to apply QPP for a streaming cipher and carry out its encryption performance and the randomness analysis of this streaming cipher. The proposed QPP streaming cipher demonstrates not only good randomness in its ciphertexts but also huge performance improvement: 13x faster than AES-256, with an overall runtime space (6.8 KB).

1. Introduction

Since the U.S. National Institute of Standards and Technology (NIST) announced the standardization of Advanced Encryption Standard or AES in 2001 [1], AES has been widely accepted as secure data encryption for data in transit or at rest. As a standard block cipher, AES accepts a fixed block size of 128 bits for three key lengths: 128, 192, and 256 bits with 10, 12, and 14 rounds respectively. Each AES round includes four steps: SubBytes, ShiftRows, MixColumns, and AddRoundKey. Over the past decade, the internet of things or IoT has captured the great attentions cross the world due to its potential to transform our daily lives through varieties of aspects such as smart home, smart city, autonomous vehicles, connected devices, etc. IoT devices are generally considered as resource constrained systems. They are often battery-powered, low computing power, and limited storages. These limitations put certain pressures on the standard AES to run in IoT devices, especially with high security requirements. The authors published their NIST report on lightweight cryptography [2], covering lightweight bloch ciphers, lightweight hash functions, lightweight message authentication codes, and lightweight streaming ciphers. In [3], the authors proposed their implementation of modified lightweight AES in FPGA, with a parallel manner for achieving better latency.

On the other hand, varieties of symmetric lightweight cryptographic algorithms have been proposed. In 2018 [4], the author have reviewed those algorithms to benchmark them on executing time, RAM memory and binary code sizes. those algorithms support the block sizes from 64 bits to 128 bits with key lengths from 80 bits to 128 bits.

AES generally faces three types of attacks: differential, linear, and integral [5]-[8]. The single static S-box representing substitutions or non-linear-transformations enables the differential analysis attacks due to some characteristic of XOR differences between input blocks and output blocks, especially impossible differences found at round 4, also called impossible differential attacks [6,7]. The differential analysis attack can be further improved with sets or multisets of input and output XOR results to create a new integral attack [8]. AddRoundKey at the end of each round contributes the linear analysis attack due to the linear transformation between rounds.

*Corresponding Author: Randy Kuang, randy.kuang@quantropi.com

https://dx.doi.org/10.25046/aj060445

In 1994, Shor proposed an algorithm to use quantum systems or qubits to perform computations called quantum computing [9]. Shor's algorithm enables a new natural parallel computing mechanism arising from the fundamental characteristic of their superpositions. With quantum computers, the classical exponential difficulty of prime factorizations becomes polynomial time, shaking the foundation of classical public key cryptography. The recent advancements in quantum computing development speeds up the urgency of quantum safe cryptography for both asymmetric and symmetric. In September 2019, Google announced their 54-qubit quantum computer called "Sycamore", marked their quantum supremacy [10]. On the other hand, in [11] the author made a milestone achievement in prime factorization with D-Wave's annealing quantum computer.

In 1996, Grover proposed his new search algorithm by using quantum computing mechanism called Grover's algorithm [12]. Grover's algorithm can achieve a square root complexity $O(\sqrt{n})$ in the unstructured search problem of size n, while any classical algorithm needs $O(n)$ queries. The squared searching power from the Grover's algorithm requires any symmetric cryptography not only to double the key length from 128 bits to 256 bits, but also to be true random. For resource constrained IoT devices, standard AES cryptography itself is already heavy so the doubled key length requires extra four rounds from 10 rounds to 14 rounds. It is necessary to explore different lightweight cryptographic algorithms in the post-quantum era. In [13], the authors proposed to use quantum cryptography with One-Time-Pad or OTP for secure encryption for power grid data. This can be considered as a hybrid quantum secure encryption combining QKD with OTP. In [14], the authors proposed a new lightweight symmetric cryptographic algorithm called Saturnin by introducing two representations of AES 256-bit internal states: the 2-dimensional and 3-dimensional notations. In its 2-dimensional representation, a 256-bit state can be expressed by sixteen 16-bit registers; but in its 3-dimensional representation, it is expressed by a 4x4x4 cube of nibbles. Through those major changes, Saturnin established a new quantum resistant lightweight cryptography for 128-bit and 256-bit block ciphers and a 256-bit hash function.

Quantum mechanics allow us to have two implementations of quantum gates: physically for quantum computing power and digitally for quantum security. In [15], the authors first attempted to present classical information quantum mechanically with a state denoted by a Dirac ket over a quantum computational basis. The well-known symmetric group S_n containing entire group actions over a set of n items has its matrix representations or permutation matrices over the corresponding quantum computational basis. The extremely large size of the quantum permutation group, $2^8!$ (factorial), for an 8-qubit quantum computational basis holds huge equivalent Shannon information entropy, desirable for information security. For an 8-bit system, the permutation group is S_{256}. Kuang and Bettenburg [15] extend the Shannon perfect secrecy of the classical one-time-pad (OTP) over GF(2^n) [16], to their proposed quantum permutation pad (QPP) over a quantum computational basis. In contrast to the one-time-use nature of OTP, QPP retains the Shannon perfect secrecy over multiple uses, thanks to the general non-commutativity properties of the symmetric group.

In [17], the authors have applied QPP for a lightweight block cipher called AES-QPP, with 16 permutation matrices selected by using the shared classical random key to replace both SubBytes and AddRoundKey. AES-QPP has a footprint below 2KB and RAM memory 4KB, with performance improvement about 3x. In this paper, we extend the work [17] further for a quantum safe lightweight streaming cipher.

This paper is organized as follows: Section 2 is for the summary of lightweight block cipher AES-QPP. Then Section 3 describes the proposed streaming cipher. The randomness and performance of the proposed streaming cipher is presented in Section 4. We will draw a conclusion at the end.

2. Quantum safe Lightweight Block Cipher AES-QPP

2.1. Quantum Permutation Pad

QPP is a pad of quantum permutation matrices randomly selected from the n-bit permutation group [15, 17] as shown in Fig. 1 for AES with 16 8-bit permutation matrices. They are all 2^8x2^8 square matrices and they are unitary and reversable so their reverse transformations are their transposes. This characteristic is great for lightweight cryptography, especially for resource constrained IoT devices. At the encryption side, we can use the selected QPP and then at receiving side we use their transposes or QPP^T. Therefore, transformations are exactly symmetric with the same computational performance for encryption and decryption. The QPP selection is a process to map classical key materials into a QPP pad over the 8-bit computational basis. There are several algorithms to be used such as RC4 key scheduling algorithm or Fisher-Yates random shuffling algorithm. We use the Fisher-Yates algorithm to map classical key into a QPP pad as shown in

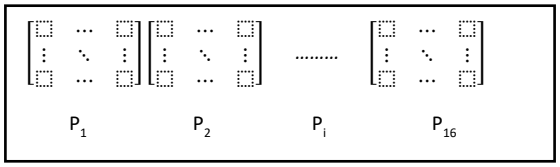


Figure 1: QPP is illustrated.

Algorithm 1.

Algorithm 1: Pseudo code of mapping a key to permutation matrix

```
Result: Permutation matrix P[256][256]
Input: 256 bytes of random key k[256]
Initialization: set S[256] and P[256][256] to 0
for i = 0 to 255
    S[i] = i;
end for
i = 255
while i-- > 1 do
    j = k[i];
    swap S[j] and S[i];
end
for i = 0 to 255
    P[i][S[i]] = 1;
end for
```

For each permutation matrix, we need 256 bytes of random numbers as shown in the pseudo-code. For a QPP pad with 16

permutation matrices, we would need to supply total 4 KB of random numbers. That means, the selected QPP holds total 4 KB of entropy. It is up to the desired security level to choose a right key length from 256 bits to 32,768 bits. To support this variable key length, a key scheduling is required to extend a variable key length to 4 KB.

2.2. AES-QPP

SubBytes in AES performs a substitution with a static S-box which is a 16x16 matrix. S-box can be converted to a 256x256 permutation matrix. The substitution can be considered as the permutation matrix multiplication with an 8-bit state vector over the 8-bit computational basis. AddRoundKey step performs byte-by-byte XOR operations between the output block from MixColumns step and a round key. XOR operation is a special case of permutation transformations. In our early block cipher, we use the QPP to replace both SubBytes and AddRoundKey in an AES round with the ShiftRows and MixColumns steps as follows:

1. 16 8-bit QPP;
2. ShiftRows;
3. MixColumns;
4. the same QPP as in the step 1.

We illustrate AES-QPP in Figure 2. In standard AES, each byte in a16-byte block is supplied to the single static S-box, but in AES-QPP, each byte in a block is supplied to its corresponding permutation matrix. The last step is performed in the same way as in the first step, unlike the standard AES in the AddRoundKey step with each byte from MixColumns to be XORed with the corresponding byte in a round key. This design of AES-QPP enables us to use variable key length without change the implementation of the cryptography.

The decryption is the same process as in the encryption with transposed QPP^T.

2.3. AES-QPP Rounds

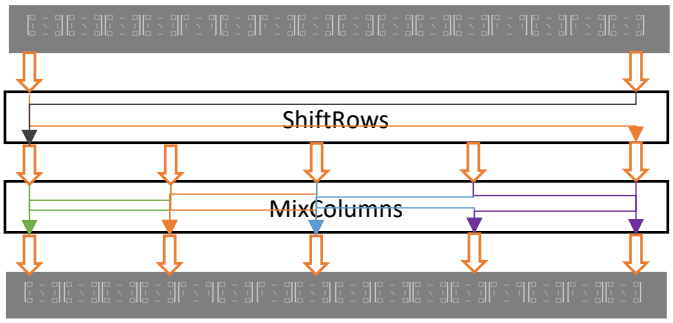


Figure 2: The proposed AES-QPP is illustrated

In comparison with AES rounds, QPP increases the diffusion capability at least 16x with 16 permutation matrices. This extra strengthened diffusion ability helps to reduce the number of rounds. AES-256 needs 14 rounds to achieve good randomness in ciphertexts. In our early report, the number of rounds in AES-QPP was reduced to 5 rounds. The ciphertext still demonstrates excellent randomness from NIST and ENT random testing suites.

2.4. Cipher Randomness with Shannon Entropy Distribution

Cipher randomness is a good measure for a cryptosystem to avoid statistical analysis attacks. We have demonstrated randomness analysis with NIST random test suite, especially with ENT testing tool to identify any byte level and bit level biases. AES-QPP shows excellent randomness in its ciphertexts. In NIST testing suites, AES-QPP ciphers with 5 rounds pass all 15 testing cases. In the sensitive testing suite ENT, the AES-QPP ciphers not only pass 6 testing cases but also demonstrate excellent Chi Square value, arithmetic means, Monte Carlo $\boldsymbol{\pi}$, as well as serial correlation. Here we want to add analysis for Shannon entropy distribution for AES-QPP in comparison with the standard AES. Shannon entropy distribution performs analysis of each 16-bit, entropy per 16-bit random data, as well as how close to the Gaussian distribution.

Figure 3 plots the Shannon entropy distribution for AES-QPP with 5 rounds, using a 16 bits Shannon Entropy calculator [18]. . The entropy is 15.999072 per 16 bits of ciphertexts, very close to the ideal entropy 16 bits. The left side of the graph displays the frequency of each 16-bit integer. The graph displays a nice symmetric behavior around an average count. The analysis shows that the median counts are 781, minimum count 656, and maximum count 896 for 100 MB files. The right-hand side of Fig. 3 shows a good Gaussian distribution with a nice symmetric shape around the median count, indicating a good randomness in AES-QPP-5 ciphertexts.

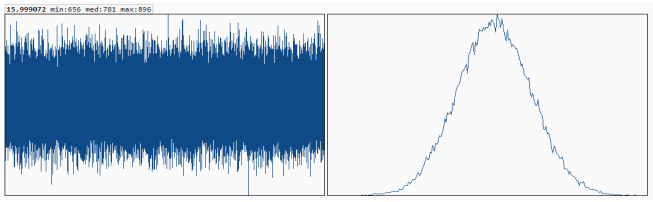
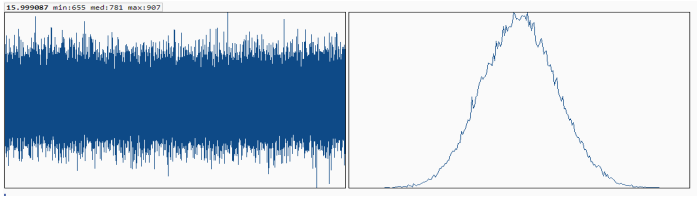


Figure 3: Shannon entropy distribution is plotted for AES-QPP-5.

Figure 4 plots the Shannon entropy distribution for AES-256. The same size of AES-256 ciphertext file as AES-QPP-5 is used. It demonstrates a very close relationship to Figure 5 with median counts 781, minimum counts 655 comparing with 656 in AES-QPP, and maximum counts 907, slightly better symmetry than AES-QPP. The Shannon entropy of AES-256 is 15.999087 per 16 bits of ciphertexts, extremely close to AES-QPP-5.



3. Quantum Safe Lightweight Streaming Cipher with QPP

In the last section, we discussed the block cipher implementation with QPP. We can also implement it in a streaming cipher with a pre-randomized dispatcher as shown in Figure 5.

Seed is an input classical key material, Init box is the scheduling to map the classical key material into a QPP pad as shown in Algorithm 1. PRNG box is a pseudo random number generator used to pre-randomize input plaintexts with a directly XOR operation before dispatching them to QPP. Dispatcher box also takes a PRNG byte and performing a 4-bit right shift as an index to the permutation matrix inside the QPP pad. A ket $|m\rangle$ represents a plaintext byte. The ciphertext denoted by a ket $|c\rangle$ can be created byte-by-byte in the same way as input plaintexts. At the receiving side, the same shared classical key material is used to establish the same QPP pad in a transposed mode because permutation matrix is unitary and reversable. The decryption is straightforward.

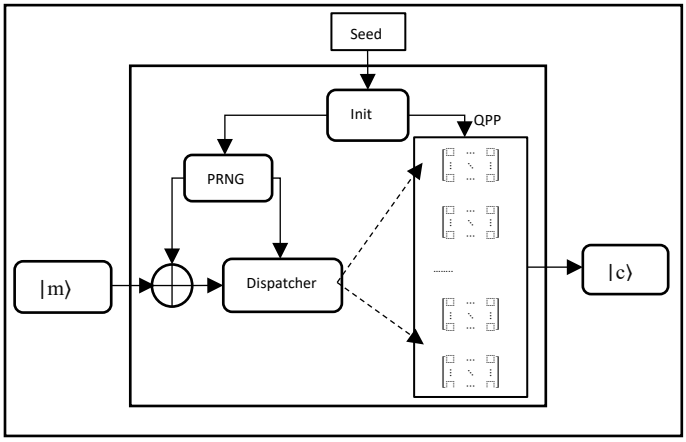


Figure 5: A streaming cipher is illustrated

QPP streaming cipher implementation eliminates the repeated rounds for better randomness ciphers, replacing with a pre-randomized process under a consideration of bijective quantum permutation transformations. We can roughly estimate the executing time for the encryption at the level of a single round AES encryption. This would dramatically improve its performance for latency and battery consumptions.

The footprint of QPP streaming cipher is about 2.5 KB, In comparison with 1.1 KB in AES-QPP because of the pre-randomization process in QPP streaming. RAM memory is the same as AES-QPP at 4KB because we still use a QPP pad with 16 permutation matrices.

4. Discussions of QPP Streaming Cipher

For the proposed QPP streaming cipher shown in Figure 5, we create a plaintext file of 120 MB by a paragraph of English sentences. Then QPP streaming encrypts the plaintext file and stores the ciphertexts into ciphertext file. The ciphertext file is passed all NIST 15 randomness test cases. We should be very interesting to see the ENT testing reports listed in Table 1, together with the results for the input plaintext file. The plaintext file comes with 4.49 bits of entropy per 8-bits, huge Chi Square value, totally wrong arithmetical mean, as well as a wrong Monte Carlo π value. All those results show that the input plaintext file is totally biased. The ciphertext file produced from our proposed QPP streaming cipher demonstrates excellent randomness: 8 bits of entropy per 8 bits of ciphertexts, Chi Square value 237.64 with a p-value 0.775, arithmetical mean 127.49 compared to 127.50, very nice Monte Carlo π value 3.141771788 compared to 3.14159265, and serial correlation value 8.2×10^{-5}. The overall ENT testing results from QPP streaming cipher is very similar to AES-QPP-5 [17].

Table 1: ENT randomness testing reports for QPP streaming cipher and plaintext files of size 120 MB

ENT	Plaintext	QPP Streaming
Entropy (bits)	4.491422	7.999999
Chi Square	1736817422.90	237.64
p-Value	0.0001	0.775
Arith. Mean	95.7060	127.49
Monte Carlo π	4.000000000	3.141771788
Serial Corr.	0.047905	0.000082

Figure 6 plots the Shannon Entropy distribution with 120 MB ciphertext file from a streaming encryption implementation of 16 permutation matrices. The Shannon entropy is 15.999244 bits per 16 bits of ciphertexts. The distribution demonstrates a nice Gaussian type with a med 957, minimum 826 and a maximum 1085, slightly better symmetry than AES-QPP-5 [17].

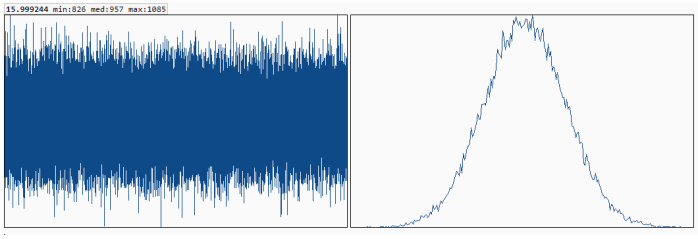


Table 2 illustrates performance comparisons among AES-256, AES-QPP-5, and QPP streaming cipher. For AES-256, we take the open-source implementation or Tiny-AES. AES-256 demonstrates a fastest key scheduling with 0.01 ms, then AES-QPP-5 with 0.120 ms, and QPP streaming with 0.235 ms, respectively. It is understandable that QPP initialization takes longer time because it processes 4KB key materials to select 16 permutation matrices, unlike in AES where key is only 32 bytes long. Also, in QPP streaming cipher, we take a special handling to increase confusion capability. This special handling would help to produce totally different QPP pad even with a single bit change in the supplied key materials.

Table 2: Performance comparisons are tabulated for key scheduling and encryptions for AES-256, AES-QPP-5, and QPP streaming cipher. Encryption speeds are tested with 16 bytes blocks in the same computer: MacBook Pro, 2.6 GHz 6-Core Intel Core i7.

	AES-256	AES-QPP-5	QPP Streaming
Key Schedule	0.01 ms	0.120 ms	0.235 ms
Encryption MB/s	51.3	115.1	672.3
Ratio	1.0	2.24	13.1
Code Footprint	11.5 KB	1.39 KB	2.5 KB
Run Time Space	12.0 KB	5.39 KB	6.5 KB

It can also be seen from Table 2 that AES-256 is the slowest among three ciphers with an encryption speed 51.3 MB/s for 16 bytes blocks, then AES-QPP is faster than AES-256 with 115.1 MB/s, finally QPP streaming cipher is the fastest cipher with 672.3 MB/s. AES-QPP is 2.24x faster than AES-256, slightly slower than what we expected 2.8x, that may be the fact that overall permutation transformations with 16 permutation matrices take longer than each step of SubBytes and AddRoundKey in AES. However, the discrepancy is very acceptable. QPP streaming cipher is 13x faster than AES-256, indicating that pre-randomization with randomly dispatching together is almost equivalent to a single round in AES. In comparison with AES-QPP-5, QPP streaming cipher is 5x faster than AES-QPP-5.

In comparison with code sizes, compiled footprints are 11.5 KB, 1.39 KB, and 2.5 KB for AES-256, AES-QPP, and QPP streaming respectively, and runtime memory spaces are 0.47 KB for AES-256, 4 KB for both AES-QPP and QPP streaming cipher. As for overall runtime space, AES-QPP takes the least runtime space at 5.39 KB, then QPP streaming cipher at 6.5 KB, after then AES-256 needs the most runtime space at 12 KB.

5. Conclusion

We have applied quantum permutation pad or QPP to establish both lightweight quantum safe block cipher and streaming cipher. In a block cipher implementation, QPP replaces both SubBytes and AddRoundKey in a standard AES or called AES-QPP. In addition to cipher randomness analysis in [17], we perform the Shannon entropy distribution for a more complete randomness analysis of this quantum safe block cipher.

In this paper We explored the QPP algorithm for a streaming cipher with a straightforward pre-randomization and random distribution process. The randomness analysis of the QPP streaming cipher demonstrates very good randomness, especially with ENT randomness testing tool. The very promising encryption speed plus overall memory space makes QPP streaming be a good candidate for quantum safe lightweight streaming cipher and AES-QPP for quantum safe lightweight block cipher.

In the future, we may extend the exploration both ciphers for 4-bit permutation matrix pad to further reduce their runtime memory spaces.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Information Technology Laboratory (National Institute of Standards and Technology), Announcing the ADVANCED ENCRYPTION STANDARD (AES), Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology Gaithersburg, MD 2001.

[2] K. McKay, L. Bassham, M. Sonmez, N. Mouha, Report on Lightweight Cryptography, NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD, [online], 2017, doi:10.6028/NIST.IR.8114 (Accessed August 23, 2021).

[3] M. James, D. S. Kumar, "An Implementation of Modified Lightweight Advanced Encryption Standard in FPGA，" Procedia Technology, **25**, 582-589，2016, doi:10.1016/ j.protcy.2016.08.148.

[4] D. Dinu, Y. L. Corre, D. Khovratovich, L. Perrin, J. Großschädl, A. Biryukov, "Triathlon of lightweight block ciphers for the Internet of things," Journal of Cryptographic Engineering, **9** (3), 283–302, 2018, doi:10.1007/s13389-018-0193-x. S2CID 1578215.

[5] J. Daemen, V. Rijmen, The Design of Rijndael, *AES - T*he Advanced. Encryption Standard, Springer-Verlag 2002.

[6] H. M. Heys, S. E. Tavares, "Substitution-permutation networks resistant to differential and linear cryptanalysis," Journal Cryptology **9,** 1–19, 1996. doi:10.1007/BF02254789.

[7] L. O'Connor, On the distribution of characteristics in bijective mappings, J. Cryptology **8,** 67–86, 1995, doi:10.1007/ BF00190756.

[8] J. Lu, O. Dunkelman, N. Keller, J. Kim, "New Impossible Differential Attacks on AES," In: Chowdhury D.R., Rijmen V., Das A. (eds) Progress in Cryptology - INDOCRYPT 2008. INDOCRYPT 2008. Lecture Notes in Computer Science, **5365**, Springer, Berlin, Heidelberg. doi：10.1007/978-3-540-89754-5_22.

[9] K. Autre, K. Arya, , *et al*， "Quantum supremacy using a programmable superconducting processor，" Nature **574** (7779)，505–510. 2019.

[10] P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring，" in Proceedings 35th Annual Symposium on Foundations of Computer Science, IEEE Computer Society Press: 124–134, 1994.

[11] B. Wang, F. Hu, H. Yao, *et al.*, "Prime factorization algorithm based on parameter optimization of Ising model," Sci Rep **10,** 7106, 2020, doi:10.1038/s41598-020-62802-5

[12] L. Grover, "A fast quantum mechanical algorithm for database search," In: Proceedings of the 28th ACM STOC, Philadelphia, Pennsylvania, 212–219, ACM Press, 1996.

[13] Y. Li, P. Zhang and R. Huang, "Lightweight Quantum Encryption for Secure Transmission of Power Data in Smart Grid," in IEEE Access, **7**, 36285-36293, 2019, doi: 10.1109/ACCESS.2019.2893056.

[14] A. Canteaut, S. Duval, G. Leurent, M. Naya-Plasencia, L. Perrin, T. Pornin, A. Schrottenloher, "Saturnin: a suite of lightweight symmetric algorithms for post-quantum security," IACR Transactions on Symmetric Cryptology, **2020**(S1), 160-207, doi:10.13154/tosc.v2020.iS1.

[15] R. Kuang, N. Bettenburg, "Shannon Perfect Secrecy in a Discrete Hilbert Space," 2020 IEEE International Conference on Quantum Computing and Engineering (QCE), Denver, CO, USA, 249-255, 2020, doi: 10.1109/QCE49297.2020.00039.

[16] C. E. Shannon, "Communication Theory of Secrecy Systems?" Bell System Technical Journal， **28** (4): 656–715， October 1949.

[17] R. Kuang, D. Lou, A. He and A. Conlon, "Quantum Safe Lightweight Cryptography with Quantum Permutation Pad," 2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS), 790-795, 2021, doi: 10.1109/ICCCS52626.2021.9449247.

[18] Server Test, https://servertest.online/entropy.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 406-413 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Personalized Clinical Treatment Selection Using Genetic Algorithm and Analytic Hierarchy Process

Olena Nosovets, Vitalii Babenko*, Ilya Davydovych, Olena Petrunina, Olga Averianova, Le Dai Zyonh

Department of Biomedical Cybernetics, Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, 03056, Ukraine

ARTICLE INFO

Article history:
Received: 13 July, 2021
Accepted: 24 August, 2021
Online: 28 August, 2021

Keywords:
Clinical Treatment
Genetic Algorithm
Analytic Hierarchy Process
Modeling

ABSTRACT

The development of Machine Learning methods and approaches offers enormous growth opportunities in the Healthcare field. One of the most exciting challenges in this field is the automation of clinical treatment selection for patient state optimization. Using necessary medical data and the application of Machine Learning methods (like the Genetic Algorithm and the Analytic Hierarchy Process) provides a solution to such a challenge. Research presented in this paper gives the general approach to solve the clinical treatment selection task, which can be used for any type of disease. The distinguishing feature of this approach is that clinical treatment is tailored to the patient's initial state, thus making treatment personalized. The article also presents a comparison of the different classification methods used to model patient indicators after treatment. Additionally, special attention was paid to the possibilities and potential of using the developed approach in real Healthcare challenges and tasks.

1. Introduction

This paper is an extension of work originally presented in the 15th International Conference on Computer Sciences and Information Technologies, held in Zbarazh (Ukraine) in September 2020 [1].

It is well known that human treatment is a delicate moment, as any wrong decision can radically affect the person's state [2], for which doctors will be primarily responsible [3]. The word 'treatment' can mean different things, and it should be understood that not all treatments are equal. For example, treating a person for influenza [4], wherein most cases it suffices to prescribe a few medications for a couple of weeks, is not equivalent to treating cancer [5], which can last for years [6-7]. Undoubtedly, this does not mean that influenza poses no risk to human health [8], but if influenza progresses to a critical stage, medical supervision and prescription of clinical treatment are mandatory [9-11].

Clinical treatment refers to the process during which the patient stays in a medical institution under the strict supervision of clinicians and specialists, and undergoes all stages of therapy (including experimental treatments) to eliminate the symptoms of disease or complications. One of the major objectives in Healthcare is to treat the disease itself (not the symptom), so treatment must be chosen carefully and without hesitation. It is therefore the regular practice to use clinical protocols (guidelines) [12-14] while prescribing treatment.

Briefly, clinical protocols are systematically developed statements, which assist clinicians and patients in making decisions about appropriate treatment for specific conditions based on the best scientific evidence at the time of development [15-17]. These protocols are the primary medico-technological documents that specialists must follow in any given clinical situation, choosing the most effective solution to cure a patient. A clinical protocol is a manual for a doctor, and it contains guidelines to treat a specific disease.

In Ukraine, Ministry of Healthcare allows international clinical protocols under Order no. 1422 of the by 29 December 2016, which entered into force on 28 April 2017 [18-19]. To get these protocols, the Ministry of Healthcare has signed an agreement with "Duodecim Medical Publications Ltd", a Finnish medical-scientific company specializing in comprehensive solutions for evidence-based medicine [18]. Since then, about a thousand international clinical protocols in English have been available online for registered users in Ukraine. Using new clinical protocols in medical practice has become one of the most important ways of implementing evidence-based medicine in Ukraine [19].

Many researchers are now focusing on clinical protocols to model the clinical treatment process [10] because these protocols do not guarantee a complete cure for the patient. The reason is that doctors do not have full knowledge of medicines, their types, and their correct use. This responsibility lies with clinical pharmacists (or clinical provisors) [20-21]. Specialist training in this field in

*Corresponding Author: Vitalii Babenko, vbabenko2191@gmail.com

Ukraine was established in 1999 by the National University of Pharmacy in Kharkiv [22]. Despite more than 20 years of training such specialists, the country simply does not have enough clinical pharmacists, which may be due to the limited financial resources of local medical institutions. This problem, and the rapid digitization of Healthcare in Ukraine [23], make it an urgent task to develop an automated decision support system for clinical treatment selection.

The making of decision support system was chosen because the selection of optimal clinical treatment should not be made solely by a machine. Medical staff is responsible for causing harm to humans [24], which is why Artificial Intelligence (AI) software must be used with utmost care for these challenges.

The basic requirements for such systems are:

- Usage of clinical protocol standards as a basis for decision support in treatment selection.
- As set out in right no. 12 of the European Charter of Patients' Rights [25], the individual characteristics of the patient must be considered.
- The availability of AI and Machine Learning (ML) techniques to provide high-quality assistance for medical staff.

This research aims to develop a general approach to selecting optimal clinical treatment based on the requirements listed above. The first thing to consider is how treatment outcomes are assessed, as that enables AI- and ML-based applications to improve efficiency in decision making.

2. Assessing the Quality of Treatment: An Overview

2.1. Scales of Patient Severity

The quality of clinical treatment should be assessed foremost by results, bearing in mind that the aim of providing medical care is to ease the patient's state. At all stages of treatment in the various fields of medicine, the determination of a patient severity objective assessment is essential for clinical decision-making. The close association of state severity with the prognosis of mortality risk further extends the application of such tools at different stages of care. For example, intensive care units (ICU) must provide prognoses for patients within the first 24 hours of admission.

The stratification of patients into risk groups according to the severity of their state is necessary to compare treatments and their quality, establish differences between different medical institutions, and evaluate the results of clinical trials in evidence-based medicine.

Establishing a prognosis comprises estimating the probability of death using indicators that are commonly used to diagnose and treat critically ill patients [26]. Severity scales are the classic tools used to establish such prognoses.

Rating systems quantify or qualify the severity of a state and classify the patient into specific risk groups, based on the analysis of anatomical, physiological and biochemical abnormalities [26-27]. Over two dozen severity scales have been developed, but only some of them can be considered universally accepted.

The most commonly used (in ICUs of USA and EU countries) scoring systems for assessing patient severity are: SAPS II, APACHE II and III, GCS, MPM II, SOFA, MODS, and LODS [28].

2.2. Models of Patient Severity

Most morbidity estimation models are based on Logistic Regression [29]. Authors of [30] analyzed papers that use SOFA models to predict mortality in ICUs. Only ten studies (56%) applied logistic regression models, and five of them had validated models with independent tests. The following models were also considered: combined with other assessment scales (APACHE, MOD) and additional measures of organ failure, time models (sequential SOFA scores), and automatically detected from the data SOFA templates. For example, the predictive ability of APACHE II was assessed using Logistic Regression analysis. For the Logistic Regression model based on the APACHE II score, the AUC of the ROC curve [31] was 0.863. Authors note that although there is heterogeneity across studies, it is impossible to say which SOFA-based model is optimal.

Decision Tree [32] methods have recently become more widespread in medical research. Clinical practitioners swallow them because they are illustrative and can turn into logical conditions (classification rules). Classification Trees have been used in critical situations, e.g., to calculate the probability of death from coronary pathologies [33], intracerebral hemorrhage [34] or craniocerebral injury [35], to predict persistent autonomic states [36], and to stratify patient groups by the likelihood of mortality in the general population of ICUs [37-38].

In [39], the author predicted the probability of hospital mortality using three Decision Tree classification algorithms: CART, CHAID, and C4.5. All models are based on estimating the severity of patients within the first 24 hours of admission only (2864 patients, 70:30). Authors of [39] point out that the chief advantages of Decision Trees are that the resulting decision rules can be easily interpreted and the composition of the patient group obtained at each final node of the tree is relatively homogeneous. All Decision Tree models achieved the AUC of 0.75-0.76, which was close to the AUC for APACHE II (0.77) but lower than the Logistic Regression AUC (0.81).

It is worth noting that such multidimensional models are not designed to handle streaming data. Modern morbidity estimation models are based on ML methods such as Support Vector Machine (SVM) [40], Bayesian models [41], Artificial Neural Network (ANN) [42], etc.

The predicted length of stay (LOS) [43], which is based on monitoring data, is seen as a target that helps plan resuscitation resources and make ICU care individualized.

2.3. Information Systems of State Assessment and Forecasting

INTCare [44] is an intelligent decision support system for intensive care medicine. It is a system based on both collecting data from monitors at the bedside and updating the model, reducing the need for human intervention. INTCare currently provides predictions about organ failure and the likelihood of in-hospital death. Reliable prognostic results contribute to improving the

quality of service. The system presents functional and structural aspects. It aims to automate the knowledge discovery process.

The most important feature of INTCare's intelligent decision support system is the ability to operate autonomously and in real-time. Two approaches were used to model and predict 2 targets (survival and LOS):

- Collected data and physiological features during the first 24 hours of inpatient treatment.
- Collected patient clinical data in real-time.

In the first approach, the achieved results were poor (73% accuracy). However, when predictions (dwell time) were made using data collected in real-time, the results became higher (model sensitivity 96.1%). Researchers in their work used these models: SVM, Decision Trees, and Naïve Bayes. For survival prediction, the Decision Trees method had the best result with a sensitivity of 87.32%.

A systematic review of the literature (2008-2018) [45], which aims to investigate the use of ML to improve patient health, analyzed 78 such studies. The conclusion is that AI techniques can analyze and learn useful standards from clinical datasets (which are stored in electronic medical records) to provide better evidence for supporting health professionals' decisions.

More recently, the work has appeared that uses Deep Learning (DL) techniques, namely, Recurrent Neural Networks (RNN) with the Long Short-Term Memory (LSTM) architecture [46]. Experiments have shown that it is possible to predict vital signs in advance with good accuracy (more than 80%) to the deterioration of the patient's state. Predicting a patient's vital signs and using them to calculate a prognostic index makes it possible to predict future severe diagnoses that would not be possible using only the patient's current vital signs (50%-60% of cases were not detected).

Also noteworthy is the Ukrainian work by Nastenko et al. [47], which used Group Method of Data Handling (GMDH) [48] models and Simplex Method [49] optimization algorithm to select the treatment strategy.

3. Mathematical Background

Let's devise the general Healthcare challenge of this research. It is to find the optimal clinical treatment for the patient. In fact, in the mathematical space of objects, the patient (object) can be represented as a multidimensional vector, where the parameters are his/her indicators. In the simplest case, 2 patient states are possible in a given space: an initial state (before treatment) and a final state (after treatment). On this basis, 2 vectors of data are given for the challenge:

- $X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}^T$ – describes the initial state of the patient (x_1, x_2, …, x_n are the patient's indicators before clinical treatment).
- $Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_m \end{pmatrix}^T$ – describes the final state of the patient (y_1, y_2, …, y_m are the patient's indicators after clinical treatment).

Finding the optimal treatment involves finding the optimum of vector Y. The clinical treatment, which is applied to a patient for getting the Y-vector, can be described by a vector $I = \begin{pmatrix} i_1 \\ i_2 \\ \dots \\ i_k \end{pmatrix}^T$, where i_1, i_2, …, i_k – are the different types of drugs (the influence parameters on the patient's state). Therefore, Y-vector directly depends on vectors X and I (Fig. 1).

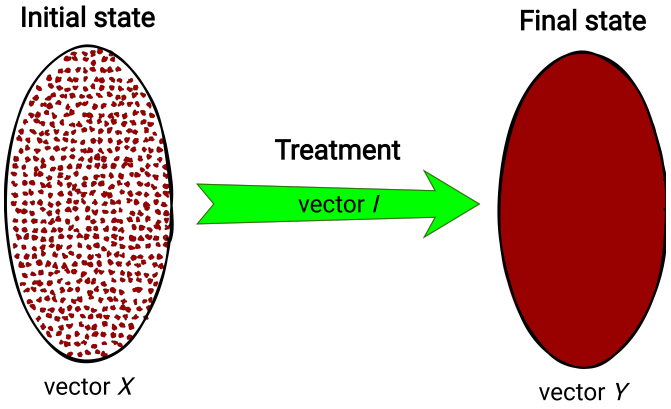


Figure 1: Visual representation of treatment process

This dependency can be described by the next general equation:

$$Y = X \times I \Leftrightarrow \begin{cases} y_1 = \sum_{j=1}^{n}\sum_{z=1}^{k} x_j i_z \\ y_2 = \sum_{j=1}^{n}\sum_{z=1}^{k} x_j i_z \\ \dots \\ y_m = \sum_{j=1}^{n}\sum_{z=1}^{k} x_j i_z \end{cases} \quad (1)$$

Thus, it can be said that $X = Y$ if no treatment is given (without regard to externalities).

Equations for $y_1, y_2, \dots, y_m$ (1) can be both linear or non-linear, parametric or non-parametric. With their usage, it seems possible to simulate (modeling) the clinical treatment process. Consequently, a multi-criteria optimization problem arises, where it is necessary to find such values of I-vector that will give the optimum of Y-vector. Values of X-vector are set by default, so the personalized solution search will be done by considering the patient's initial state indicators (this idea was proposed in [47] for a single-criteria problem). So, it is necessary to create an algorithm that will solve such a task.

4. Personalized Clinical Treatment Selection Algorithm

As mentioned earlier, finding the optimal clinical treatment is a multi-criteria optimization problem. However, creating the algorithm for solving this problem raises the following issues:

- Simultaneous optimization of the *Y*-vector parameters (patient's indicators after treatment).
- Searching the values of *I*-vector (influence parameters on the patient's state), which gives the global optimum (NP-complete problem).

The first issue can be solved using Multi-Criteria Decision Making (MCDM) methods [50]. Since in most cases the final state of a patient is described by two or more *Y*-vector parameters (in this problem – the criteria for optimization), it is worthwhile to assess the patient's state after treatment in the right way. MCDM methods allow getting a convolution of several criteria into one so-called "supercriterion". Apart from solving simultaneous optimization, this supercriterion can be used as an assessment metric to describe the final state of the patient.

One of the simplest and most easily interpreted methods of MCDM is the Analytic Hierarchy Process (AHP) [51-52], invented by Thomas L. Saaty in the 70s. This method allows getting a function of additive convolution by pairwise comparison of criteria priorities. The comparison mechanism by AHP in general form is shown in Table 1.

Table 1: The General Form of Criteria Pairwise Comparison

	y_1	y_2	…	y_m
y_1	$\frac{v_1}{v_1}$	$\frac{v_2}{v_1}$	…	$\frac{v_m}{v_1}$
y_2	$\frac{v_1}{v_2}$	$\frac{v_2}{v_2}$	…	$\frac{v_m}{v_2}$
…	…	…	…	…
y_m	$\frac{v_1}{v_m}$	$\frac{v_2}{v_m}$	…	$\frac{v_m}{v_m}$

where: v_i – the sequential number in the criteria list of *Y*-vector, ranked by importance.

The above table interprets the matrix of criteria pairwise comparison. To obtain a metric for the final patient state using AHP, the geometric mean for each matrix row is calculated. Then, the obtained values should be normalized; they will be the weights (w) of each criterion of the *Y*-vector, so the metric can be represented as follows:

$$F_{ac} = w_1 y_1 \pm w_2 y_2 \pm \ldots \pm w_m y_m \quad (2)$$

This metric is a function of additive convolution (F_{ac}) of the criteria. It has the advantage of flexibility because it depends on priorities set out in Table 1 by the decision-maker. The signs in (2) are placed depending on whether it is necessary to maximize (then the "+" sign) or minimize (then the "-" sign) y_i.

That solves the first issue of the given multi-criteria optimization problem, which allows converting it to the single-criteria optimization problem, where it is necessary to find the maximum of F_{ac}. Solving it can be done by many optimization approaches. One of the most famous methods is the Genetic Algorithm [53-55] – a stochastic method for finding the required solution. The ideas of natural selection and genetics provide a fast search for the global optimum, thus solving the second issue of the NP-complete problem. The algorithm is shown schematically in Fig. 2.

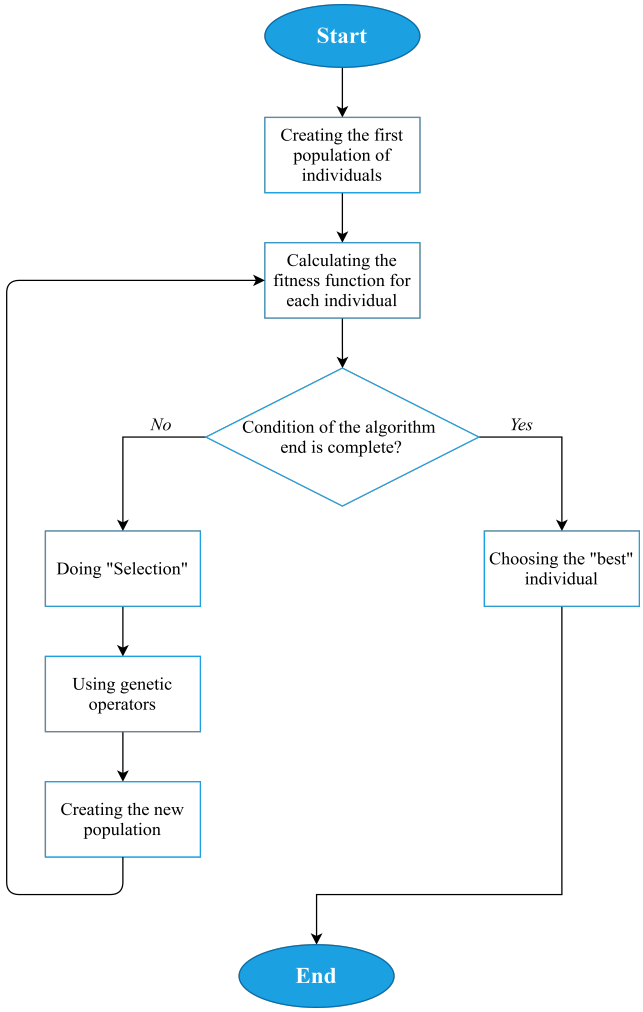


Figure 2: Genetic Algorithm scheme

To describe the Genetic Algorithm in more detail:

1. A random sample ("population") of *N* arrays ("individuals", or "chromosomes") that contain values of *I*-vector parameters ("genes") is created (Figure 3).

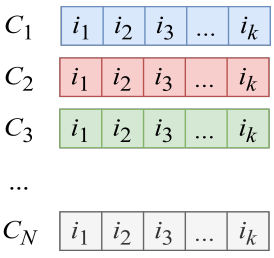


Figure 3: Population in general form

The number N, as well as the boundaries in which the values of genes will lie, are set directly by the researcher.

2. "Fitness function" for each individual is calculated. In the current research, the fitness function is F_{ac} (2).

3. The condition of the algorithm end is checked (it can be the presence of the preassigned value of F_{ac} or exceeding the time limit of the algorithm).

3.1. If the condition is complete, the Genetic Algorithm returns the "best" individual (optimal clinical treatment strategy).

3.2. If the condition is incomplete, the formation of a new population begins.

3.2.1. The "selection" [53-55] of individuals from the current population is carried out. This procedure aims to select individuals for the next generation creation, and the chance of selecting each individual directly depends on the value of his fitness function. The selected individuals form N pairs, which will then give "new" individuals.

3.2.2. The usage of "crossover" (one of the genetic operators) [53-55] for crossing the resulting pairs of individuals. A mixing of "genes" (parameters of I-vector) occurs between a pair of individuals, thus forming a new individual that stores the information about his "ancestors". In the general case, randomly the "crossover point" is determined, which allows mixing a pair of "parents": the genes of the first parent are before the crossover point, and after it – the genes of the second parent:

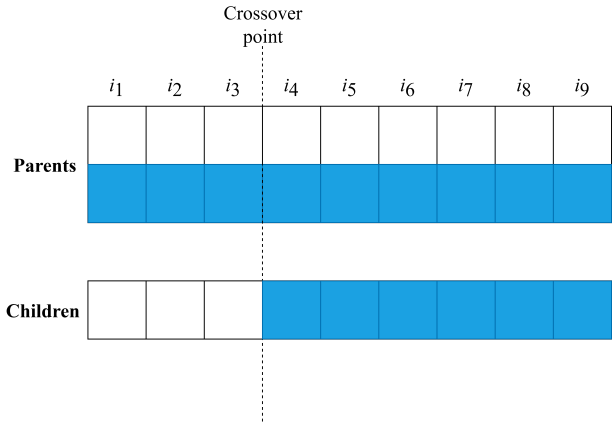


Figure 4: Example of crossover

That creates a new individual. To ensure the diversity of the population during the entire algorithm's operation, another genetic operator called "mutation" [53-55] is also used. This operator can be triggered with low probability instead of crossover, and its main purpose is to replace randomly selected genes of individuals with completely new ones.

3.2.3. Back to point number 2.

4. Using the best individual as a recommendation for personalized treatment. Multiple choices can be derived so that the doctor has a choice.

As a result, the Genetic Algorithm has been got, where F_{ac} is used as a fitness function, derived from the AHP ideology. The idea of using the convolution function (obtained by the MCDM method) as a fitness function of the Genetic Algorithm is not new. In [56], the authors used Weight Sum Approach and Tchebycheff Approach to get the convolution function. The authors of [57] were comparing Non-Dominated Sorting Genetic Algorithm II (NSGA-II), Multi-Objective Differential Evolution (MODE), and Multi-Objective Particle Swarm Optimization (MOPSO) algorithms. Such approaches are rather difficult to interpret, which makes it more complex to explain to the doctor the principle of the algorithm for finding the optimal clinical treatment. Therefore AHP was chosen to obtain a convolution function.

5. Statement of Findings

5.1. Description of Clinical Data

To test the performance of the algorithm, 2 clinical databases of patients with congenital heart defects [58] were used, which were provided by Amosov National Institute of Cardiovascular Surgery [59].

The first database ("DB1") has 128 patients from 3 to 28 years. They underwent a total cavopulmonary connection (TCPC) in an extracardiac conduit modification as the final stage of hemodynamic correction between January 2005 and September 2016. Patients were treated in two phases: surgical treatment (various types of surgery were performed, including TCPC) and conservative treatment (use of medication). Only conservative treatment is considered for the research. With that in mind, the database has the following variables:

- **7** patient indicators before conservative treatment – the vector X.
- **22** types of drugs (that were used to treat patients) – the vector I.
- **38** patient indicators after treatment – the vector Y.

It is worth mentioning that patient indicators before treatment were selected with the help of doctors specifically for the research.

The second database ("DB2") has 144 patients from 1 to 18 years. As in the first case, patients were treated in two stages (the only difference was the methodology). Variables of this database:

- **10** patient indicators before conservative treatment.
- **10** types of drugs (that were used to treat patients).
- **9** patient indicators after treatment.

It is also worth emphasizing that the variables' names are not given so that people cannot use this research for self-medication.

5.2. Modeling the Patient Final State

As mentioned earlier, to perform an optimal clinical treatment selection, it is necessary to obtain models of the patient final state parameters (1). Both patient indicators before and after conservative treatment are either quantitative or qualitative features. Regression methods can be used to model quantitative features. However, doctors are not so much interested in what a particular value will equal a feature whether it will be in the normal range. Therefore, the authors of this research proposed another unique approach, namely, the binarization of quantitative features.

In this way, it is necessary to get models of the binary features (0 – patient indicator after treatment is normal, 1 – patient indicator is abnormal). Patient indicators before treatment and drugs indicators will be used as models' predictors. That requires the use of classification methods. The following algorithms were chosen:

- Linear Discriminant Analysis (LDA) [60].
- Logistic Regression [29].
- Naïve Bayes [40].
- Linear SVM [41].
- SVM with Radial Basis Function (RBF) kernel [41].
- Gaussian Process Classifier (GPC) [61].
- Random Forest Classifier (RFC) [62].
- Adaptive Boosting (AdaBoost) [63].
- Multilayer Perceptron (MLP) [64].

All these algorithms were implemented using the *Python* programming language. Models were built for all final patient parameters in the "DB1" (38 indicators) and "DB2" (9 indicators) databases. To evaluate the models more adequately, the total data samples were split into training (80%) and test samples (20%). The models were evaluated according to their:

- accuracy (percentage of correctly classified objects):

$$\frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$

- sensitivity (share of correctly classified objects of the first class):

$$\frac{TP}{TP+FP} \quad (4)$$

- specificity (share of correctly classified objects of the second class):

$$\frac{TN}{TN+FN} \quad (5)$$

- Matthews Correlation Coefficient (MCC):

$$\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (6)$$

where: TP – true positives; FP – false positives (type I error); FN – false negatives (type II error); TN – true negatives.

The last metric (6) is a measure of binary classification quality. Its peculiarity is to consider positive and negative results, both true and false. MCC is a balanced measure that is used even for unbalanced classification. It is a correlation coefficient between real and predicted objects: it returns a value from -1 (complete mismatch) to 1 (perfect match). At 0, the classifier is considered to have made the prediction "by chance".

Tables 2 and 3 show results of the classification, namely, the average values of the model classification metrics for each algorithm. As seen from the tables below, the RBF SVM classification algorithm performed best, showing an average model accuracy of around 100% on the test sample.

The resulting models are mathematical equations for patient indicators after treatment (1), which can be substituted in formula (2) to get F_{ac}. This allows using the Genetic Algorithm to derive the best clinical treatment options for patients. These options will be personalized as each of the models is substituted for patient indicators before treatment.

Table 2: Final State Indicators Classification Results ("DB1")

Classification algorithm	Accuracy	Sensitivity	Specificity	MCC
Training sample (80%)				
LDA	89.6%	0.89	0.7	0.611
Logistic Regression	73.7%	0.742	0.779	0.321
Naïve Bayes	74.4%	0.807	0.758	0.429
Linear SVM	79.8%	0.801	0.838	0.474
RBF SVM	100%	1	1	1
GPC	98.6%	0.949	0.715	0.679
RFC	99.9%	1	0.996	0.998
AdaBoost	98.7%	0.985	0.98	0.97
MLP	87.1%	0.862	0.391	0.323
Test sample (20%)				
LDA	87.7	0.859	0.673	0.508
Logistic Regression	71.6	0.722	0.717	0.288
Naïve Bayes	75.3	0.811	0.74	0.412
Linear SVM	77.1	0.77	0.82	0.431
RBF SVM	99.2	0.987	0.99	0.982
GPC	97.6	0.936	0.728	0.661
RFC	99.1	0.995	0.972	0.975
AdaBoost	96.4	0.957	0.95	0.918
MLP	86.3	0.856	0.413	0.314

Table 3: Final State Indicators Classification Results ("DB2")

Classification algorithm	Accuracy	Sensitivity	Specificity	MCC
Training sample (80%)				
LDA	79.6	0.639	0.713	0.413
Logistic Regression	65.6	0.656	0.665	0.242
Naïve Bayes	54.5	0.321	0.901	0.215
Linear SVM	68.6	0.699	0.691	0.307
RBF SVM	100	1	1	1
GPC	100	1	1	1
RFC	99.9	0.999	0.996	0.997
AdaBoost	95.1	0.917	0.972	0.891
MLP	90.1	0.892	0.604	0.511
Test sample (20%)				
LDA	76.4	0.6	0.646	0.267
Logistic Regression	63	0.625	0.616	0.192
Naïve Bayes	55.4	0.364	0.843	0.172
Linear SVM	64.1	0.644	0.618	0.204
RBF SVM	99.6	1	0.989	0.992
GPC	98.5	1	0.972	0.967
RFC	99	0.991	0.971	0.973
AdaBoost	89.8	0.852	0.886	0.748
MLP	88.8	0.846	0.598	0.46

6. Conclusions and Future Work

The research described the development of an algorithm for personalized clinical treatment selection by using the principles of the Genetic Algorithm (for quick treatment variant searching) and the Analytic Hierarchy Process (for patient final state indicators simultaneous optimization). It was detailed from start to finish how the algorithm performs the selection of optimal treatment, including such steps as binarizing the quantitative features of the patient after treatment, and further modeling them with different classification algorithms (in the conference paper [1] only Group Method of Data Handling algorithm was used for modeling). A comparative analysis of classification algorithms showed that the best option for obtaining patient indicators after treatment models is the Support Vector Machine classifier with Radial Basis Function kernel.

The resulting classification models are substituted into the function of additive convolution formula (obtained by Analytic Hierarchy Process), which is used as an optimization function that estimates the final state of the patient. The values of this function range from 0 to 1, and the higher the function value, the better the patient's state. Such a feature could be used as a state-of-the-art metric for patient assessment.

This paper can be described as the beginning of creating a decision support system for personalized clinical treatment selection in Ukraine. It is necessary to carefully elaborate on all stages of the system to provide effective support for the doctor in deciding on a clinical treatment strategy. In this way, the system will be able to fill the absence of clinical pharmacists and optimize the work of medical institutions.

Also, despite the excellent results that were obtained, there were a few data to complete a full study. The problem is that Ukrainian medical institutions do not yet have much confidence in such Artificial Intelligence and Machine Learning methods, and with the medical liability legislation few institutions will provide data for comprehensive research and implementation. The decision support system requires a considerable number of resources and finances, which will be reviewed by authors for the future development of Medicine and Healthcare in Ukraine.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to express their gratitude to the experts from Amosov National Institute of Cardiovascular Surgery who provided clinical data and assist with the research.

References

[1] V. Babenko, O. Nosovets, "Calculating the Personalized Treatment Strategy by Genetic Algorithm Using Optimal Complexity Models," in 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT), 1-4, 2020, doi: 10.1109/CSIT49958.2020.9321947.

[2] "Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013," **385**(9963), 117-171, 2015, doi: 10.1016/s0140-6736(14)61682-2.

[3] S. Holm, "Final responsibility for treatment choice: the proper role of medical doctors?," Health Expectations, **14**(2), 201-209, 2011, doi: 10.1111/j.1369-7625.2011.00673.x.

[4] H. Mousa, "Prevention and Treatment of Influenza, Influenza-Like Illness, and Common Cold by Herbal, Complementary, and Natural Therapies," Journal Of Evidence-Based Complementary & Alternative Medicine, **22**(1), 166-174, 2016, doi: 10.1177/2156587216641831.

[5] M. Arruebo, N. Vilaboa, B. Sáez-Gutierrez, J. Lambea, A. Tres, M. Valladares, Á. González-Fernández, "Assessment of the Evolution of Cancer Treatment Therapies," Cancers, **3**(3), 3279-333, 2011, doi: 10.3390/cancers3033279.

[6] C. Pucci, C. Martinelli, G. Ciofani, "Innovative approaches for cancer treatment: current perspectives and new challenges," Ecancermedicalscience, **13**, 2019, doi: 10.3332/ecancer.2019.961.

[7] E. Crimini, M. Repetto, P. Aftimos, A. Botticelli, P. Marchetti, G. Curigliano, "Precision medicine in breast cancer: From clinical trials to clinical practice," Cancer Treatment Reviews, **98**, 102223, 2021, doi: 10.1016/j.ctrv.2021.102223.

[8] E. Chow, J. Doyle, T. Uyeki, "Influenza virus-related critical illness: prevention, diagnosis, treatment," Critical Care, **23**(1), 2019, doi: 10.1186/s13054-019-2491-9.

[9] D. Hall, A. Prochazka, A. Fink, "Informed consent for clinical treatment," Canadian Medical Association Journal, **184**(5), 533-540, 2012, doi: 10.1503/cmaj.112120.

[10] X. Lu, Z. Huang, H. Duan, "Supporting adaptive clinical treatment processes through recommendations," Computer Methods And Programs In Biomedicine, **107**(3), 413-424, 2012, doi: 10.1016/j.cmpb.2010.12.005.

[11] Z. Huang, W. Dong, L. Ji, H. Duan, "Outcome Prediction in Clinical Treatment Processes," Journal Of Medical Systems, **40**(1), 2015, doi: 10.1007/s10916-015-0380-6.

[12] S. Woolf, R. Grol, A. Hutchinson, M. Eccles, J. Grimshaw, "Clinical guidelines: Potential benefits, limitations, and harms of clinical guidelines," BMJ, **318**(7182), 527-530, 1999, doi: 10.1136/bmj.318.7182.527.

[13] L. Hughes, M. McMurdo, B. Guthrie, "Guidelines for people not for diseases: the challenges of applying UK clinical guidelines to people with multimorbidity," Age And Ageing, **42**(1), 62-69, 2012, doi: 10.1093/ageing/afs100.

[14] R. Rosenfeld, J. Shin, S. Schwartz, R. Coggins, L. Gagnon, J. Hackell et al., "Clinical Practice Guideline," Otolaryngology–Head And Neck Surgery, **154**(2), 201-214, 2016, doi: 10.1177/0194599815624407.

[15] L. Duff, E. Mclnnes, N. Cullum, A. Nelson, K. Luker, "Clinical guidelines," Primary Health Care, **9**(1), 28-30, 1999, doi: 10.7748/phc.9.1.28.s14.

[16] "Clinical Guidelines". Physiopedia. (2012). Retrieved 24 March 2021, from https://bit.ly/3hPPpFY.

[17] "National Clinical Effectiveness Committee Standards for Clinical Practical Guidance". Nursing and Midwifery Board of Ireland. (2015). Retrieved 24 March 2021, from https://bit.ly/3jY921d.

[18] "Clinical protocols" (in Ukrainian). Ministry of Healthcare of Ukraine. (2017). Retrieved 23 December 2020, from https://bit.ly/2T4LLj8.

[19] "Клінічні протоколи: що це і чи є вони в Україні" (in Ukrainian). Ministry of Healthcare of Ukraine. (2017). Retrieved 23 December 2020, from https://bit.ly/2T26BzF.

[20] J. Saseen, T. Ripley, D. Bondi, J. Burke, L. Cohen, S. McBane et al., "ACCP Clinical Pharmacist Competencies," Pharmacotherapy: The Journal Of Human Pharmacology And Drug Therapy, **37**(5), 630-636, 2017, doi: 10.1002/phar.1923.

[21] F. Khan, N. Waqas, A. Ihsan, P. Khongorzul, J. Wazir, W. Gang et al., "Analysis of the Qualities Matching New Classification of Clinical Pharmacist," Indian Journal Of Pharmaceutical Sciences, **81**(1), 2019, doi: 10.4172/pharmaceutical-sciences.1000473.

[22] "Clinical Pharmacy (Educational Programme)". National University of Pharmacy. (2015). Retrieved 15 October 2020, from https://bit.ly/3hZ0h46.

[23] "Ukraine to receive UAH 128 million for digitalization of services provided by Digital Transformation Ministry". UNDP. (2020). Retrieved 12 March 2021, from https://bit.ly/3e4czHm.

[24] "Відповідальність медичних працівників" (in Ukrainian). Legislation of Ukraine. (2011). Retrieved 14 March 2021, from https://bit.ly/3AOe9Hg.

[25] Ö. Emre, G. Sert, "European Charter of Patients' Rights," Turkish Journal Of Bioethics, **1**(4), 198-205, 2014, doi: 10.5505/tjob.2014.69775.

[26] "ASA Physical Status Classification System | American Society of Anesthesiologists (ASA)". American Society of Anesthesiologists. (2020). Retrieved 28 November 2020, from https://bit.ly/3xx4DGr.

[27] W. Owens, J. Felts, E. Spitznagel, "ASA Physical Status Classifications," Anesthesiology, **49**(4), 239-243, 1978, doi: 10.1097/00000542-197810000-00003.

[28] J. Havens, A. Columbus, A. Seshadri, C. Brown, G. Tominaga, N. Mowery, M. Crandall, "Risk stratification tools in emergency general surgery,"

Trauma Surgery & Acute Care Open, **3**(1), e000160, 2018, doi: 10.1136/tsaco-2017-000160.
[29] S. Sperandei, "Understanding logistic regression analysis," Biochemia Medica, 12-18, 2014, doi: 10.11613/bm.2014.003.
[30] L. Minne, A. Abu-Hanna, E. de Jonge, "Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review," Critical Care, **12**(6), R161, 2009, doi: 10.1186/cc7160.
[31] J. Muschelli, "ROC and AUC with a Binary Predictor: a Potentially Misleading Metric," Journal Of Classification, **37**(3), 696-708, 2019, doi: 10.1007/s00357-019-09345-1.
[32] C. Bulac, A. Bulac, "Decision Trees," Advanced Solutions In Power Systems: HVDC, FACTS, And Artificial Intelligence, 819-844, 2016, doi: 10.1002/9781119175391.ch18.
[33] P. Austin, "A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality," Statistics In Medicine, **26**(15), 2937-2957, 2007, doi: 10.1002/sim.2770.
[34] O. Takahashi, E. Cook, T. Nakamura, J. Saito, F. Ikawa, T. Fukui, "Risk stratification for in-hospital mortality in spontaneous intracerebral haemorrhage: A Classification and Regression Tree Analysis," QJM, **99**(11), 743-750, 2006, doi: 10.1093/qjmed/hcl107.
[35] A. Rovlias, S. Kotsou, "Classification and Regression Tree for Prediction of Outcome after Severe Head Injury Using Simple Clinical and Laboratory Variables," Journal Of Neurotrauma, **21**(7), 886-893, 2004, doi: 10.1089/0897715041526249.
[36] G. Dolce, M. Quintieri, S. Serra, V. Lagani, L. Pignolo, "Clinical signs and early prognosis in vegetative state: A decisional tree, data-mining study," Brain Injury, **22**(7-8), 617-623, 2008, doi: 10.1080/02699050802132503.
[37] A. Abu-Hanna, N. de Keizer, "Integrating classification trees with local logistic regression in Intensive Care prognosis," Artificial Intelligence In Medicine, **29**(1-2), 5-23, 2003, doi: 10.1016/s0933-3657(03)00047-2.
[38] L. Gortzis, F. Sakellaropoulos, I. Ilias, K. Stamoulis, I. Dimopoulou, "Predicting ICU survival: A meta-level approach," BMC Health Services Research, **8**(1), 2008, doi: 10.1186/1472-6963-8-157.
[39] J. Trujillano, M. Badia, L. Serviá, J. March, A. Rodriguez-Pozo, "Stratification of the severity of critically ill patients with classification trees," BMC Medical Research Methodology, **9**(1), 2009, doi: 10.1186/1471-2288-9-83.
[40] V. Chauhan, K. Dahiya, A. Sharma, "Problem formulations and solvers in linear SVM: a review," Artificial Intelligence Review, **52**(2), 803-855, 2018, doi: 10.1007/s10462-018-9614-6.
[41] T. Fragoso, W. Bertoli, F. Louzada, "Bayesian Model Averaging: A Systematic Review and Conceptual Classification," International Statistical Review, **86**(1), 1-28, 2017, doi: 10.1111/insr.12243.
[42] R. Houthooft, J. Ruyssinck, J. van der Herten, S. Stijven, I. Couckuyt, B. Gadeyne et al., "Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores," Artificial Intelligence In Medicine, **63**(3), 191-207, 2015, doi: 10.1016/j.artmed.2014.12.009.
[43] Moore, D., Keegan, T., Dunleavy, L., & Froggatt, K. (2019). Factors associated with length of stay in care homes: a systematic review of international literature. Systematic Reviews, **8**(1). doi: 10.1186/s13643-019-0973-0
[44] P. Gago, M. Santos, A. Silva, P. Cortez, J. Neves, L. Gomes, "INTCare: a Knowledge Discovery Based Intelligent Decision Support System for Intensive Care Medicine," Journal Of Decision Systems, **14**(3), 241-259, 2005, doi: 10.3166/jds.14.241-259.
[45] N. Kaieski, C. da Costa, R. da Rosa Righi, P. Lora, B. Eskofier, "Application of artificial intelligence methods in vital signs analysis of hospitalized patients: A systematic literature review," Applied Soft Computing, **96**, 106612, 2020, doi: 10.1016/j.asoc.2020.106612.
[46] D. da Silva, D. Schmidt, C. da Costa, R. da Rosa Righi, B. Eskofier, "DeepSigns: A predictive model based on Deep Learning for the early detection of patient health deterioration," Expert Systems With Applications, **165**, 113905, 2021, doi: 10.1016/j.eswa.2020.113905.
[47] I. Nastenko, V. Pavlov, O. Nosovets, K. Zelensky, O. Davidko, O. Pavlov, "Solving the Individual Control Strategy Tasks Using the Optimal Complexity Models Built on the Class of Similar Objects," Advances in Intelligent Systems and Computing IV, **1080**, 535-546, 2020, doi: 10.1007/978-3-030-33695-0_36.
[48] V. Vaishnav, J. Vajpai, "Assessment of impact of relaxation in lockdown and forecast of preparation for combating COVID-19 pandemic in India using Group Method of Data Handling," Chaos, Solitons & Fractals, **140**, 110191, 2020, doi: 10.1016/j.chaos.2020.110191.
[49] R. Vanderbei, "The Simplex Method," International Series In Operations Research & Management Science, 11-25, 2020, doi: 10.1007/978-3-030-39415-8_2.
[50] S. Panjwani, S. Naresh Kumar, L. Ahuja, "Multi-criteria decision making and its applications," International Journal of Innovative Technology Exploring Engeering **8**(9 Special Issue 4), 2019, doi: 10.35940/ijitee.I1122.0789S419.
[51] T. L. Saaty, "Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World," RWS Publications, 1990, doi: 10.1016/0377-2217(89)90066-0.
[52] T. L. Saaty, "THE ANALYTIC HIERARCHY PROCESS WITHOUT THE THEORY OF OSKAR PERRON," International Journal Of The Analytic Hierarchy Process, **5**(2), 2014, doi: 10.13033/ijahp.v5i2.191.
[53] D. Whitley, "A genetic algorithm tutorial," Statistics And Computing, **4**(2), 1994, doi: 10.1007/bf00175354.
[54] J. García, C. Acosta, M. Mesa, "Genetic algorithms for mathematical optimization," Journal of Physics: Conference Series, 5-5, 2020, doi: 10.1088/1742-6596/1448/1/012020.
[55] S. Katoch, S. Chauhan, V. Kumar, "A review on genetic algorithm: past, present, and future," Multimedia Tools And Applications, **80**(5), 8091-8126, 2020, doi: 10.1007/s11042-020-10139-6.
[56] Q. Zhang, H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," IEEE Transactions on Evolutionary Computation, **11**(6), 712–731, 2007, doi: 10.1109/TEVC.2007.892759.
[57] H. Monsef, M. Naghashzadegan, A. Jamali, R. Farmani, "Comparison of evolutionary multi objective optimization algorithms in optimum design of water distribution network," Ain Shams Engineering Journal, **10**(1),103–111, 2019, doi: 10.1016/j.asej.2018.04.003.
[58] A. Dydyk, O. Nosovets, V. Babenko, "Setting Up the Genetic Algorithm for the Individualized Treatment Strategy Searching," Herald of Advanced Information Technology, **3**(3), 125–135, 2020, doi: 10.15276/hait.03.2020.2.
[59] "Amosov National Institute of Cardiovascular Surgery". Retrieved 1 October 2018, from https://bit.ly/2T9vWrq.
[60] D. HU, X. LI, F. NIE, "Deep linear discriminant analysis hashing," SCIENTIA SINICA Informationis, **51**(2), 279-279, 2021, doi: 10.1360/ssi-2019-0175.
[61] D. Mackay, M. Gibbs, "Variational Gaussian process classifiers. IEEE Transactions On Neural Networks, **11**(6), 1458-1464, 2000, doi: 10.1109/72.883477.
[62] C. Mantas, J. Castellano, S. Moral-García, J. Abellán, "A comparison of random forest based algorithms: random credal random forest versus oblique random forest," Soft Computing, **23**(21), 10739-10754, 2018, doi: 10.1007/s00500-018-3628-5.
[63] D. Feng, Z. Liu, X. Wang, Y. Chen, J. Chang, D. Wei, Z. Jiang, "Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach," Construction And Building Materials, **230**, 117000, 2020, doi: 10.1016/j.conbuildmat.2019.117000.
[64] J. Tang, C. Deng, G. Huang, "Extreme Learning Machine for Multilayer Perceptron," IEEE Transactions On Neural Networks And Learning Systems, **27**(4), 809-821, 2016, doi: 10.1109/tnnls.2015.2424995.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 6, No. 4, 414-430 (2021)
www.astesj.com

ASTES Journal
ISSN: 2415-6698

Data Stream Summary in Big Data Context: Challenges and Opportunities

Jean Gane Sarr*, Aliou Boly, Ndiouma Bame

Faculté des Sciences et Techniques (FST) / Département Mathématiques et Informatique, Université Cheikh Anta Diop de Dakar, Dakar-Fann BP 5005, Senegal

ARTICLE INFO

Article history:
Received: 27 July, 2021
Accepted: 17 August, 2021
Online: 28 August, 2021

Keywords:
Data streams
Summaries
NoSQL
Big Data
Real time

ABSTRACT

With the advent of Big Data, we are witnessing a rapid and varied production of huge amounts of sequential data that can have multiple dimensions, we speak of data streams. The characteristics of these data streams make their processing and storage very difficult and at the same time reduce the possibilities of querying them a posteriori. Thus, it has become necessary to set up so-called summary structures, equivalent to views on the data streams which take into account these constraints and allow querying the data already pruned from the system. In order to take into account the aspect of volume, speed and variety of data streams, new methods have appeared in the field of Big Data and NoSQL. These solutions combined make it possible now to set up summaries that make it possible to store and process different types of data streams with more efficiency and representativeness and which best meet the constraints of memory and CPU resources necessary for processing data streams but also with some limits.

1 Introduction

In the development of the research in Databases, Statistics, Telecommunication, etc, and especially with the advent of Big Data, new applications which product intensive data have arisen in various areas [1]–[5].

These data are no more modelizied in the same manner like in classical databases (relational databases), but on form of transitory data streams. They have an ephemeral, continuous, veloce and mutidimentional caracter and their data are queried as soon as they arrived in the processing system by queries in continuous execution before to be discarded. If there is no well defined regrouping mechanism in place on part of these streams and that allow future re-execution of queries, data will be forever lost. Indeed, the data streams characteristics not allow to plan their storage in their globality, and make their processing relatively dependant on CPU and memory resources. However, these data could hold required informations for analytics traitement later. That way, it becomes essential to know how to retain these informations or a part of them for a future exploitation.

In this context, many works [6]–[9] have been done to make avalaible structures or algorithms that allow to have views over data streams for posteriori queries execution over them. These structures are known under the denomination of data stream summaries [10]. Thus in the literature, we find two approaches of data stream summary, generalist approach and specific approach. The generalist summary approaches of data streams allow to answer aproximatively in posteriori to any kind of queries over the data stream in an optimal manner and while respecting the storage and computation constraints required. Specific summary approaches are tailored to specific needs and oriented to more specific areas. In the front of these approaches, other works have been done combining the Big Data technologies with NoSQL databases [8], [9]. They have as goal to bring durable solutions to processing and storage constraints of data streams. Indeed, these technologies give largest capabilities of computation and storage and offer performances in terms of treatement that can be done over clusters of machines allowing thus to scale-up.

The need to be able to perform real-time analysis of the data available at all times is becoming more and more urgent in all areas. Thus, in order to make a contribution to the question, we have as objectives in this work (i) to make an in-depth state of the art of the techniques and Big Data tools put in place to summarize the data stream, (ii) to study the advantages, disadvantages and use cases of these tools, (iii) to describe what are the different phases of the data stream processing and how these tools are used in these phases (iv) and finally , to see what perspectives are opening up for a future contribution for example to see the possibility to do OLAP analysis

*Corresponding Author: Jean Gane SARR, Dakar Senegal,Phone, Email : jeangane.sarr@ucad.edu.sn

over Big Data streams.

Thus, in this paper, while taking into account the opportunities offered by Big Data technologies, we abord in the section 2 the problematic of storage and computation of data streams. In the section 3, we present a state of art of the literature over the data streams summarization. This section treats generalist and specific approaches for data streams summaries. In the section 4, we study the use of the Big Data technologies to produce and treat summaries of data streams. And to finish, in the section 6, we dress the bilan and the future works relatives to this paper.

2 Problematic of data streams storage and processing

With the advent of intensive applications which produce huge volumes of data like fraud detection, roads traffic monitoring, the management of smart electricity meters, etc.[1]–[3], processing of data streams is often confronted to storage and computation constraints caused by the fact that they are generated in swift and continuous manner with variables velocities. The unlimited nature of data streams, produced in a rapid and continuous manner make that the Data Streams Management Systems (DSMS) do not have sufficient resources for their storage and processing. Indeed, the characters (rapid, continuous, infinite, etc.) of data streams do that, their storage in their totality is not conceivable. In the operation of a Data Stream Management System (DSMS), data are continuously treated on the fly relatively of a temporal windows defined in priori. After expiration of this windows, the data expire and are discarded from the system (lost). However, it can be very useful to submit posteriori queries on these data. Indeed, if new needs are declared that include a particular task that requires data already discarded from the system, then this task will not be realizable. For example, when a decision maker wants to know sales trends of the last day or the power produced the last two days an equipment of the distribution system, within a temporal window of one hour, all of the data of these last periods would no longer exist in the system if it doesn't exists any mechanism, like low cost storage technologies (but we have to take into account the cost that could rise over time), to retain these data. In that fact, one will not be able to satisfy these kinds of requests. By the fact that the data of streams are discarded from the system at the end of the window, some tasks that require data of the past (no covered by the window) will never be satisfied. This problem still remains when it comes to meeting needs combining both unloaded data and current data (from a valid window). Indeed, an aggregation query or a join query that contains so-called blocking operators like MIN, MAX, AVG, Order By, Group By, etc. (for example the maximum of sales or the sum of amounts of sales by city of the last 48 hours by considering a windows of an hour) require that all of required data to be available. Thus, it becomes impossible to give satisfaction to theses kinds of queries. To efficiently answer to these kind of needs, the ideal would be to dispose of an unlimited memory to permanently store data. However, we are limited by a bounded memory length. Thus, it becomes necessary to put in place mechanisms that allow to conserve a summary of expired data in order to provide approximate answers (acceptable) instead of the exact answers (which are impossible to obtain).

In this study, the method followed to respond to theses needs is defined by a series of questions:

- What are the classic data stream summary techniques with their advantages and limitations ?
- What are the new available architectures in the context of Big Data to summarize data streams with a comparative study of them ?
- What tools or technologies can be used in the different layers of these architectures?
- What are the advantages, disadvantages and use cases of these tools?
- What are the perspectives of contribution after this study?

3 Data streams summary

A data stream summary can be defined as a structure or an algorithm that allows to permanently store a portion of data of a stream [11], [12]. Data streams summaries have the advantage to allow processing of analytics queries posed in posteriori over the data streams. Multiple researches works [10], [11], [13]have been proposed in the scope of the data streams summaries and that can be grouped in two majors approaches: the generalist data streams summary approach [11], [14], [15] and the specific data streams summary approach [16]–[18].

3.1 generalist summaries of data streams

A generalist summary of data streams [10], [13] is a data structure updated as the events of a stream data arrive in the system. This type of summary allows to answer, in optimal way, in posteriori and approximatively to all kind of queries that can interest the final user and that address the stream's data. It also allows to deal with the resources constraints (storage and computation power) subject to data streams. To be generalist, a summary of data streams must align with a certain number of points [13] :

- to respect the constraints of espace-memory and computation power (1).
- to exprime itself with variables belonging to cartesian product TXD where T is the timestamps'space and D is the set of qualitatives or quantitatives values (2).
- to allow an approximation of all queries of type SELECT AND COUNT over the TxD space (3).
- to allow the calculation of the approximation error in function of CPU and memory resources available (4).

This structure must allow in the same time to calculate bounds of approximative answers's precision of these queries. A generalist summary of data stream particularly looks for to control the lose of information between the data stream and the summary produced. Many works [11], [14], [15] have addressed the problematic of the generalist summaries of data streams.

- The authors of [11] proposed a general description of a summary that supports two operations *update(tuple)* and *computeAnswer()*. The *update* operation is called to update the summary's structure at each arrival of a new event and the *computeAnswer()* function update or produce new results for answering the posed query.

  During processing of the continuous queries, the optimal scenario is this where the two operations are running in a fast way compared with the arrival delay of events into the data stream. In that case, it is not necessary to provide special techniques to be in phase with the data stream (no blocking factor or bottleneck) and to produce approximative answers in real time : as new event arrives, it will be used to update the data structure and then new results are calculated by the way of this structure, all of this during a time period less important than the delay between arrivals of two successives events (i).

  The algorithm proposed does not respect the constraint (3) because the one or the two operations are slows; it then becomes impossible to continually produce an exact updated answer. The (1) constraint, in this side, is satisfied by the mean of technic describe in (i).

  By the same method, we can see that constraints (2) and (4) are in turn verified by this proposition.

- The authors of [14], in their side, proposed the StreamSamp algorithm based on the fundamental technic of random sampling of the entering data stream, followed by an intelligent storage of generated summaries that allow to analyze the data stream in ints entirety or in a part of it.

  This process allows to this algorithm to no depend on the data stream events arrival rate. However, the efficiency of StreamSamp deteriorates with time. Indeed, the weight of old events grows over time with a fixed sample size. Consequently, if the old events are putted together with others having a lower important weight, then they increase negatively the variance of the sample.

  With this algorithm, the (1) constraint is assured but with a logarithmic growth of the size of the summary according to the stream size. Even if the data stream size is potentially infinite, the logarithmic growth is a good solution in the practice. The (2), (3) and (4) constraints are verified by the application of the poll theory [19] that allows to pose bounds to the approximations's quality with condition to take in consideration the events weight, these masses depend to order of the sample in wich they are keeped.

- The authors of [15] proposed the CluStream algorithm that is divided in two phases, the On-line phase and the Offline phase. The On-line phase or micro-clustering phase is the portion of collection on line of statistics data. This processing do not depend of any user entry like the time window upper bound or the required granularity for the clusters construction. The goal, here, is to maintain some statistics in a level of granularity (spatial and temporal) sufficiently high to be used by the second phase or macro-clustering phase with a specific time bound as good as an evolutive analyze. This phase is inspired by the k-means and the nearest neighbour algorithms. Thus, the on-line phase iteratively operates by always maintening a set of micro-clusters updated by integrating the new events arrived in the system. These micro-clusters represent the snapshots of the clusters that change in each new occurring event of the stream.

  The off-line phase begins by creating a certain number of initials micro-clusters at the start of the data stream. Thus, at the start of the data stream, a number *InitNumber* of events is stored over hard drive and a k-mean algorithm is then used to create the initials micro-clusters. The number *InitNumber* is chosen in order to be the larger possible authorized by the k-mean algorithm processing complexity by creating initial micro-clusters.

  This proposition satisfies the (4) constraint because it allows to obtain approximative answers by using of Cluster Feature Vector (CFV). For the (2) constraint, it appears satisfied by the fact that in function of stream's characteristics (the space dimension of values) and of the number of micro-clusters, the structure update time at new event arrival can be bounded : this allows to reduce the needs in term of computation resources[13]. However, the CluStream algorithm does not match with the definition of a generalist summary of data stream over the (3) constraint because this algorithm doesn't allow to know the bounds for approximation of answers of SELECT AND COUNT queries type and also eventually over the (1) constraint (for the required memory-espace)[13].

3.2 *Specific summary approach of data streams*

Generalists data streams summaries studied in 3.1 constitute an ideal. However, applications used in different domains can require a more specific type of summary : i.e. a data stream summary specialized in a well precise domain. Thus, in the literature we find different summary techniques usable depending on needs. These mainly consiste of probabilistics techniques or data mining (sketchs, echantilloning, clustering, etc.) [16]–[24] where the summary has a probability to be selected (if these techniques are applied many times on the same dataset, the result (a summary) may vary) and the deterministics or statistics techniques [25]–[27](histograms, wavelets, etc.) which when applied on the same dataset will always give the same result (identical summaries).

3.2.1 *Probabilistics or data mining techniques*

From the point of view of the results they produce, data mining techniques can be likened to summaries. Indeed, a lot of research has been carried out to extract sequential patterns, or frequent items-sets by using of sliding windows, etc., allowing to capture trends in data streams.Thus, we can note techniques such as sketches, clustering, sampling that we propose to study in this section.

The Sketches

In their seminal article, the [16] authors introduced the first time the randomized sketching notion, which is since then, widely used. These are small data structures and provide very compact data

stream summaries by using few memory resources. This is a notable point in data stream field which is characterized by the memory space constraint. The authors of [17] proposed the following formalism of a sketch :
a sketch S(A) is a compressed form of a given sequence A, providing the operations :

- *INIT(S(A))* which defines how the sketch is initialized;
- *UPDATE(S(A),e)* which describes how to modify a sketch when a new event e arrives in the A sequence;
- *UNION(S(A),S(B))* given two sketchs for two sequences A and B, provides the sketch of their union $A \cup B$
- *SIZE(S(A))* which estimates the distinct number of events of the sequence A.

Their use makes it possible to respond to queries over all of the data stream by providing approximative answers. The main idea of this technique is to randomly project each event in a particular space using hash function and keep only the most relevant components, thus saving space while preserving most of the information. There are different implementations of these sketchs notably [20], Count sketches [21], the Bloom's filter [22], [23], Count-Min sketch (CM Sketch)[24], etc.

The sampling methods

The sampling techniques [18], [28]–[30] are also other probabilistics summary methods. The sampling over data streams is based on traditionals sampling techniques. However, it requires significants innovations, like sequential sampling, to prevent the unbounded size of data streams because generally requiring all the data in order to select a representative sample. The sampling techniques can also be coupled with windowing techniques in order to prevent the data expiration phenomenon of the stream by allowing to consider an interval in which to define the sample. These windows can be of two types : sliding or landmark. Thus, certains of these mains methods have been adaptated to data stream context. Within these last, we can cite the Ramdom sampling [28] or the Reservoir sampling[28] [18]. The Random sampling [28] uses a little sample to capture the essentials characteristics of a dataset. This can be in the simplest form of a summary to implement and others samples can be constructed from this one. This method seems to be inadapted when we process data streams with a certain complexity for example health data characterizied by a certain number of variables. Concerning the Reservoir sampling [18], [28] the basic idea is to select a sample size $\geq n$, from which a ramdom sample of size n can be generated. However, this method has some disadvantages such as the size of the reservoir which can be very large. This can turn this procedure costly. Furthermore, this method is useful for an insertion or an update but find its limits at data expiration in a sliding window. Thus, it must be implemented actualization algorithms of the sample without affecting the representativity. Indeed, in this type of window, the events no longer part of the current window become exceeded, and if they belong to the sample, they must be replaced. In that sens, many technics have been developed known as sliding windows sampling for processing the case of sliding windows which, for remember, can be logical (defined over a time period) or physical (based on the number of the events) [29], [30].

- For the logical windows, we find the periodic sampling [29] which consists to maintain a reservoir sample type for the K first events of the data stream constituting the first window. Then, when an event expires, it is replaced by the new arriving event.
  This procedure thus maintains an uniform random sample for the first sliding window of size L and do not demand much memory (K events each moment into the memory). However, this method has the drawback to be highly periodic. It is that the reservoir sampling searches to resolve by adding all new arrived events in the system to a backing sample by affecting to it a given probability $(2.\theta.K \log(L))/L$. Then it generates a random sample from this one. When an event expires, it is immediately deleted from the reservoir. However, this method does not determine the index of the event which must replace the event to erase. It is the same principle of operation that the chain sampling follows by with a $1/min(i, L)$ probability where L represents the window size. These different methods requiring a piori knowledge of the number of events cannot be applied to the case of the physical windows. Indeed, these consist of a number of events that vary over time that cannot be guessed a priori. In addition, several window events can vary at the same time (for example when sliding the window). To work around this problem, there are sampling methods based on physical windows.

- Among the sampling methods using physical windows [28]–[30], we find the sampling by priority which assigns to each event i a priority p_i between 0 and 1 and a replacer chain. What distinguish it from the on-chain sampling. The events in the sample are chosen by considering only those with a higher priority and with a more recent timestamp. We also find the Random pairing sampling [30] which allows to maintain an uniform sample over a sliding window combining Vitter's sampling algorithm [28] and whose of Babcock [29]. The algorithm can process any structure S_c that tolerates insertions and deletions but with a size that is always fixed.

As another sampling method, we have the join sampling [31], [32] which attempts to connect distributed data streams, for example, from meteorological measurement sensors from several stations. This method has the advantage of drawing each event in the sample in a single pass. However, its use requires having all the frequencies of the join key in the second relation. A last class of the probabilistic methods is the clustering techniques set [33]. The clustering problematic have been widely studied in Databases, Data mining and statistics. Indeed, the clustering is widely used by many applications in different domains. However, many of these methods cannot be applied to data stream management because they must be adapted to the data streams volumetry and to the computation power of the systems, in short, to all constraints generally required by the data streams. Furthermore, these methods must operate in one pass instead of many like in classical systems. Thus, viewed the important number of applications appearing in data streams, other researches have been done in the goal to provide new propositions

responding to the need of using clustering. This is how the authors of [33] present a distributed version of the clustering algorithm based on the density that they call DbScan [18]. We also find DenStream, StreamKM++ [34], etc. The particularity of these different algorithms is that they divide the clustering process in two major steps, to know, the on-line phase in which data are summarized into micro-clusters by conserving the data's temporal information (timestamp) and the off-line phase which uses the summaries also called quantifications of the first step to compute the final clusters. In the objective to evaluate these different algorithms, three measures are generally used [18] :

- The Accuracy that measures the clusters purity generated by the studied or provided algorithm with clusters having the labels that are of the dataset.
- The Normalized Mutual Information (NMI) that provides an independant measure of the number of clusters. It takes the maximal value of 1 only when two sets of labels have a perfect two-to-two match.
- The Rand Index that measures the accuracy which is used by a cluster to be able to classify the data elements by comparing the labels of underlying classes.

3.2.2 *Deterministic or statistic techniques of summaries*

A data stream can be defined by data of different nature i.e. qualitative or quantitative. In the latter case, we find numerical data streams which can be likened to time series whose size is unbounded. In other words, they are streams having a constantly evolving size and whose values are taken in the space of real numbers. As example, we have the phones calls, meteorological data from sensors, etc. Immediately, it is necessary to apply the methods of signal theory to them to summarize such data [35]. Among these efficient and robust mathematical tools, we find histograms [25], [26], wavelet compression, Fourier transformations, discrete cosine transforms, curve segmentation, etc.[27], [36], [37].

The Histograms

Histograms are commonly used in data summaries structures to succinctly capture the distribution of the values (discrete or continuous) in a dataset (a column or a tuple of a table). They have been used for a multitude of tasks such as estimating query sizes, queries responses approximation, as well as in data mining. Furthermore, they can be used in order to summarize data which come from streams. The literature offers different types of histograms like Histogrammes V-optimal [11], [12], Equi-Width Histograms, End-Biased Histograms)[25], [26] or Compressed histogram [27].

The comparaison by wavelets

The wavelet transform, like the Fourier transform, is a mathematical tool for capturing the evolution of digital functions (signal processing). They are often used as techniques to provide a rough representation of data, the [38], [39] data cube approximation, etc. In the context of data streams, the constraints known as the large volume and the often high rate with which events occur in the stream still apply to wavelets taken as a data processing algorithm. And for their use in this field, it becomes necessary to design techniques for processing wavelets in data streams. With this in mind, the authors of [40] show how to dynamically maintain the best wavelet coefficients efficiently as the underlying stream data is updated. While in [41], the authors propose a technique to get closer to the best dyadic (bipolar) interval which best reduces the error. This gave birth to a light algorithm to find the best wavelet representation denoted B-term Haar. It is for this reason that the Haar wavelet representation method relies on constant-valued dyadic intervals. In [12], the authors attempt to construct and update an optimal wavelet summary by considering time series modeling where new events are inserted at the end of the series (stream from a sensor of temperature measured at each instant). And since only the coefficients belonging to the path from the root to the new elements will be changed, most of the wavelet coefficients will not be affected by the insertion of new events. This implies a very simple construction of an algorithm with B coefficients by the use of a metric L_2. (Maintain the highest B coefficients in terms of absolute normalized values among the finalized values, as well as the updated coefficients ($log(N)$ coefficients, N initial size of the series). These are then compared with the B coefficients to construct a new subset of coefficients of high values and the turnstile modeling where the elements of the stream update the data of the series (distributed sensor streams accumulating their measurements before sending them to a central server) allow a more general wavelet decomposition insofar as all the coefficients can be affected by the arrival of a new event because this can be linked to any event already present in the series. Which makes difficult to maintain the coefficients and therefore to build the summary. Others authors like in [42], use a sketch in order to maintain, according to probabilities, an incremental summary of the data stream which will then be used as a basis for calculate the B coefficients. Those of [43], do not decompose the sketch in wavelets but directly construct their sketch by mean of the data of the stream decomposed in wavelets, afterward the obtained sketch is updated by incremental way. Their algorithm est thus applicable to many areas and is usable for to extend multidimensionnal data streams. In this section, we have seen that with the growth of applications producing massive, fast and varied data streams, computer systems are subject to very great storage and processing constraints [12]. These data can be structured, semi-structured or unstructured and their type can be either qualitative or quantitative. Thus, the need arose to have structures called summaries to store and query the pruned data of the system. These summaries can be generic in order to respond to any kind of query or specific to respond to specific queries only. Different methods such as probabilities, statistics, etc., allow these summaries to be made. Thus, the choice of a summary technique is often guided by the field one wishes to study. However, their excessive memory resource requirements and processing time due to their complexity mean that these techniques do not appear to be the best solution to adopt for the construction of data stream summaries in a large scale context. Furthermore, their use does not always make it possible to have sufficiently representative summaries which can help decision-makers to always be as well informed as possible. These constraints are however nowadays more and more well apprehended by Big data tools.

4 Big Data tools for data streams summaries

As discussed in the previous sections, many applications in different fields such as social networks, or IoT, produce huge amounts of varied data in a rapid way; we are talking about Big Data [1]–[3]. "Big Data" represents large volumes or streams of data that can be structured, semi-structured or unstructured. These data streams are generated quickly so that traditional databases systems do not have sufficient processing and storage resources to support them. This is why new tools adapted to this context have emerged. Generally, these techniques provide high performance, fault tolerance and can operate on distributed architecture systems. For each well-defined stage of the data stream processing cycle, several technologies are available.

4.1 Data streams processing architecture

Each big data solution acts in a well-defined phase of the big data stream processing cycle [44]–[46]. These phases range from the collection or ingestion of data streams to the analysis of these streams, including the data processing and storage management. These different phases define a multi-layer architecture with those of the highest level strongly dependent on the low level layers. In general, we have 4 main types of architectures (figure 1) for processing Big Data data flows having the following layers :

- Data retrieval layer which takes care off the collection and transfer of data streams to the processing layer;
- Data processing layer which is responsible for performing processing operations on the streams and preparing them to be summarized;
- Data storage layer which stores summaries from generated data streams;
- Data analysis layer in which the visualization of data is defined for analysis and decision-making.

Each of these different layers has its own characteristics and involves its own tools.

4.2 Data streams collect and ingestion

The ingestion [47] step is the entry point for the entire data processing system. Indeed, this step links all data streams sources such as the electricity consumption collected in real time from all meters of the network structures (substations, transformers, feeders, etc.) [2] to the storage layer. These streams are collected and then injected into the system by various tools which operate in producer/consumer approach. Given the large amount of data collected, this phase will have to eliminate some unnecessary data through filtering. All this, taking into account significant information such as outliers [48] which may reflect anomalies or matters useful for decision-making. It is also in this step that we have to ensure the generation of metadata on the structure and origin of the data, but also on the details of the collection. These will be of capital importance for the rest of the phases, more particularly, the data analysis.

In this step can intervene tools such as Kafka [49], Flume [50] or even Nifi [51], etc.

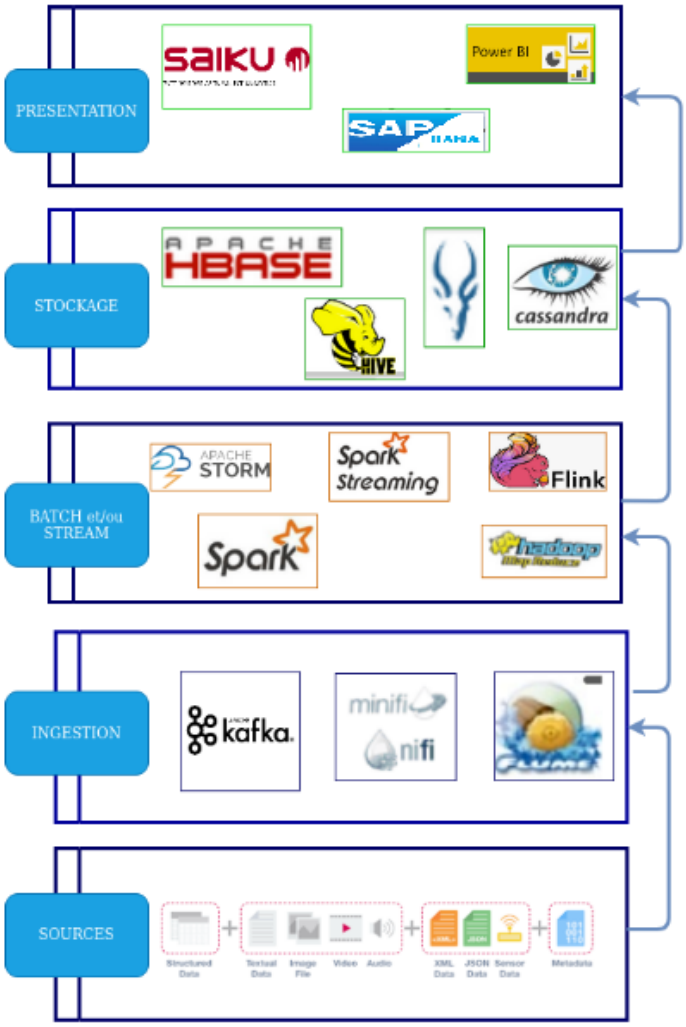


Figure 1: Data Stream processing architecture layers

- Kafka is a distributed messaging system that collects and delivers large volumes of data with low latency [49]. Kafka's architecture is essentially made up of three parts, namely a producer which collects data from different sources then injects them into a "topic". The topic describes the events of the data stream and has a queue structure. These queues are managed by brokers who are simple systems responsible for maintaining published data. Each broker can have zero or more partitions per topic. If a topic admits N partitions and N brokers, each broker will have a single partition. If the number of brokers is greater than the number of partitions, some will not have any partition from this topic. The third element in the Kafka architecture is the consumer which retrieves the elements of the topic then injects them into a processing layer upstream. Producer and consumer can be written in different languages like Java, Python or Scala etc. There are also producers already defined (owners) such as in Azure [52], CDC (Change Data Capture) techniques [53] which make it possible to detect changes (insert, update, etc.) on the rela-

Table 1: Big Data streams processing main architectures

Technology	Principle	Advantages	Incovenients	Usage
Traditionnal Big data architecture	• Positionned to solve the problems of traditional BI • It can be seen that it still retains the ETL action and enters the data storage through the ETL action.	• Simple and easy to implement as per BI system concerns • The basic methodology has not changed. • The only change is the selection of technology, replacing the BI components with the big data architecture.	• For big data, there is no such complete cube architecture under BI. • At the same time, the architecture is still mainly batch processing and lacks real-time support.	• Data analysis needs are still dominated by BI scenarios • But due to issues such as data volume and performance, they cannot meet daily use.
Data streaming architecture	• The batch processing is directly removed • And the data is processed in the form of streams throughout the entire process. • The ETL is replaced with a data channel. • The data processed by stream processing is directly pushed to consumers in the form of messages. • Although there is a storage part, the storage is more stored in the form of windows, so the storage does not occur in the data lake, but in the peripheral system.	• There is no bloated ETL process, • The effectiveness of the data is very high	• For streaming architecture, there is no batch processing, so data replay and historical statistics cannot be well supported. • For offline analysis, only analysis within the window is supported.	• One can use this as an early warning • The different monitoring aspects, and the data validity period requirements.
Lambda architecture	• Lambda's data channel is divided into two branches real-time streaming and offline. • Real-time streaming basically depends on much of the streaming architecture to ensure its real-time performance • While offline is mainly batch processing to ensure final consistency.	• Both real-time and offline, covering the data analysis scenarios very well.	• Although the offline layer and the real-time stream face different scenarios • Their internal processing logic is the same, so there are a lot of honors and duplicate modules.	• There are both real-time and offline requirements.
Kappa architecture	• The Kappa architecture is optimized on the basis of Lambda • Combining the real-time and streaming parts • And replacing the data channel with a message queue.	• The Kappa architecture solves the redundant part of the Lambda architecture. • It is designed with an extraordinary idea of replaying data. • The entire architecture is very simple.	• Although the Kappa architecture looks concise • It is relatively difficult to implement, especially for the data replay part.	• It provides features like Lambda architecture.

tional databases like Debezium [54] in the NoSQL, the Neo4j Stream connector which links Kafka to the graph-oriented database Neo4j [55]. In addition, in order to be able to query the data of the stream passing through the topic, Kafka now has the KSQL language [56] which is an SQL-like language for streaming. The kafka project is now supported by the Apache foundation.

- Flume is a collecting, aggregating and transferring framework of large volumes of data in HDFS (Hadoop Distributed File System) [50] file systems such as Hadoop, HBase or Spark [47]. In addition to the Hadoop ecosystem, Flume also allows the injection of data stream from social networks such as twitter, facebook, etc. Like kafka, Flume's architecture consists mainly of three parts, namely the source, the chain and the

sink. The source retrieves the data streams to put them in the chain or channel. The Flume sink component ensures that the data in the chain has been transmitted to the destination which can be HBase, Hadoop, etc. In this architecture, the source, the chain and the sink are together called Agents. The figure 3 describes this architecture.

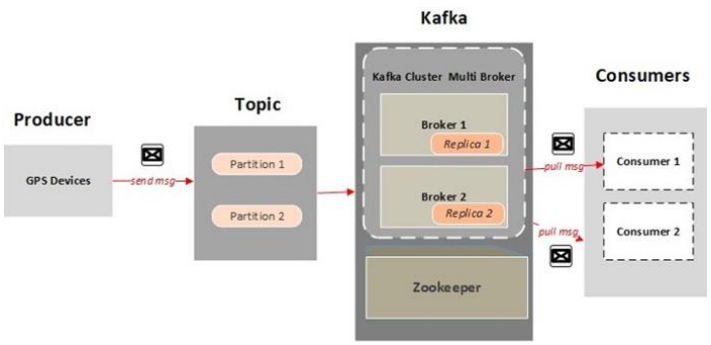


Figure 2: Kafka Architecture

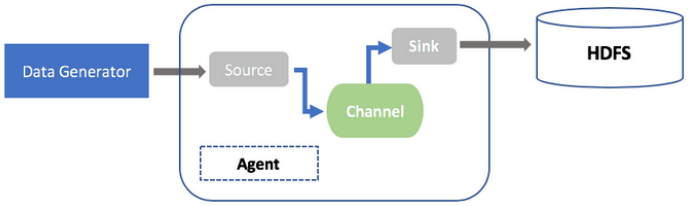


Figure 3: Flume Architecture

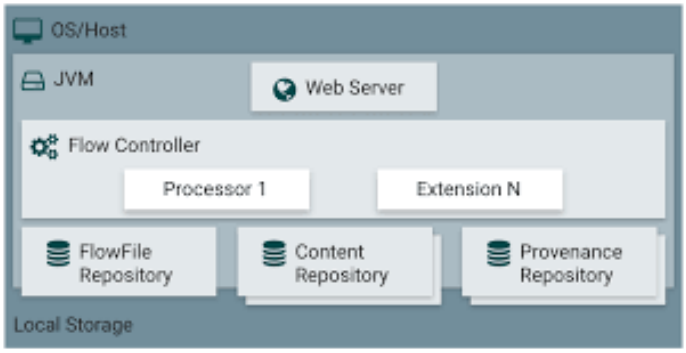


Figure 4: Nifi Architecture

- Nifi [51] is a data ingestion technology that uses data stream oriented processing. It enables data acquisition, basic event processing and a data distribution mechanism. NiFi provides organizations with a distributed platform for building [57] enterprise workflows. It provides the ability to accommodate various data streams generated by IoT. NiFi enables seamless connections to databases, big data clusters, event (message) queues and devices. It incorporates tools for visual command, control, provenance (lineage or data traceability), prioritization, buffering (back pressure), latency, throughput, security, scalability and extensibility [58]. NiFi is highly configurable and provides a scalable and robust solution to process and integrate data streams of various formats from different sources on a cluster of machines. It allows to manipulate data of network failure, bad data, security, etc. MiNiFi2 [59], [47], a sub-project of Apache NiFi is a complementary approach to NiFi fundamentals in data stream management, focusing on data collection at the source of their creation.

These different tools can be used separately but are not universal. In that way and according to the needs or scenarios, we can combine them. Thus Flume or Nifi can be used as a producer or consumer of Kafka. The combination of Flume and Kafka allows Kafka to avoid custom coding and take advantage of Flume's sources and strings, while Flume events passing through Kafka's topic are stored and replicated between Kafka's brokers for more resilience [60]. The combination of tools might seem unnecessary, as it seems to introduce some overlap in functionality. For example, NiFi and Kafka provide brokers to connect producers and consumers. However, they do it differently: in NiFi, most of the data stream logic does not reside inside the producer / consumer, but in the broker; which allows centralized control [60]. NiFi was designed primarily for data stream management. With the two tools combined, NiFi can take advantage of Kafka's reliable storage of stream data, while overcoming some Kafka [60] limitations such as lack of monitoring tools, reduced performance when a message has need to be touched up.
Table 2 compares these different tools.

4.3 *Data stream processing*

This step is responsible for standardizing the formats unsuitable for analyzing the data collected and extracting relevant information. It is also responsible for eliminating potentially erroneous data. Indeed, the [61] veracity criterion of Big Data requires verifying whether the data received is reliable and must therefore be verified before analysis. This processing can be done in two ways, namely stream processing or batch processing [44]–[46].

4.3.1 *The batch processing*

In the batch processing [62], [63], the data is collected and grouped into blocks of a certain duration (minute, hour, etc.) then injected into a processing system. For example, processing all measurements taken by the sensors after 10 minutes [64]. So, rather than processing data streams as streams, current configurations ignore the continuous and timely nature of data production. Data collection tools, workflow managers and planners orchestrate batch creation and processing [47]. This constitutes a continuous line of data processing [46]. Batch processing is best suited when the data streams are received offline (the data source only delivers its information every 30 minutes for example) and when it is more important to process large volumes of data to obtain more detailed information than to get quick scan results. For batch processing, there are different distributed platforms that provide scalable processing on clusters. Among these tools we find :

Table 2: Kaka, Flume, Nifi comparison

Ingestion tool	Out-of-the-box	Limites	Uses cases
Flume	• Configuration-based • Sources, channels & sinks • Interceptors	• Data loss scenarios when not using Kafka channel • Data size (KB) • No data replication	• Collecting • Aggregating and moving high volume streaming events into hadoop
Kafka	• Back-pressure • Reliable stream data storage • Kafka stream • Sources/skins with Kafa connect	• Custom coding often need • Data size (KB) • Fixed protocol/format/schema	• Streaming • Messaging • Systems integration • Commit log
Nifi	• Configuration-based UI • Many drag & drop processors • Back-pressure • Prioritized queuing • Data provenance • Flow template	• Not for CEP or windowed computations • No data replication	• Dataflow management with visual control • Data routing between disparate systems • Arbitrary data size

- Apache Hadoop [65] which is one of the most popular batch processing frameworks. Hadoop uses a master / slave architecture. The master is named namenode and is responsible for keeping the metadata on the cluster and distributes the data on the slaves. These are called datanodes and are used for storage in HDFS (its distributed file system). The namenode can also have a copy called secondary namenode which allows the cluster to be maintained in the event of the first on falling and thus becomes the main node [66]. By default, Hadoop stores the data to be processed in its distributed file system (HDFS) in CSV, Parquet or ORC format. However, it provides tools to store data on external databases such as NoSQL HBase or Cassandra databases. Hadoop uses the Map-Reduce [67] programming model for parallel data processing.

- We can also use Apache Spark [62], [68] which is a data processing framework that implements an execution engine based on direct acyclic graphs (DAG) and divides the processing into micro-batches. It also provides an SQL optimizer. In order to maximize the performance of Big Data applications, Spark supports in-memory processing but also on-disk processing when the data does not fit in memory. A Spark application takes its data from a collection of sources (HDFS, NoSQL, relational databases, etc.) or by interacting with data ingestion tools like Kafka, then applies a set of processes to them to generate interpretable results. For this, Spark uses an architecture made up of: [69] i) Spark driver which is the main node of a Spark cluster and which converts the processing into a set of tasks and then transmits them to ii) Cluster Manager which will carry them out on a set of iii) Executors which will be responsible for executing them.

There are several customized distributed computing architectures for batch processing [62]–[64]. However, they are not suitable for stream processing because in the Map-Reduce paradigm, all input data must be stored on a distributed file system (like HDFS) before starting processing. However, with data streams, there always arises the main storage constraint which means that this is not possible with the unlimited size of the data streams; hence the need to have streaming processing tools.

4.3.2 The stream processing

To process the data streams as they arrive in the system, it is necessary to carry out ongoing processing on them in order to be able to draw knowledge from them since their storage is not allowed by available resources [62]. This stream processing is generally carried out on clusters of distributed machines to allow to scale-up. This provides a certain availability of processing resources (memory and CPU). It is in this sense that recent Big Data platforms have been born such as Apache Storm [70], Spark Streaming [71], Apache Flink [64], Samza [72] which process continuous streams messages on distributed resources with low latency and high throughput [44]–[46].

- Apache Storm [70] is a stream processing framework developed by the company Backtype acquired by Twitter. The main goal of Storm is the processing of data streams in a distributed way with fault tolerance. To achieve this, it provides a framework for hosting applications and two approaches for creating these applications. The first approach called ”classic” composes the application according to a directed acyclic graph (DAC) called topology. The topmost part of this graph takes input from sources like Kafka. These data sources are called spouts. The latter therefore pass the data to processing units called bolts which will be responsible for executing the requests on the data streams. The second approach is a model for building applications called Trident [73]. This is a high level of abstraction at the top of the topology. This model is more for familiar operations like aggregations and persistence. It provides primitives intended for this type of processing. A Trident therefore calculates a topology by combining and splitting operations into appropriate collections of bolts.

- Spark [68] makes possible to process data which is frozen at an instant T. Thanks with the Spark Streaming module,

Table 3: Processing tools comparaison (1)

	Storm	Trident	Samza	Spark streaming	Flink (streaming)
Strictest guarentee	at-least-once	exactly-once	at-least-once	exactly-once	exactly-once
Achievable latency	<< 100 ms	< 100 ms	< 100 ms	< 1second	<100 ms
State Management	Small state	Small state	YES	YES	YES
Processing model	one-at-a-time	mico-batch	one-at-a-time	mico-batch	one-at-a-time
Back pressure	YES	YES	No buffering	YES	YES
Ordering	NO	between batches	within partitions	between batches	within partition
Elasticity	YES	YES	NO	YES	YES

it is possible to process data streams which arrive continuously, and therefore to process this data as and when as they arrive. Spark Streaming [71] is therefore an extension of Spark for streaming processing. It provides fault tolerance in real-time processing of data streams. The framework will accumulate data for a certain period of time and then produce a small RDD (Resilient Distributed Dataset). This RDD accumulation / production cycle will recur until the program is terminated. We are talking here about micro-batches as opposed to processing events one by one. Spark Streaming divides the incoming stream into these micro-batches of specific intervals and then returns a Dstream. The latter represents a continuous stream of data ingested by a source like Kafka, Flume, Twitter, etc. Dstreams are processed and then sent to a file system, database, real-time dashboard, etc. With micro-batch processing, Spark Streaming will add a delay between the arrival of a message and its processing. This therefore opposes it here to Apache Storm which offers real-time processing of events and non-compliance with the constraint (1) presented in section 3. This difference in processing, however, allows Spark Streaming to offer a guarantee of exactly once message processing under normal conditions (each message is delivered once and only once to the program, without loss of messages), and at least once in degraded conditions. (a message can be delivered several times, but always without losses). Storm in turn allows to adjust the guarantee level but, to optimize performance, the at most once mode (each message is delivered at most once but losses are possible) must be used. Another advantage of Spark Streaming is its API which is identical to the classic Spark API. It is thus possible to manipulate data streams in the same way as we manipulate frozen data.

- Apache Flink [64] follows a paradigm that embraces data streams processing as a unified model of real-time analysis, streaming data stream, and batch processing in a single programming model and with a only execution engine. In comparison with ingestion techniques like Kafka which allow a quasi-arbitrary reproduction of data streams, the data stream processing programs do not distinguish between the fact of processing the latest events in real time (Storm), the continuous aggregation of data periodically in windows (Spark Stream). Rather, they just start processing at different points in the continuous data stream and maintain states during [64] processing. While Flink on the other hand, through a highly flexible windowing mechanism, can process both early and approximate results, as well as delayed and precise results, in the same operation, thus avoiding the need to combine different systems for them for the two use cases [64]. The architecture of a Flink cluster includes three types of processes: the client, the Job Manager and at least one Task Manager. The client takes the program code, transforms it into a data stream graph and submits it to the Job Manager. This transformation phase also examines the data types (schema) of the data exchanged between operators and creates serializers and other type/schema specific codes. The Job Manager coordinates the distributed execution of the data stream. It tracks the status and progress of each operator and each stream, schedules new operators, and coordinates checkpoints and recovery. In a high availability configuration, the Job Manager maintains a minimal set of metadata at each checkpoint to fault-tolerant storage, so that a standby Job Manager can rebuild the checkpoint and recover the data stream execution from there. The actual data processing takes place in the Task Managers. A Task Manager executes one or more operators that produce streams and report their status to the JobManager. Task Managers manage buffer memory pools to buffer or materialize streams, and network connections to exchange data streams between cluster operators. In comparison to Spark, Flink incorporates i) an execution dataflow that leverages pipelined execution of distributed batch and streamed data flows, ii) native iterative processing, iii) sophisticated windowing semantics.

- Samza [72] is another LinkedIn project in the real-time data stream processing space. Become open source and added to the incubator of the Apache family of projects, Samza is a framework for real-time processing of data streams built on top of the Apache YARN infrastructure as well used by Spark or in Hadoop. Like Storm with Trident, Samza provides some primitives for building common types of data streams processing applications and for maintaining states within those same applications. The Samza application is based on the Application Manager which is used in order to manage the Samza Task Runners which are hosted in containers called YARN Containers. These Task Runners perform Stream Tasks which are the equivalent of Storm bolts for Samza. In other words, they take care of doing the desired processing on the events of the data stream such as the computation of aggregate functions (sum, min, max, avg, count, etc.). All communications

Table 4: Processing tools comparaison (2)

Technology	Principle	Advantages	Incovenients	Languages
Apache Spark	• The Apache Spark Architecture is founded on Resilient Distributed Datasets (RDDs). • These are distributed immutable tables of data, which are split up and allocated to workers. • The worker executors implement the data.	• Apache Spark is a mature product with a large community • Proven in production for many use cases • And readily supports SQL querying.	• Spark can be complex to set up and implement • It is not a true streaming engine (it performs very fast batch processing) • Limited language support • Latency of a few seconds, which eliminates some real-time analytics use cases	• Python • Java • Scala • R • SQL.
Apache Storm	• The Apache Storm Architecture is founded on spouts and bolts. • Spouts are origins of information and transfer information to one or more bolts.	• Probably the best technical solution for true real-time processing • Use of micro-batches provides flexibility in adapting the tool for different use cases • Very wide language support	• Does not guarantee ordering of messages, may compromise reliability • Highly complex to implement	• Java
Apache Samza	• Apache Samza uses the Apache Kafka messaging system, architecture, and guarantees, to offer buffering, fault tolerance, and state storage.	• Offers replicated storage that provides reliable persistency with low latency. • Easy and inexpensive multi-subscriber model • Can eliminate backpressure, allowing data to be persisted and processed later	• Only supports JVM languages • Does not support very low latency • Does not support exactly-once semantics	• JVM languages
Apache Flink	• Apache Flink is a stream processing framework that also handles batch tasks. • Flink approaches batches as data streams with finite boundaries.	• Stream-first approach offers low latency, high throughput • Real entry-by-entry processing • Does not require manual optimization and adjustment to data it processes • Dynamically analyzes and optimizes tasks.	• Some scaling limitations • A relatively new project with less production deployments than other frameworks	• Java • Maven

from Samza are submitted to Kafka brokers. Similar to Data nodes in a Hadoop Map-Reduce application, these brokers are often co-located on the same machines hosting Samza's containers. Samza therefore uses the topics or subjects of Kafka and a natural partitioning to implement most of the grouping modes found in data streams processing applications.

Both modes of processing have their advantages and disadvantages. The major advantage of streaming processing is its lack of complexity since it processes the data stream as it receives it. Also in comparison in terms of processing time, the stream processing has a lower latency time since the tuples are processed immediately after their arrival. However, they often have a low output rate. In addition, fault tolerance and load balancing are more expensive in stream processing than in batch processing [74], [75].

In batch processing, splitting data streams into micro-batches reduces [76] costs. Certain operations like state management are the most difficult to implement because the system will then have to consider the whole [77] batch. We can also note that the batch processing can also be incorporated into a stream processing as in Flink [64] or Apache Streaming [71], [78]. The choice of the type of processing and the tools will be guided by the nature of the application to be implemented. Thus, in the literature there are still other technologies that we have not studied in this paper such as [63], [79]–[84] for Stream processing. The figure 3 provides some elements of comparison of these different tools. In general, they meet the constraints given in section 3.

The table 4 is a comparison of these differents tools.

4.4 Summary or partial storage of data streams

In this layer, the processed data streams upstream are then integrated. After this processing, they are likely to be aggregated according to different temporal granularities. Then, they can be modeled according to a more suitable format to build a data stream summary which will be used for the partial storage of the stream and for its analysis later. This layer involves solutions such as HBase, Hive or even Cassandra for NoSQL databases management [44]–[46]. NoSQL databases are most often used to store Big Data [85], [86]. They are schema-free and allow the storage of many data formats without prior structural declarations. They are grouped into four categories according to the difference in implemented data models.

Table 5: Storage tools comparaison

Technology	Principle	Advantages	Inconvenients	Usage
HBase	• Distributed and scalable big data store newline • Strong consistency newline • Built on top of Hadoop HDFS newline • CP on CAP	• Optimized for read • Well suited for range based scan • Strict consistency • Fast read and write with scalability	• Classic transactional applications or even relational analytics • Applications need full table scan • Data to be aggregated, rolled up, analyzed cross rows	• Facebook message.
Cassandra	• High availability • Incremental scalability • Eventually consistent • Trade-offs between consistency and latency • Minimal administration • No SPF (Single point of failure) – all nodes are the same in Cassandra • AP on CAP	• Simple setup, maintenance code • Fast random read/write • Flexible parsing/wide column requirement • No multiple secondary index needed	• Secondary index • Relational data • Transactional operations (Rollback, Commit) • Primary & Financial record • Stringent and authorization needed on data • Dynamic queries/searching on column data • Low latency	• Twitter • Travel portal
Hive	• Hive can help the SQL savvy query data in various data stores that integrate with Hadoop. • Hive's partitioning feature limits the amount of data. Partitioning allows running a filter query over data stored in separate folders and only reads the data which matches the query.	• It uses SQL. • Fantastic Apache Spark and Tez Integration. • You can play with User Defined Functions (UDF). • It has great ACID tables with Hive 3+. • You can query huge Hadoop datasets. • Plenty of integrations (e.g., BI tools, Pig, Spark, HBase, etc.). • Other Hive-based features like Hive Mall can provide some additional unique functions.	• Very basic ACID functions • High latency • Hive isn't the best at small data queries (especially in large volume)	• Hive should be used for analytical querying of data collected over a period—for instance, to calculate trends or website logs.

Thus, we distinguish between key-value oriented , column-oriented, document-oriented and graph-oriented models. In the rest of this section, we will introduce some NoSQL database management systems.

- Hbase [85] is a distributed, column-oriented database management system built on the Hadoop [65] HDFS file system on which it constitutes a major component of the ecosystem by providing real-time read / write access to HDFS files. The architecture of HBase is made up of worker nodes in HBase also called Region Servers. Each Region Server contains an arbitrary number of regions. Each region is responsible for storing rows from a specific table, based on an interval of row keys. The actual contents of the lines are stored in HF files on the underlying HDFS file system. An HBase master node coordinates the Region Servers and assigns their row key intervals. HBase provides fault tolerance for storing large volumes of sparse data. It includes an environment allowing compression, in-memory processing and filters on database columns. HBase is used more and more in different systems like facebook messaging system. It also provides an API in Java [85]. When processing [66] data streams, HBase is used to store the results or summaries obtained either in batch with Spark or in Stream with Storm for analysis purposes. HBase performs very well for real-time analysis of big data streams, and is thus used by Facebook for message processing and real-time analysis [46], [87].

- Hive [88] is a data warehouse structure for analyzing structured data stored in Hadoop. These data from the HDFS system can be in CSV, Avro, Parquet or even ORC format. Hive is most often used for the purpose of creating big data stream summaries and making them easy to query by providing an SQL like language called HQL or HiveQL. It also provides ODBC drivers for accessing data by tools such as Power BI Desktop etc. Hive supports [89] online scanning. The architecture of Hive is composed of an interface with the user (either in command line or in graphical mode), of a metastore which is a database of metadata on the tables and databases created in Hive, a Thrift Server allowing the operation of Hive

and executing the queries, a Driver which manages the life cycles of HiveQL queries during compilation, optimization and execution, a Compiler which is invoked by the driver by passing it the HQL request. The latter transcribes this into a plan or a DAG of map-reduce jobs [67]. This plan is then transmitted to the execution engine in topological order. Hive uses Hadoop as the [90] runtime engine.

- Apache Cassandra [86] is a NoSQL database management system that provides a distributed and decentralized storage system for managing large volumes of data. It is column oriented and provides scalability, fault tolerance and consistency [91]. In Cassandra's architecture, there is no master node to manage all the nodes of the network like the HBase namenode [87]. The distribution of the data on the different nodes is done in an equivalent manner. Cassandra defines this environment to guarantee more consistency of data as well as for the availability of resources [92]. For its good writing performance, Cassandra is increasingly used for processing [87] data streams by different organizations like Facebook and for IoT [86], etc.

These tools that we have just mentioned are not the only ones. There are others like MongoDB [93] a document oriented database, Neo4j [55] graph oriented from NoSQL technologies . The choice of one or the other of these systems must be made according to the needs of the application.

Also, as discussed in section 3, data streams summaries are either generalist or specific in nature. This separation is also followed by Big Data tools which can often be used in both summary type cases. Thus there are some implementations such as the MongoDB [93] database used in [94] in order to analyze the data streams collected from sensors positioned on sick subjects allowing the doctor to make the right diagnoses in order to administer the best treatments and be more reactive in the event of attacks (respiratory, cardiac or stroke, etc.) [95]–[97] leader in data-oriented graph which has developed for its server a connector for Kafka and Confluent called Neo4j Streams [98] allowing to integrate streaming events for the analysis of financial frauds, knowledge graphs and a vision to all levels of clients of the [99] system. There are also key-value oriented NoSQL databases like Redis [100] coupled with Storm in [101] and [102] to keep static data on clients who have visited a website in order to be able to enrich the analyzes carried out on the sections on which they had to click, we speak of clickstream as well as in [103] for the analysis of Twitter data. Table 5 provides a comparison of these different tools.

4.5 Data stream visualization and analysis

In order to facilitate decision making from data streams, they must be described according to a certain number of representation models such as tables, graphs, curves, etc. And these different visuals when put together, make it possible to build dashboards. In this sense, this phase of data analysis is one in which we can detect intrinsic patterns, extract relationships and knowledge, but also correct errors and eliminate ambiguities. It then makes it possible to be able to interpret the data streams more easily. This is because decision-makers must interpret the results of an analysis of data stream. This is necessary in order to rule out errors. SAP Hana [104], Power BI [105] Saiku [106] are different tools for carrying out this analysis. Table 6 provides a comparison of these different tools.

- Saiku [106] offers a user-friendly web analytics solution that allows users to quickly and easily analyze their data and create or share reports. Saiku connects a wide range of OLAP servers including Mondrian, Microsoft Analysis Services, SAP BW and Oracle Hyperion and can be deployed quickly and inexpensively to allow users to explore data in real time.
- SAP Hana [104] is a column-oriented and in-memory relational database management system and has a database server to store and retrieve data requested by applications in real time. In addition, SAP Hana performs advanced real-time analysis of big data streams (prescriptive and predictive analysis, sentiment analysis, spatial data processing, continuous analysis, text analysis, social network analysis, text search, processing. graphics data) and contains ETL capabilities as well as a real-time application server [107].
- Power BI [105] is a self-service Business Intelligence solution produced by Microsoft. It provides business user-oriented data visualization and analysis capabilities to upgrade the decision-making process and business visions. Power BI is a cloud-based, self-service BI solution. This means that we can build and deploy a solution immediately with data from cloud and on-premises data sources, systems and functions. Power BI is composed of two main parts namely a server installed locally or at Microsoft to host the reports and a tool to create and publish these [105] reports. It provides connectors to access different data sources like Hive, etc. For example, when a visualization (table, graph, etc.) in a dashboard is connected to a real-time data source (Direct Query), the visualization updates continuously, allowing faster information [105]. When dealing with data arriving in real time, this ability to automatically update reports offered by Power BI has a real advantage. Thus, it allows decision-makers to be informed at all times about the overall state of the subject studied.

As for the other phases of the architecture, there are a very large number of tools in the literature for visualizing data streams such as Splunk [44]–[46], [108].

5 Future works

It is possible to note that even if the works presented in this study provide rather interesting results, they do not always make it possible to satisfy the requirements subject to data streams. Indeed, if we consider data streams with several dimensions the requirements of representativeness which means that a model must always remain faithful to the source data is not always guaranteed. Also the requirement of compactness which makes it possible to guarantee that the model will be able to hold in memory can be not assured. The genericity prerequisite used in order to optimize the processing and storage times and to respond to different types of requests can also be not. The dynamicity criterion making it possible to take into account new events that have arisen in the stream or even rapid

Table 6: Visualization tools comparaison

Technology	Principle	Advantages	Incovenients	Usage
Saiku	• A lightweight open-source analytical tool which is written in HTML/JavaScript (jQuery) • Focuses on fast and configurable OLAP analysis.	• Undertake complex, powerful analysis using an easy to use, drag and drop interface, via the browser • Diverses data source. Deployement quickly with the graphical • User friendly Schema Designer for designing data models. • Creation consistent and re-usable meta data • RESTful web-services provider with JSON data payload.	• Pricing : to take advantage of all the features of Saiku Enterprise every user requires a licence	• Add reporting and analysis to any application or website • Explore data in MongoDB, Spark and more, directly from the browser. • Caching to address performance and speed up analysis.
Power BI	• Power BI is a cloud-based business analysis and intelligence service by Microsoft. • It is a collection of business intelligence and data visualization tools such as software services, apps and data connectors.	• Cost-effective • Custom visualization • High data connectivity • Regular updates • Integration with Excel • Attractive visualizations	• Crowed user interface • Difficult to understand • Rigid formulas (DAX)	• Real-time analysis • custom visualizations • Quick Insights
SAP HANA	SAP HANA is a tool, which comprises: • An in-memory HANA database • Data modeling tools • Data provisioning • And HANA administration, making it a suite.	• Provides real-time analysis and decision-making capability. • It enables processing of large amounts of data while the business is going on. Thus, it provides instant real-time insights.	• SAP HANA is only compatible to and thus will run only on SAP or SUSE Linux certified hardware. • Limited hardware compatibility makes wanting to use SAP HANA a costly investment.	• Core process accelerators • Planning, Optimization applications • Sense and response applications i.e. SAP HANA works as a digital core for an Internet of things (IOT).

updating in order to be able to provide responses quickly and not constitute a bottleneck or a blocking factor for the stream can also be not guaranteed. All of these factors would increase processing times and the difficulty to storage some data for the aim of analysis.

Thus, to overcome these various concerns, we will soon be proposing a generic data stream summary model based on Big Data technologies as well. This model should make it possible to collect, transform, store, process and present data streams. It would be defined by different storage structures in cascade where each level of the cascade would correspond to a time granularity that would define when new measures would be calculated. For data collection, we would use technologies as Apache Kafka combined with some technics like random functions to produce stream. Form the streams processing, we would aim to use streaming technologies like Apache Storm and for the storage, we would use NoSQL technologies as Apache HBase. This model would be based over Titled Time Windows to manage the space and time dimensions.

6 Conclusion

With the advent of intensive applications which produce huge volumes of data like fraud detection, roads traffic monitoring, the management of smart electricity meters, etc. it becomes necessary for companies, science, finance, medicine, etc. to be able to analyses and use the results obtained by the mean of this Big data for decisions making. However, processing of these data streams is often confronted to storage and computation constraints caused by the fact that they are generated in swift and continuous manner with variables velocities. In this paper, our main goals were to study and evaluate classical techniques and Big data tools used to generate data stream summaries, the architectures defined and tools that can be used and in what layer of these architectures for to answer user's queries. Thus we have drawn up a state of the art on data streams summaries. Classical methods of stream data summary have many benefits but their implementation entails constraints which are the limits of storage and processing capacities available in traditional systems which that Big data tools can deal with. Indeed, with Big Data solutions, new possibilities are opening up to better understand these limits. In this sense, we have presented different architectures for the processing of data stream of a Big Data nature which involves in each of their layers a certain number of tools (Section 4). These architectures have thus been implemented by various approaches combining Big Data technologies with those of NoSQL in order to overcome the problem of processing and storage relating

to data stream. These different proposals are most often distributed over clusters.

References

[1] A. Murphy, C. Taylor, C. Acheson, J. Butterfield, Y. Jin, P. Higgins, R. Collins, C. Higgins, "Representing financial data streams in digital simulations to support data flow design for a future Digital Twin," Robotics and Computer-Integrated Manufacturing, **61**, 101853, 2020, doi:https://doi.org/10.1016/j.rcim.2019.101853.

[2] R. Verde, A. Balzanella, A Spatial Dependence Measure and Prediction of Georeferenced Data Streams Summarized by Histograms, chapter 5, 99–117, John Wiley & Sons, Ltd, 2020, doi:https://doi.org/10.1002/9781119695110.ch5.

[3] P. K. Padigela, R. Suguna, "A Survey on Analysis of User Behavior on Digital Market by Mining Clickstream Data," in K. S. Raju, A. Govardhan, B. P. Rani, R. Sridevi, M. R. Murty, editors, Proceedings of the Third International Conference on Computational Intelligence and Informatics, 535–545, Springer Singapore, Singapore, 2020.

[4] R. Sahal, J. G. Breslin, M. I. Ali, "Big data and stream processing platforms for Industry 4.0 requirements mapping for a predictive maintenance use case," Journal of Manufacturing Systems, **54**, 138–151, 2020, doi:https://doi.org/10.1016/j.jmsy.2019.11.004.

[5] F. Zheng, Q. Liu, "Anomalous Telecom Customer Behavior Detection and Clustering Analysis Based on ISP's Operating Data," IEEE Access, **PP**, 1–1, 2020, doi:10.1109/ACCESS.2020.2976898.

[6] L. Rutkowski, M. Jaworski, P. Duda, Basic Concepts of Data Stream Mining, 13–33, Springer International Publishing, Cham, 2020, doi:10.1007/978-3-030-13962-9_2.

[7] S. Zhang, F. Zhang, Y. Wu, B. He, P. Johns, "Hardware-Conscious Stream Processing: A Survey," 2020.

[8] T. Dilova, Y. Anastasov, V. Yordanov, N. Milovanov, D. Dzhonova, "Arrangement for enriching data stream in a communications network and related method," 2020.

[9] B. Tidke, R. Mehta, D. Rana, H. Jangir, "Topic Sensitive User Clustering Using Sentiment Score and Similarity Measures: Big Data and Social Network," International journal of web-based learning and teaching technologies, **15**(2), 34–45, 2020.

[10] B. Csernel, Résumé généraliste de flux de données. PhD thesis, Ph.D. thesis, 2008.

[11] B. Babcock, S. Babu, M. Datar, R. Motwani, J. Widom, ""Models and Issues in Data Stream Systems."," 1–16, 2002, doi:10.1145/543613.543615.

[12] R. Chiky, Résumé généraliste de flux de données. PhD thesis, Ph.D. thesis, 2009.

[13] C. Midas, A. Inria, "Résumé généraliste de flux de données," Extraction et gestion des connaissances (EGC'2010), Actes, 26 au 29 janvier 2010, Hammamet, Tunisie, 2010.

[14] B. Csernel, F. Clerot, G. Hébrail, "Datastream clustering over tilted windows through sampling," 2006.

[15] C. Aggarwal, J. Han, J. Wang, P. Yu, T. Watson, R. Ctr, "A Framework for Clustering Evolving Data Streams," 2003.

[16] N. Alon, Y. Matias, M. Szegedy, "The Space Complexity of Approximating the Frequency Moments," Journal of Computer and System Sciences, **58**, 137–147, 1999, doi:10.1006/jcss.1997.1545.

[17] P. Jukna, Probabilistic Counting, 41–51, 2011, doi:10.1007/978-3-642-17364-6_3.

[18] C. Aggarwal, J. Han, J. Wang, P. Yu, T. Watson, R. Ctr, "A Framework for Clustering Evolving Data Streams," 2003.

[19] Y. Tillé, Théorie des sondages: échantillonnage et estimation en populations finies, 2001.

[20] P. Flajolet, G. Martin, "Probabilistic Counting Algorithms for Data Base Applications," Journal of Computer and System Sciences, **31**, 182–209, 1985, doi:10.1016/0022-0000(85)90041-8.

[21] A. Ebrahim, J. Khlaifat, "An Efficient Hardware Architecture for Finding Frequent Items in Data Streams," 113–119, 2020, doi:10.1109/ICCD50377.2020.00034.

[22] B. Bloom, "Space/Time Trade-offs in," 2002.

[23] R. Chiky, B. Defude, G. Hébrail, "Définition et diffusion de signatures sémantiques dans les systèmes pair-à-pair." 463–468, 2006.

[24] E. Eydi, D. Medjedovic, E. Mekić, E. Selmanovic, Buffered Count-Min Sketch, 249–255, 2018, doi:10.1007/978-3-319-71321-2_22.

[25] I. Kutsenko, "OPTIMIZATION OF QUERY IN RELATIONAL DATABASES," EurasianUnionScientists, **2**, 52–58, 2018, doi:10.31618/ESU.2413-9335.2018.2.56.52-58.

[26] P. Ravi, D. Haritha, P. Niranjan, "A Survey: Computing Iceberg Queries," International Journal of Engineering & Technology, **7**, 791, 2018, doi:10.14419/ijet.v7i2.7.10979.

[27] A. Calì, "Querying web data," **9050**, 2015.

[28] C. Myers, Random Sampling, 91–106, 2020, doi:10.1201/9780429292002-ch07.

[29] B. Babcock, M. Datar, R. Motwani, "Sampling From a Moving Window Over Streaming Data," Proc. SODA, 2001, doi:10.1145/545381.545465.

[30] R. Gemulla, W. Lehner, P. Haas, "A Dip in the Reservoir: Maintaining Sample Synopses of Evolving Datasets." 595–606, 2006.

[31] K. Ying, J. Wu, G. Dai, C. Miao, C. Fan, Q. Chen, "A kind of data aggregation technology based on linear time probabilistic counting algorithm," Chinese Journal of Sensors and Actuators, **28**, 99–106, 2015, doi:10.3969/j.issn.1004-1699.2015.01.018.

[32] W. Ren, X. Lian, K. Ghazinour, "Efficient Join Processing Over Incomplete Data Streams," 209–218, 2019, doi:10.1145/3357384.3357863.

[33] Y. Wang, T. Peng, J.-Y. Han, L. Liu, "Density-Based Distributed Clustering Method," Ruan Jian Xue Bao/Journal of Software, **28**, 2836–2850, 2017, doi:10.13328/j.cnki.jos.005343.

[34] N. Goyal, P. Goyal, K. Venkatramaiah, "An Efficient Density Based Incremental Clustering Algorithm in Data Warehousing Environment," 2021.

[35] Y. Pitarch, Résumé de Flots de Données : motifs, Cubes et Hiérarchies, Ph.D. thesis, 2011, thèse de doctorat dirigée par Poncelet, Pascal et Poncelet, Pascal Informatique Montpellier 2 2011.

[36] K.-P. Chan, A. Fu, "Efficient time series matching by wavelets," 126–133, 1999, doi:10.1109/ICDE.1999.754915.

[37] E. Stollnitz, T. Derose, D. Salesin, Wavelets for computer graphics - theory and applications, 1996.

[38] P. Rathika, G. Sreeja, "Wavelet Based Histogram Technique for Tumour Detection in Digital Mammograms," 2012.

[39] J. Vitter, M. Wang, B. Iyer, "Data Cube Approximation and Histograms via Wavelets (Extended Abstract)," 1998.

[40] Y. Matias, J. Vitter, M. Wang, "Dynamic Maintenance of Wavelet-Based Histograms," 2000.

[41] A. Gilbert, S. Guha, P. Indyk, Y. Kotidis, S. Muthukrishnan, M. Strauss, "Fast, Small-Space Algorithms for Approximate Histogram Maintenance," Conference Proceedings of the Annual ACM Symposium on Theory of Computing, 2001, doi:10.1145/509907.509966.

[42] G. Cormode, M. Garofalakis, D. Sacharidis, "Fast Approximate Wavelet Tracking on Streams," volume 3896, 4–22, 2006, doi:10.1007/11687238_4.

[43] A. Gilbert, Y. Kotidis, S. Muthukrishnan, M. Strauss, "Surfing Wavelets on Streams: One-Pass Summaries for Approximate Aggregate Queries," 2001.

[44] N. Singh, "Emerging Trends in Technologies for Big Data," The International Technology Management Review, **5**, 202, 2015, doi:10.2991/itmr.2015.5.4.4.

[45] S. Ahsaan, H. Kaur, S. Naaz, Real-Time Data Processing Topology, 2020.

[46] F. Gurcan, M. Berigel, "Real-Time Processing of Big Data Streams: Lifecycle, Tools, Tasks, and Challenges," 1–6, 2018, doi:10.1109/ISMSIT.2018.8567061.

[47] H. Isah, F. Zulkernine, "A Scalable and Robust Framework for Data Stream Ingestion," 2018.

[48] C. Lavergne, "Statistique et analyse des données," 2021.

[49] J. Kreps, "Kafka : a Distributed Messaging System for Log Processing," 2011.

[50] D. Vohra, Apache Flume, 287–300, 2016, doi:10.1007/978-1-4842-2199-0_6.

[51] R. Young, S. Fallon, P. Jacob, "An Architecture for Intelligent Data Processing on IoT Edge Devices," 2017, doi:10.1109/UKSim.2017.19.

[52] F. Khan, "Apache kafka with real-time data streaming," 2021.

[53] S. Rooney, P. Urbanetz, C. Giblin, D. Bauer, F. Froese, L. Garcés-Erice, S. Tomic, "Kafka: the Database Inverted, but Not Garbled or Compromised," 3874–3880, 2019, doi:10.1109/BigData47090.2019.9005583.

[54] S. Rooney, P. Urbanetz, C. Giblin, D. Bauer, F. Froese, L. Garcés-Erice, S. Tomic, "Debezium stream changes from your database," 2019, doi:https://debezium.io/doc.

[55] F. López, E. Cruz, "Literature review about Neo4j graph database as a feasible alternative for replacing RDBMS," Industrial Data, **18**, 135, 2015, doi:10.15381/idata.v18i2.12106.

[56] R. D. e. D. G. Hojjat Jafarpour, "KSQL: Streaming SQL Engine for Apache Kafka," in Proceedings of the 22nd International Conference on Extending Database Technology (EDBT) ISBN, **18**, 2019, doi:978-3-89318-081-3onOpenProceedings.org.

[57] B. M. West, "Integrating IBM Streams with Apache NiFi - Streamsdev," https://developer.ibm.com/streamsdev/docs/integrating-ibm-streams-apache-nifi/, 2018.

[58] D. Baev, "Managing Data in Motion with the Connected Data Architecture," in 4th Big Data & Business Analytics Symposium, **2**, 2017.

[59] J. P. Young Roger, Fallon Sheila, "An Architecture for Intelligent Data Processing on IoT Edge Devices," 10.1109/UKSim.2017.19., **2**, 2017.

[60] P. H. et Mohamed Rilwan, "Big Data Ingestion: Flume, Kafka and NiFi," Big Data, Software Architect, **2**, 2017, doi:https://tsicilian.wordpress.com/2017/07/06/big-data-ingestion-flume-kafka-and-nifi/.

[61] J. Hiba, H. Hadi, A. Hameed Shnain, S. Hadishaheed, A. Haji, "BIG DATA AND FIVE V'S CHARACTERISTICS," 2393–2835, 2015.

[62] M. Goudarzi, "Heterogeneous Architectures for Big Data Batch Processing in MapReduce Paradigm," in IEEE Transactions on Big Data, **65**, 18–33, 2019, doi:10.1109/TBDATA.2017.2736557.

[63] N. S. e. G. M. Nasiri H., "Evaluation of distributed stream processing frameworks for IoT applications in Smart Cities," J Big Data, **6**, 52, 2019, doi:10.31618/ESU.2413-9335.2018.2.56.52-58.

[64] S. E. e. S. H. Paris Carbone, "Apache Flink: Stream and Batch Processing in a Single Engine," Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, **6**, 52, 2015.

[65] G. Yanfei, J. Rao, C. Jiang, X. Zhou, "Moving MapReduce into the Cloud with Flexible Slot Management and Speculative Execution," IEEE Transactions on Parallel and Distributed Systems, **28**, 1–1, 2016, doi:10.1109/TPDS.2016.2587641.

[66] A. S. e. S. M. Toshifa, "Big Data Hadoop Tools and Technologies: A Review," International Conference on Advancements in Computing & Management (ICACM-2019), 2019.

[67] M. Dayalan, "MapReduce: Simplified Data Processing on Large Cluster," International Journal of Research and Engineering, **5**, 399–403, 2018, doi:10.21276/ijre.2018.5.5.4.

[68] N. S. I. e. a. Koo J., Faseeh Qureshi, "IoT-enabled directed acyclic graph in spark cluster," J Cloud Comp, **9**, 50, 2020, doi:https://doi.org/10.1186/s13677-020-00195-6.

[69] R. W. P. D. T. A. e. a. Zaharia, Matei Xin, "Apache spark: A unified engine for big data processing," Communications of the ACM, **59**, 56–65, 2016, doi:10.1145/2934664.

[70] A. B. et V. Voityshyn, "Apache storm based on topology for real-time processing of streaming data from social networks," Lviv I. EEE First International Conference on Data Stream Mining & Processing (DSMP), 345–349, 2016, doi:10.1109/DSMP.2016.7583573.

[71] L. H. S. S. S. I. Zaharia M, Das T, "Discretized streams: an efficient and fault-tolerant model for stream processing on large clusters." HotCloud 12:10, 2012.

[72] Y. P. N. R. J. B. I. G. Shadi A. Noghabi, Kartik Paramasivam, R. H. Campbell, "Samza: Stateful Scalable Stream Processing at LinkedIn," Proceedings of the VLDB Endowment, **10 No. 12**, 345–349, 2017, doi:Copyright2017VLDBEndowment2150-8097/17/08.

[73] e. a. Ankit Toshniwal, "Storm @Twitter," SIGMOD'14, June 22–27, 2014, Snowbird, Utah, USA.Copyright, **10 No. 12**, 345–349, 2014, doi:2014ACM978-1-4503-2376-5/14/06...\$15.00.http://dx.doi.org/10.1145/2588555.2595641.

[74] T. S. Singh MP, Hoque MA, "A survey of systems for massive stream analytics," arXiv preprint, 2016, doi:arXiv:1605.09021.

[75] S. S. G. B. G. R. Hirzel M, Soulé R, "A catalog of stream processing optimizations," CM Comput Surv CSUR, **46(4)**, 46–50, 2014, doi:Returntoref28inarticle.

[76] W. P. D. T. A. M. D. A. M. X. R. J. V. S. F. M. e. a. Zaharia M, Xin RS, "Apache spark: a unified engine for big data processing," Commun ACM, **59(11)**, 2016, doi:arXiv:1605.09021.

[77] O. AC., "Storm or spark: choose your real-time weapon." 2018, doi:http://www.infoworld.com/article/2854894/application-development/spark-and-storm-for-real-time-computation.html.

[78] S. D. e. S. D. Lekha R.Nair, "Applying spark based machine learning model on streaming big data for health status prediction," Computers & Electrical Engineering, **65**, 393–399, 2018, doi:http://www.infoworld.com/article/2854894/application-development/spark-and-storm-for-real-time-computation.html.

[79] S. S. G. V. A. M. e. C. C. B. Saha, H. Shah, "Apache tez: A unifying framework formodeling and building data processing applications," ACM SIGMOD, 2015.

[80] V. B. T. B. C. C. A. C. J. E. M. G. D. H. M. J. e. a. M. Kornacker, A. Behm, "Impala: A modern, open-source sql engine for hadoop," CIDR, 2015.

[81] D. S. A. B. H.-I. H. R. R. e. a. D. J. DeWitt, S. Ghandeharizadeh, "The gamma databasemachine project," IEEE TKDE, 1990.

[82] F. M. K. V. K. C.-M. S. P. J. R. K. T. S. Kulkarni S, Bhagat N, "witter heron: stream processing at scale," In: Proceedings of the 2015 ACM SIGMOD international conference on management of data, 239–50, 2015.

[83] R. K. ejariwal A, Kulkarni S, "KReal time analytics: algorithms and systems," Proc VLDB Endow., **8(12)**, 2040–1, 2015.

[84] Z. P., "Comparison of apache stream processing frameworks," Cakesolutions, **8(12)**, 2040–1, 2018, doi:http://www.cakesolutions.net/teamblogs/comparison-of-apache-streamprocessing-frameworks-part-1.

[85] N. I. e. K. T. Xiufeng Liu, "Survey of real-time processing systems for big data," IDEAS '14 - ACM 18th International Database Engineering & Applications Symposium, **8(12)**, 2040–1, 2014, doi:10.1145/2628194.2628251.

[86] A. Lakshman, P. Malik, "Cassandra — A Decentralized Structured Storage System," Operating Systems Review, **44**, 35–40, 2010, doi:10.1145/1773912.1773922.

[87] O. M. E. P. Henrique Zanúz, Bruno Raffin, "In-Transit Molecular Dynamics Analysiswith Apache Flink," In Situ Infrastructures for Enabling Extreme-scale Analysis andVisualization, DALLAS, United States, 1–8, 2018, doi: ISAV2018.

[88] N. J. Z. S. P. C. S. A. H. L. P. W. Ashish Thusoo, Joydeep Sen Sarma, R. Murthy, "Hive - A Warehousing Solution Over a Map-Reduce Framework," VLDB'09 Lyon, France, 239–50, 2009.

[89] e. a. Jesús Camacho-Rodríguez, "Apache Hive: From MapReduce toEnterprise-grade Big Data Warehousing," SIGMOD '19 Amsterdam, Netherlands, 239–50, 2019.

[90] M. M. N. Dr MD NADEEM AHMED, AASIF AFTAB, "A Technological Survey On Apache Spark And Hadoop Technologies." INTERNATIONAL JOURNAL OF SCIENTIFIC &TECHNOLOGY RESEARCH, **9 ISSUE 01**, 239–50, 2020.

[91] G. L. G. Pal, K. Atkinson, "Near Real-Time Big Data Stream Processing," 4th International Conference for Convergence in Technology (I2CT), Mangalore, India, **9 ISSUE 01**, 1–7, 2018, doi:10.1109/I2CT42659.2018.9058101.

[92] K. K. Wahid A., "Cassandra—A Distributed Database System: An Overview." In: Abraham A., Dutta P., Mandal J., Bhattacharya A., Dutta S. (eds) Emerging Technologies in Data Mining and Information Security. Advances in Intelligent Systems and Computing, **755**, 1–7, 2019, doi:Springer,Singapore.https://doi.org/10.1007/978-981-13-1951-8_47.

[93] S. T. Krishnan Hema, Elayidom M.Sudheep, "MongoDB – a comparison with NoSQL databases," International Journal of Scientific and Engineering Research, **7**, 1035–1037, 2016.

[94] C. e. L. M. Nadeem Qaisar, Mehmood et Rosario, "Modeling temporal aspects of sensor data for MongoDB NoSQL database," Journal of Big Data, **9 ISSUE 01**, 239–50, 2017.

[95] S. D. e. M. L. S.K. Goyal, "Effective handling of patient health records using NoSQL: MONGOD," Research Journal of Pharmaceutical, Biological and Chemical Sciences, **7(4)**, 134–138, 2016.

[96] P. S. et Stirman K., "Schema design for time series data in mongodb," http://blog.mongodb.org, **30**, 134–138, 2013.

[97] N. Team, "Neo4j Streams," https://neo4j.com/press-releases/neo4j-streams-kafta-confluent/, **30**, 134–138, 2020.

[98] N. Team, "Neo4j Streams," https://neo4j.com/press-releases/neo4j-streams-kafta-confluent/, **30**, 134–138, 2020.

[99] M. S. et Narendra Singh Yadav, "A Practical Approach to Process Streaming Data using Graph Database," RInternational Journal of Computer Applications, **117(23)**, 28–32, 2015, doi:10.5120/20695-3588.

[100] Redis, "A Practical Approach to Process Streaming Data using Graph Database," Site web de Redis, 2020, doi:https://redis.io/.

[101] A. Quinton, "Storm Real-time Processing Cookbook Efficiently process unbounded streams of data in real time," Packt Publishing, 2013, doi: https://redis.io/.

[102] G. Shilpi, Saxena et Saurabh, "Practical Real-Time Data Processing and Analytics, Distributed Computing and Event Processing using Apache Spark, Flink, Storm, and Kafka," Packt Publishing, 2017, doi:https://redis.io/.

[103] G. Shilpi, Saxena et Saurabh, "Big Data Analysis: Apache Storm Perspective," International Journal of Computer Trends and Technology, **19(1)**, 9–14, 2015, doi:10.14445/22312803/IJCTT-V19P103.

[104] L. W. G. P. M. I. R. H. D. Färber Franz, May Norman, "The SAP HANA database - An architecture overview," IEEE Data Eng. Bull, **35**, 28–33, 2012.

[105] A. K. U. Amrapali Bansal, "Microsoft Power BI," International Journal of Soft Computing and Engineering (IJSCE) ISSN, **7**, 2231–2307, 2017.

[106] S. Thabet, "Concepts and Tools for Marketing Intelligence Development," International Journal of Innovation in the Digital Economy, **4**, 15–3, 2013, doi:10.4018/jide.2013070102.

[107] e. a. Färber Franz, "SAP HANA database: Data management for modern business applications," SIGMOD Record, **40**, 45–51, 2011, doi:10.1145/2094114.2094126.

[108] M. F. e. M. F. K. D. K. D. Fischer Fabian, Fischeruni Fabian, "Real-Time Visual Analytics for Event Data Streams," Proceedings of the ACM Symposium on Applied Computing, 2011, doi:10.1145/2245276.2245432.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 4, 431-438 (2021)

www.astesj.com

Special Issue on Innovation in Computing, Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

A Design of Anthropomorphic Hand based on Human Finger Anatomy

Zixun He[1], Yousun Kang[2], Duk Shin[3,*]

[1]*Department of Human Centered Science and Biomedical Engineering, Tokyo Institute of Technology, Yokohama, 226-8503, Japan*

[2]*Department of Applied Computer Science, Tokyo Polytechnic University, Kanagawa, 243-0297, Japan*

[3]*Department of Electronics and Mechatronics, Tokyo Polytechnic University, Kanagawa, 243-0297, Japan*

ARTICLE INFO

Article history:
Received: 21 April, 2021
Accepted: 09 August, 2021
Online: 28 August, 2021

Keywords:
Human hand biomechanics
Prosthetic hand
EMG signals
Quality of life

ABSTRACT

In the past decade, multiple anthropomorphic prosthetic hands have been developed to replace the role of human hands. Prostheses should not only replace the functions of human hands in functionality but also replicate human hands in appearance and sense of body-belonging intuitively. Human fingers have very delicate and complex structures, and it is these complex structures that make our hands dexterity. This study proposes a design based on the anatomical characteristics of the human hand. The proposed design replicates human fingers from bones, ligaments, extensor hoods, and extensor mechanism of tendon, intended to develop a prosthesis that has the same flexibility and appearance as human hand. To evaluate the performance of the proposed prosthetic in daily life, we conducted grasping experiments on common objects. It is successfully proved that the proposed design helps to improve the grasping performance of the artificial hand and has a natural appearance. In this paper, our design succeeds to improve the grasping performance of the artificial hand and gain natural appearance.

1. Introduction

This paper is an extended paper of our work initially presented in International Symposium on Community-centric Systems (CcS 2020) [1]. The human hand is an important tool for us to interact with and perceive the physical environment. Upper limbs loss is one of the most disabling diseases that a person may experience, which can severely affect amputees Activities of daily living (ADLs) and working abilities. In the United States alone, there are approximately 50,000 people who have lost upper limbs, and it is predicted that the numbers of upper limb loss will increase over time [2]. According to the investigation [3], different amputees may have different requirements for prostheses. Besides performance of prosthetics devices, a high level of personification and simplified operating system are also required. Prosthetic developers need to consider these different requirements, which bring challenges to the development of prosthetic hands.

In order to improve the acceptability of the prosthesis [4,5], various artificial hand designs and models have been proposed. Commercially available advanced prosthetic hands, such as Michelangelo hand (by Otto Bock), are easy to use and can help amputee grip the objects. However, the price of commercially available prosthetics is very expensive. It is difficult for people who have lost upper limbs to afford this financial burden. In recent years, with the development of new manufacturing technologies such as fused deposition modeling (FDM) 3D printers, low-cost and easy-to-make open-source prostheses have also been designed [6,7].

Most research groups aim to replicate the movement mechanism of human hands in a mechanized design, improving the flexibility and performance of the prosthetics. In the past two decades, the research community has developed some novel artificial hands [8-10]. Lee et al. proposed the design of a 9- DOF bio-mimetic robot hand. This hand has four under-actuated fingers, each of finger is equipped with a tactile sensor and driven by two linear actuators coupled (together) [11]. In order to further reduce the weight of the prosthetic hand, Mohammadi et al. used soft materials and developed a lightweight soft robotic prosthetic hand that has synergy-based motion and cable-driven actuation system. In addition, successfully performed three different grips with high dynamic power (21.5N) and fast finger bending speed. Although these artificial hands have excellent grip and performance, they

*Corresponding Author: Duk Shin, 1583, Iiyama, Atsugi, Kanagawa, 243-0297, Japan, Email: d.shin@eng.t-kougei.ac.jp

have not yet reached the level of human dexterity. The reason for that is these designs adopt the structure of the human body in the mechanized way, the incomplete human hand-like motion mechanism limited to the motion performance of the manipulator [12].

In [13], the author developed the anatomically corrected tested (ACT) hand by replicating the bone structures of the fingers, which is the first robotic hand developed based on human hand interpretation. Although its mechanism is still based on hinges and universal joints, it still exhibits a more flexible grip than previous manipulators. In [14], the author designed a robot hand through high bionics, and a flexor mechanism for an elastic pulley mechanism to realize the joint movement of human fingers. These studies of highly bionic anthropomorphic robotic hands have showed the possibility of replicating the structure of the human hand on the prosthesis to make the movements more dexterity.

Electromyography signal (EMG) is a signal that can reflect the user's movement intention. In recent years, it has been widely used to control prostheses or robots. Although this kind of EMG-based control is widely used for pattern recognition and rehabilitation, the nature of EMG signal is random or non-stationary and will be affected by nearby muscles that generate additional noise [15]. Therefore, in order to control the prosthetic hand accurately, it is necessary to extract the features related to the selected action patterns, and then use support vector machine (SVM), linear discriminant analysis (LDA), Artificial neural network (ANN) and other machine learning to classify the action. The correct selection of features is critical to classification performance. However, because of different electrode placement schemes, number of channels, and EMG signal preprocessing methods in different studies, it is difficult to compare clearly which features can significantly improve the accuracy of classification [16]. Deep learning methods have recently been used in the classification of EMG signals and have shown strong performance [17]. Côté-Allard et al. use Convolutional Neural Networks (CNN) to classify the actions of 7 forelimbs with an accuracy of 97.9% [18]. It can be said that in a laboratory environment, such a classifier with high computational cost and powerful performance is satisfactory, but it is an important practical consideration to ensure the effectiveness while reducing the computational burden when building the system. In addition, reducing the number of electrodes to improve comfort also needs to be considered for the prosthetic hand system. Tavakoli et al. use only 2 EMG channels to classify four types of gestures, and the classification accuracy rate exceeds to 90% [19]. In daily life, we use hand movements to grasp objects or interact with the environment [20,21]. Therefore, by defining the minimum number of daily grasping patterns, the number of electrodes can be reduced, and the simplicity of the system can be maintained.

In this article, we designed an anthropomorphic prosthetic hand based on anatomical structure. The four-finger and thumb model of the prosthetic hand is modeled based on laser scanned hand bone data on cadaveric. This can reduce the rejection of the amputee to the prosthesis after the artificial skin is worn on the prosthetic hand, and the movement angle of each finger will not exceed the movable range of the original finger. Incorporate soft tissue structures such as ligaments, tendons and tendon sheaths that have a significant impact on the flexibility of the fingers at each joint to replicate the movement characteristics of the human hand. Then he conducted a grasping experiment on the prosthetic hand to test the performance of our prosthesis in daily life. In order to control the prosthetic hand using Electromyogram (EMG) signals, an EMG control system is designed for the artificial hand by an artificial neural network (ANN). The EMG control system is used to control the four main hand positions used in daily life: Power grasps, Precision grasps, Lateral grasps and relax state [21].

2. Development of prosthetics

2.1. Hand structures and the Bones Models

We designed highly anthropomorphic prosthetic hand based on the anatomy. First, the skeletal structure of the human finger movement mechanism is mainly determined by the metacarpal and phalangeal bones. The human finger comprises the distal, middle, proximal phalanges and metacarpals. These four bones make up the distal interphalangeal (DIP), proximal interphalangeal (PIP) and metacarpophalangeal (MCP) joints of the hand. The thumb finger has only two phalanges, so it has only one interphalangeal joint. Every finger joint has soft tissues called ligaments [22].

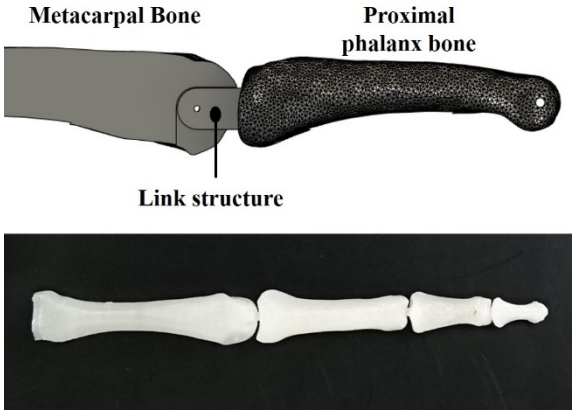


Figure 1: Designed Finger model

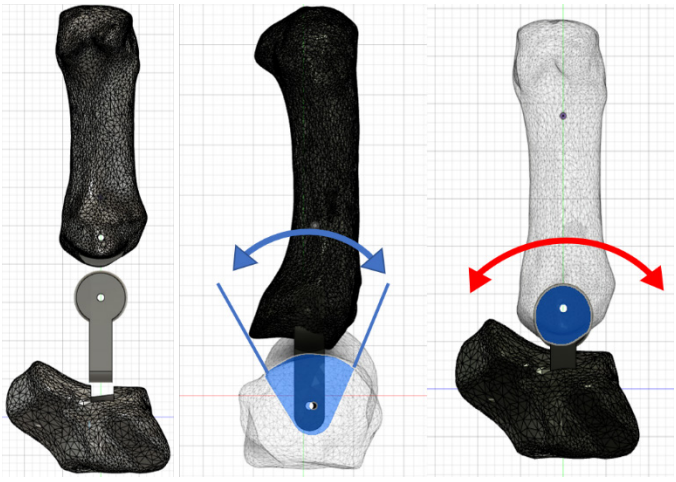


Figure 2: Proposed thumb design

We used laser-scanned bone data as the basic model of the prosthesis [23]. Since the size of the prosthetic hand skin in the market vendor is larger than the bone model of this hand, we adjusted the size of the model. In order to improve the efficiency of power transmission, a link structure is designed between each joint as shown as Figure 1.

Table 1: Finger Flexion degree [deg]

DIP Joint	PIP Joint	MCP Joint
0 - 77	0 - 105	0 - 90

The range of motion of the finger joints will vary between different fingers and different people, depending on the bone shape, tendons and muscle structure characteristics of the hand. By referring to the previous research on joint structure [24], in this research, their range of motion is uniformly designed, as shown in Table 1.

Although the thumb comprises only two bones, it has great freedom of movement. The reason is attributed to the carpometacarpal (CMC) joint composed of the thumb metacarpal bone and trapezium bone. The trapezium bone is shaped like a saddle used for horse riding, allowing the thumb to move in a larger range. The movement of the thumb is caused by sliding the metacarpal bone of the thumb along the trapezium bone at the CMC joint back and forth (extension / flexion), left and right (abduction / adduction) or both at the same time [25]. The following method is used to reproduce the thumb structure to ensure correct thumb movement. In order to reproduce the adduction and abduction functions of the thumb, the connecting rod structure on the metacarpal model additionally uses a rotary joint, as shown in Figure 2. And a groove is designed on the trapezium bone to achieve flexion and extension movement.

2.2. Design of Joint Ligament and Tendon

Ligaments are a type of fibrous connective tissue that connects bone to bone. These tissues can be found in all joints. The range of motion of each finger joint is limited by the length of the ligament. There are collateral ligaments on both sides of the hand bones, which are attached to the volar plates. Cartilage structures called volar plates are inserted on both sides of the bone joints. Volar plates and ligaments form a joint pack to prevent joint overextend and enhance joint stability. There are two sets of tendons in the human hand to stretch and bend the fingers. They are the extensor and flexor tendons. The flexor tendons extend from the forearm and, finally, branches become the flexor digitorum superficialis (FDS) tendon and the flexor digitorum profundus (FDP) tendon. The FDS tendon is fixed to the intermediate phalange bones of the PIP joint. The FDP tendon is inserted into the bottom of the palm of the distal phalanges. When the finger is bent, the PIP joint attached to the FDS tendon will bend first. As the movement progresses, the FDP tendon will play a role in bending the DIP joint. These two flexor tendons are firmly attached to the phalanx through the tendon sheath. The extensor tendons are combined with extensor hood that regulate joint extension and flexion. Extensor hood is a complex mesh structure that wraps the phalanx directly from the back of the hand. Its structure is shown in Figure 5 (a) [26]. The first layer of the dorsal aponeurosis of the extensor hood is inserted into the base of the DIP joint and divided into two small tendons at the PIP joint. The second layer of the dorsal aponeurosis of the extensor hood is located at the bottom of the PIP joint.

According to the ligament, the replicated model of the structure of the volar plate is shown in Figure 3. The volar plate is made of High Elasticity Rubber Plate. Anchor the rubber volar plate to each joint with screws. The ligament is made of rubber thread, and the natural length of the ligament is adjusted when inserted into the volar plate and both sides of the hand bone, so that the fingers of the artificial hand can be naturally bent like a human finger when they are naturally relaxed. Thus, the prosthetic hand will not be kept in a straight and rigid motion, and the prosthetic hand will be more personified.

Figure 3: Structure in anatomy and reproduced model

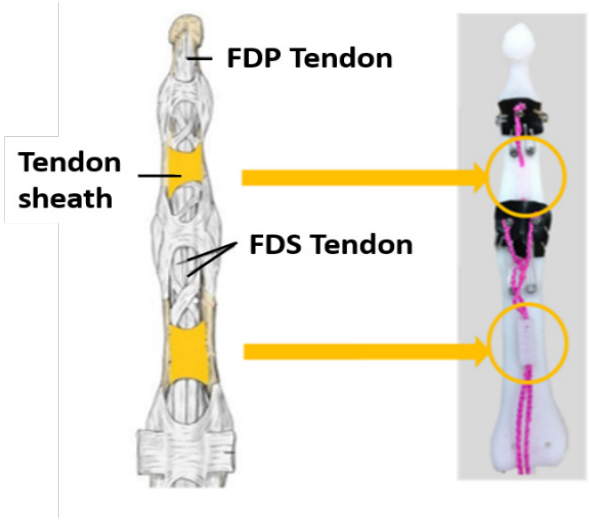
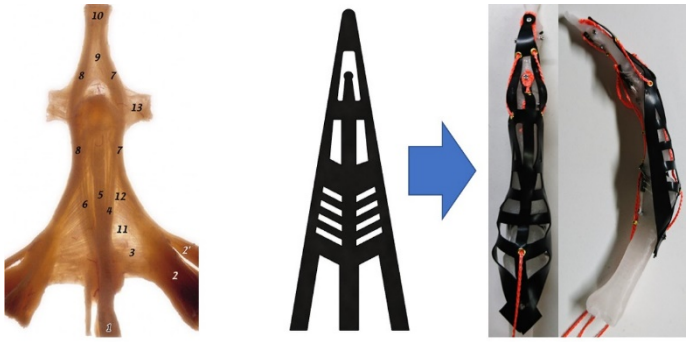


Figure 4: Tendon and Tendon sheaths

The flexor and extensor tendons composed of FDS tendons and FDP tendons are made of 0.33 mm polyethylene wire (Figure 4), with chief strength (250 N breaking strength) and good flexibility. The tendons are woven to prevent the volar plate and tendon sheath from being worn out because of the long-term movement of the tendons.



(a) (b)

Figure 5: Hood parts; (a) Extensor hood Structure [26] (b) Reproduced extensor hood sheath

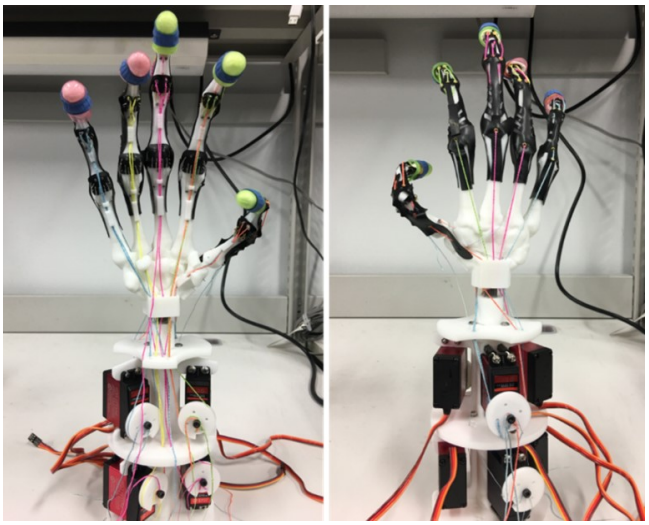


Figure 6: Configuration of the prosthesis

Channel	Muscle
1	Flexor digitorum superficialis muscle
2	Flexor digitorum profundus muscle
3	Extensor pollicis longus muscle
4	Extensor digitorum muscle

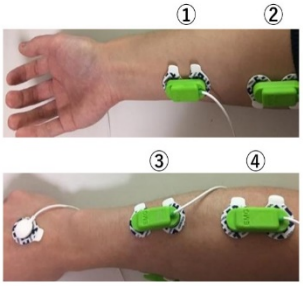


Figure 7: Location of the EMG electrodes

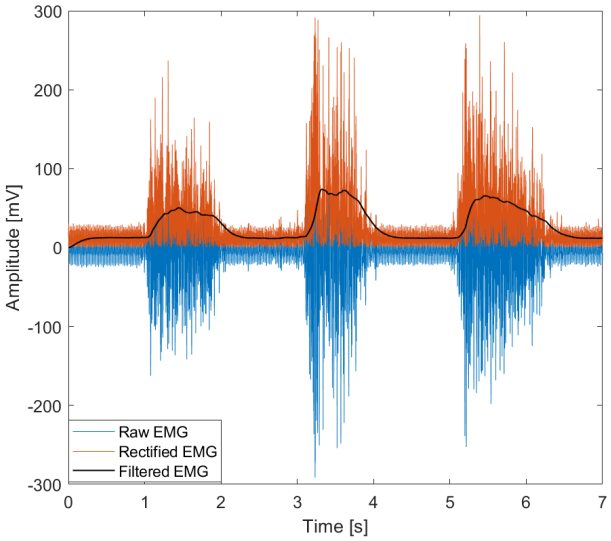


Figure 8: Pre-Processing of EMG signal

Table 2: The specifications of DS3218 Servo

Dimensions	40 x 20 x 40.5 mm
Weight	60 g
Stall Torque	21.5 kg・cm (6.0V)

The extensor tendon has a central fiber bundle that reaches the base of the intermediate phalanges while two outer fiber bundles are anchored at the base of the distal phalanx. We simplified the structure of the extensor hood and made the extensor hood using a highly elastic rubber sheet to reproduce the function of the extensor hood to adjust joint extending or bending. Figure 5 shows the finger after all components have been assembled.

2.3. Configuration of the prosthesis and Control Method

After all fingers are assembled, combined them with the wrist part. The complete prosthesis is shown in Figure 6. The tendon of the finger is fixed on the pulley, and the pulley is connected to the servo. By pulling the tendon with the servo motor, the finger joint bent or extended. Six Servo motors (Ds servo DS3218) are used to control the prosthetic hand. The servo specifications are shown in Table 2. Considering Thumb has 2 sets of movement, we use 2 servos to control Adduction or Abduction, and Flexion or Extension of the thumb, respectively. Other fingers are each controlled by one servo. In order to use the intuitive operation method to control the prosthetic hand in the grasping experiment, an operating system was developed using the right-hand glove part of Perception Neuron (Noitom, Beijing, China). Perception Neuron's joint obtain the movement of the finger detected in three-dimensional coordinates. The servo is controlled by the DIP joint angle detected by the sensor neuron. Calculate the bending angle of the finger according to formula (1).

$$\theta = 180 - arccos\frac{z}{\sqrt{x^2 + y^2 + z^2}} \quad (1)$$

Here, x, y, and z are three-dimensional coordinates detected from the sensory neuron of the DIP joint. The calculated angle data will be sent to Arduino UNO to control the servo.

2.4. EMG signals Recording and Processing

The author showed that increasing the number of electrodes does not always improve the accuracy of gesture classification [27]. In addition, problems such as using many electrodes or the long time for classification can also make amputee annoying.

Based on these basic demands, we decided on the number of electrodes and the classification time. The connection points of the EMG sensor and the surface muscles selected to measure the EMG signal are shown in Figure 7. These muscles control the four movements (power grasp, precision grasp, lateral grasp and relax) used for classification. The reason for choosing these 4 kinds of motions is because power grasps, precision grasps, Lateral grasps are most used in activities of daily living and user will also need a relax mode for an inactive position.

We gained EMG signals using the BITalino EMG sensor. By using MATLAB's Bitalino'Toolbox, the EMG signal is recorded on MATLAB at a sampling rate of 1000 Hz [28]. Before calculating the features of the EMG signal, we rectified and filtered the four EMG signal channels. All preprocessing are performed using MATLAB. A second-order Butterworth low-pass filter with a cutoff frequency of 5 Hz is used to filter the EMG signal to reduce the noise of the EMG signal [29]. Figure 8 shows the raw EMG signal and filtered EMG signal. During training and experiment, all features were analyzed using a sliding time

window of 200 (ms) with an overlap of 100 (ms). The calculation processing delay is about 300 (ms). The control is quasi real-time.

2.5. Feature Extraction

Three types of signal features: time-domain features, frequency-domain features, and time-frequency features are usually used for EMG classification. In order to improve the accuracy of classification, two or more features are usually combined. We tested 8 EMG signal features proposed by Phinyomark et al. in the previous work to select features [30]. These features are: Integrated EMG (IEMG), Mean Absolute Value (MAV), Simple Square Integral (SSI), Variance of EMG (VAR), Root Mean Square (RMS), Waveform Length (WL), Mean Frequency (MNF) and Median Frequency (MDF). Among them, we found that the combination of the three features of RMS, WL and MNF(o) showed better results. These results are consistent with previous research results, root mean square (RMS) and wavelength (WL) can get good classification results and require less computational cost [31]. Compared with time-domain features such as RMS and WL, frequency-domain features require more computational costs. In [32], the author reported that the frequency-domain features are better than the time-domain features when predicting the angle and flexion of the thumb. Therefore, this study uses a combination of two time domain features RMS, WL and a frequency-domain features Mean Frequency (MNF) for classification.

The WL, RMS, MNF are defined as follows:

$$WL = \sum_{n=1}^{n-1} |x_{n+1} - x_n| \tag{2}$$

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{3}$$

$$MNF = \sum^{M} F\,P \Big/ \sum^{M} P \tag{4}$$

here, x_n denotes the n^{th} of EMG signal, N is the length of signals. F is the frequency, P is the Power spectrum.

Table 3: The list of objects used in the Experiment

Number	Name	Weight (g)	Size(cm)
1	TV Remote	75	length 13 width 5 thickness 3
2	Pill	1	diameter 1
3	Telephone	137	length 17 width 5 thickness 2
4	Bottle	40	length 13 width 5 thickness 3
5	Glasses	38	length 15 width 3 thickness 3
6	Spoon	5	length 15 width 3 thickness 2
7	Phone	190	length 16 width 7 thickness 1
8	Tooth Paste	238	length 13 width 5 thickness 3
9	Envelope	14	length 23 width 12
10	Drink (500ml)	534	length 21 width 7 thickness 7
11	Bowl	57	length 11 width 11 thickness 5
12	Key	9	length 7 width 3 thickness 0.5
13	Pen	15	length 14 width 1 thickness 1
14	Paper box	73	length 18 width 11 thickness 8
15	Bill	1	length 15 width 7
16	Straw	2	length 10 width 0.5
17	Coin	5	diameter 2.5
18	Bag	897	length 40 width 25 thickness 15
19	DVD	16	diameter 11
20	Small Ball	12	diameter 3

3. Evaluation Experiment

3.1. Grasping experiment

In order to evaluate the applicability of the proposed prosthesis to daily life, we used the items in the list of items that amyotrophic lateral sclerosis (ALS) patients are difficult to pick up in daily life for experiments [33]. We tested the feasibility of the prosthetic hand grasped and pick up easily some items as shown in Table 3. The gripping and dexterity must be needed the equivalent strength of strings driven like an actual hand. We have selected 20 items in the item list above. Among them, 19 items can be easily found in the laboratory. In addition, a school bag of about 1 kg was selected as the object to observe the performance of the artificial hand against heavy objects in Appendices. In grasping an item, if the prosthesis can grip for the object over 5 seconds, it will be recorded as a success, otherwise, it will be recorded as a failure.

3.2. Classification experiment

In order to classify four gestures, we used MATLAB's neural network toolbox to build a three-layer artificial neural network (ANN) as a classifier. This ANN consists of twelve (four channels x three features) nodes of input layers, one hidden layer composed of twelve nodes, and four nodes output layers. The activation functions of the middle layer and the output layer use hyperbolic tangent functions. The training method uses a learning algorithm called Levenberg-Marquardt backpropagation (LMBP). LMBP is a general non-linear least-squares optimization method, its major advantage lies in the speed of convergence. Before starting the Classification Experiment, we will train their own ANN for each subject in advance. The dataset used for training ANN contains 10 sets of data for each motion, and each set of data contains 5 seconds of EMG signal. In order to prevent over fitting, the training of ANN will stop if Mean Square Error (MSE) $\leqq$ 0.001. Before performing the Classification experiment, use 10-fold cross-validation on the collected dataset to test the accuracy of the classifier as shown as Figure 8. For each subject's dataset, 80 % of the data is used to train the classifier, and the remaining 20 % are used to validate the performance of the classifier. Figure 9 shows the average accuracy of the four motions of 4 subjects after using 10-fold cross-validation. The highest accuracy rate is 94.7% (Subject4), the lowest accuracy rate is 91.4% (Subject2).

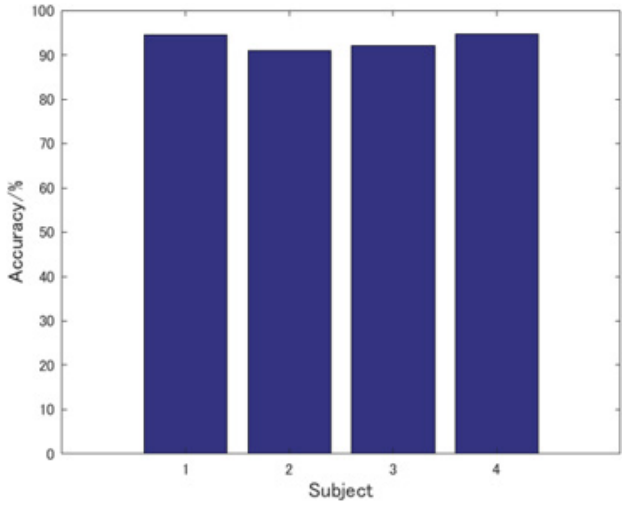


Figure 9: 10-fold cross validation results

4. Results

4.1. Result of Grasping experiment

In the experiment, our prosthetic hand successfully grasped selected 20 objects (Figure 10). Different from the traditional ON/OFF clip-shaped prosthetic hand or the prosthetic hand that grasps in a fixed posture, the proposed anthropomorphic design allows our prosthetic hand to grasp objects in the suitable grasping posture.

Figure 10: Result of grasping experiment

4.2. Result of Classification experiment

It showed the results of the movement Classification experiment in Table 4. The highest average accuracy rate was 93.7%. The lowest average accuracy rate was 87.5%. All subjects achieved 100% accuracy when Classifying "Power grasp" Motion. In addition, the classification of "Precision grasp" and "Lateral grasp" is slightly lower than the accuracy rate of other motions, with a minimum of 80%. During the experiment, the researchers observed that the subject's 'Relax' movements were classified as 'Lateral grasp' movements multiple times. This might because the

Table 4: Classification accuracy

Subject / Motion	1	2	3	4
Power grasp	20/20 100%	20/20 100%	20/20 100%	20/20 100%
Precision grasp	18/20 90%	18/20 90%	17/20 85%	16/20 80%
Lateral grasp	20/20 100%	16/20 80%	18/20 90%	17/20 85%
Relax	17/20 85%	19/20 95%	18/20 90%	17/20 85%
AVG.	93.7%	91.2%	91.2%	87.5%

EMG signal generated by thumb adduction is difficult to collect from superficial muscles.

5. Discussion

In the Grasping experiment, we tested the gripping performance of the proposed prosthesis against objects commonly found in daily environments. When grasping objects such as cylindrical objects or large objects, we usually use the Power grasp posture. The 500ml PET bottle and the 18x11x8cm paper bag are very large for the hand, so it is necessary to grasp them with your fingertips. In addition, the artificial hand can lift a bag of about 1 kg and hold the posture for over 5 seconds. From the above results, it can be found that the anatomically reproduced finger structure sufficiently contributes to the transmission of force and transmits the torque of the motor to the fingers and fingertips. Although a strong torque moment applies to the fingertips, the shape of the finger joints is not deformed, and the proposed artificial hand has sufficient force and rigidity to grasp heavy objects. By using precision grasp and Lateral grasp, the prosthetic hand can pick up small objects such as balls and coins with fingertips during the experiment.

These results shown that our proposed design shows highly coordinated motions of the thumb and index finger, suggested that the proposed thumb structure has the same kinetic features as a human thumb. It is also one of our design goals to use bones as a model to give the artificial hand a natural appearance. As shown in Figure 11, our prosthetic hand also holds Natural-Appearance when wearing a silicon skin glove. It should be noted that the silicon skin glove worn is only a general commercial product.

For healthy subjects, the ANN classification system can achieve high-precision classification of 4 motions. But before the system is used in amputees, it is still necessary to continue to improve performance. From the experimental results, the accuracy of the two motions of "precise grasp" and "side grasp" is lower than other motions. The reason may be that "precise grip" and "side grip" include thumb movement, so the feature sets of these two motions are similar. In our experiment, the EMG signal is measured by two EMG sensors in the flexor group and two EMG sensors in the extensor group. However, it may be difficult to classify detailed finger movements from the extracted EMG signals. Because there are many muscles in the forearm (Flexor carpi radialis, flexor carpi ulnar is, Flexor digitorum superficialis muscle, Deep digital flexor muscle, flexor digitorum longus) located deep in the measurement position. This will affect the accuracy and precision of classification.

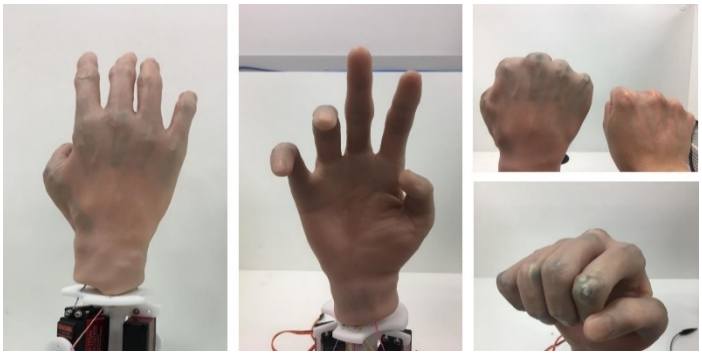


Figure 11: Appearance of prosthetic hand wearing silicon skin glove

6. Conclusion

In this study, we developed a biomimetic prosthetic hand based on human anatomy. By reproducing critical soft tissues such as tendons and ligaments in the human hand, the developed artificial hand has the same dexterity as a human finger. For this reason, our prosthetic hand can stably grasp all the object types proposed in the experiment. In order to control the prosthetic hand by EMG signal, an ANN has been designed for classifying four motions. Four subjects tested the classification performance of ANN. The classification accuracy average rate was 91%.

The limitations of this study are as follows. The overall system of the proposed bionic prosthesis is about 560 grams. A survey of prosthetic users pointed out that if the weight of the prosthetic hand exceeds 400 grams, and it is considered being heavy [34]. Currently, our biomimetic prosthesis does not include the wrist part. In order to use artificial hands in daily life, the ability of the wrist part also plays a vital role. Currently, the proposed EMG control method only tests the performance of healthy subjects, and needs to be tested on amputees to investigate whether the performance will change. In the next phase of this research, we will make improvements to address these issues.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 19K11428.

References

[1] Z. He, R.R. Yurievich, S. Shimizu, M. Fukuda, Y. Kang, D. Shin, "A Design of Anthropomorphic Hand based on Human Finger Anatomy," in 2020 International Symposium on Community-centric Systems (CcS), IEEE: 1–5, 2020, doi:10.1109/CcS49175.2020.9231423.

[2] K. Ziegler-Graham, E.J. MacKenzie, P.L. Ephraim, T.G. Travison, R. Brookmeyer, "Estimating the Prevalence of Limb Loss in the United States: 2005 to 2050," Archives of Physical Medicine and Rehabilitation, **89**(3), 422–429, 2008, doi:10.1016/J.APMR.2007.11.005.

[3] F. Cordella, A.L. Ciancio, R. Sacchetti, A. Davalli, A.G. Cutti, E. Guglielmelli, L. Zollo, "Literature Review on Needs of Upper Limb Prosthesis Users," Frontiers in Neuroscience, **0**(MAY), 209, 2016, doi:10.3389/FNINS.2016.00209.

[4] E.A. Biddiss, T.T. Chau, "Upper limb prosthesis use and abandonment: A survey of the last 25 years," doi:10.1080/03093640600994581.

[5] I. Vujaklija, D. Farina, O.C. Aszmann, "New developments in prosthetic arm systems," Orthopedic Research and Reviews, **8**, 31, 2016, doi:10.2147/ORR.S71468.

[6] G. Langevin. InMoov – open-source 3D printed life-size robot, http://inmoov.fr/, Aug. 2021.

[7] Exiii Inc. HACKberry |3D-printable open-source bionic arm, http://exiii-hackberry.com/, Aug. 2021.

[8] J. Zuniga, D. Katsavelis, J. Peck, J. Stollberg, M. Petrykowski, A. Carson, C. Fernandez, "Cyborg beast: a low-cost 3d-printed prosthetic hand for children with upper-limb differences," BMC Research Notes 2015 8:1, **8**(1), 1–9, 2015, doi:10.1186/S13104-015-0971-9.

[9] P. Weiner, J. Starke, F. Hundhausen, J. Beil, T. Asfour, "The KIT Prosthetic Hand: Design and Control," IEEE International Conference on Intelligent Robots and Systems, 3328–3334, 2018, doi:10.1109/IROS.2018.8593851.

[10] P. Weiner, C. Neef, Y. Shibata, Y. Nakamura, T. Asfour, "An Embedded, Multi-Modal Sensor System for Scalable Robotic and Prosthetic Hand Fingers," Sensors 2020, Vol. 20, Page 101, **20**(1), 101, 2019, doi:10.3390/S20010101.

[11] S. Lee, S. Noh, Y.K. Lee, J.H. Park, "Development of bio-mimetic robot hand using parallel mechanisms," 2009 IEEE International Conference on Robotics and Biomimetics, ROBIO 2009, 550–555, 2009, doi:10.1109/ROBIO.2009.5420706.

[12] A. Mohammadi, J. Lavranos, H. Zhou, R. Mutlu, G. Alici, Y. Tan, P. Choong, D. Oetomo, "A practical 3D-printed soft robotic prosthetic hand with multi-articulating capabilities," PLOS ONE, **15**(5), e0232766, 2020, doi:10.1371/JOURNAL.PONE.0232766.

[13] A.D. Deshpande, Z. Xu, M.J.V. Weghe, B.H. Brown, J. Ko, L.Y. Chang, D.D. Wilkinson, S.M. Bidic, Y. Matsuoka, "Mechanisms of the anatomically correct testbed hand," IEEE/ASME Transactions on Mechatronics, **18**(1), 238–250, 2013, doi:10.1109/TMECH.2011.2166801.

[14] Z. Xu, E. Todorov, "Design of a highly biomimetic anthropomorphic robotic hand towards artificial limb regeneration," Proceedings - IEEE International Conference on Robotics and Automation, **2016-June**, 3485–3492, 2016, doi:10.1109/ICRA.2016.7487528.

[15] M.B.I. Reaz, M.S. Hussain, F. Mohd-Yasin, M.B.I. Raez, "Techniques of EMG signal analysis: detection, processing, classification and applications," Biol. Proced. Online, **8**(1), 11–35, 2006, doi:10.1251/bpo115.

[16] O. Faust, Y. Hagiwara, T.J. Hong, O.S. Lih, U.R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," Computer Methods and Programs in Biomedicine, **161**, 1–13, 2018, doi:10.1016/J.CMPB.2018.04.005.

[17] A. Kalantari, A. Kamsin, S. Shamshirband, A. Gani, H. Alinejad-Rokny, A.T. Chronopoulos, "Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions," Neurocomputing, **276**, 2–22, 2018, doi:10.1016/J.NEUCOM.2017.01.126.

[18] U. Côtéallard, F. Nougarou, C.L. Fall, P. Gigu'ere, C. Gosselin, F. Laviolette, B. Gosselin, "A Convolutional Neural Network for robotic arm guidance using sEMG based frequency-features," IEEE International Conference on Intelligent Robots and Systems, **2016-November**, 2464–2470, 2016, doi:10.1109/IROS.2016.7759384.

[19] M. Tavakoli, C. Benussi, P. Alhais Lopes, L.B. Osorio, A.T. de Almeida, "Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier," Biomedical Signal Processing and Control, **46**, 121–130, 2018, doi:10.1016/J.BSPC.2018.07.010.

[20] I.M. Bullock, J.Z. Zheng, S. De La Rosa, C. Guertler, A.M. Dollar, "Grasp frequency and usage in daily household and machine shop tasks," IEEE Transactions on Haptics, **6**(3), 296–308, 2013, doi:10.1109/TOH.2013.6.

[21] C.C. Scuola, S. Sant'anna, M.C. Scuola, M. Chiara, C. Scuola, C. Cipriani, M. Controzzi, M. Chiara Carrozza, "Objectives, criteria and methods for the design of the SmartHand transradial prosthesis Sensory feedback and closed loop control of artificial limbs View project Artificial arms components View project Objectives, criteria and methods for the design of the SmartHand transradial prosthesis," Robotica, **28**, 919–927, 2010, doi:10.1017/S0263574709990750.

[22] C.L. Taylor, R.J. Schwarz, "The Anatomy and Mechanics of the Human Hand," Artificial Limbs VOL. 2 MAY 1955 NO. 2 CONTENTS, 2, 1955.

[23] Human Hand Bones - Thumb by siderits - Thingiverse, https://www.thingiverse.com/thing:15342, Aug. 2021.

[24] J.C. Becker, N. V. Thakor, "A Study of the Range of Motion of Human Fingers with Application to Anthropomorphic Designs," IEEE Transactions on Biomedical Engineering, **35**(2), 110–117, 1988, doi:10.1109/10.1348.

[25] I. Komatsu, J.D. Lubahn, "Anatomy and Biomechanics of the Thumb Carpometacarpal Joint," Operative Techniques in Orthopaedics, **28**(1), 1–5, 2018, doi:10.1053/J.OTO.2017.12.002.

[26] EDUCATION EXHIBIT Extensor Mechanism of the Fingers: MR Imaging-Anatomic Correlation 1, Aug. 2021, doi:10.1148/rg.233025079.

[27] A. Gailey, P. Artemiadis, M. Santello, "Proof of Concept of an Online EMG-Based Decoding of Hand Postures and Individual Digit Forces for Prosthetic Hand Control," Frontiers in Neurology, **0**(FEB), 7, 2017, doi:10.3389/FNEUR.2017.00007.

[28] MathWorks Instrument Control Toolbox Team (2021). BITalino Toolbox, https://www.mathworks.com/matlabcentral/fileexchange/53983-bitalino-toolbox, Aug. 2021.

[29] Y. Koike, M. Kawato, "Estimation of dynamic joint torques and trajectory formation from surface electromyography signals using a neural network model," Biological Cybernetics 1995 73:4, **73**(4), 291–300, 1995, doi:10.1007/BF00199465.

[30] A. Phinyomark, C. Limsakul, P. Phukpattaranont, "A Novel Feature Extraction for Robust EMG Pattern Recognition," Journal of Medical Engineering and Technology, **40**(4), 149–154, 2009.

[31] W. Chen, M. Liu, S. Yue, W. Chen, C.-H. Xiong, W.-R. Chen, B.-Y. Sun, M.-J. Liu, S.-G. Yue, W.-B. Chen, "Design and Implementation of an Anthropomorphic Hand for Replicating Human Grasping Functions," IEEE TRANSACTIONS ON ROBOTICS, **32**(3), 2016, doi:10.1109/TRO.2016.2558193.

[32] A.R. Siddiqi, S.N. Sidek, "Estimation of continuous thumb angle and force using electromyogram classification," Aug. 2021, doi:10.1177/1729881416658179.

[33] S.C. Young, T. Deyle, T. Chen, J.D. Glass, C.C. Kemp, "A list of household objects for robotic retrieval prioritized by people with ALS," 2009 IEEE International Conference on Rehabilitation Robotics, ICORR 2009, 510–517, 2009, doi:10.1109/ICORR.2009.5209484.

[34] J.T. Belter, A.M. Dollar, "Performance characteristics of anthropomorphic prosthetic hands," IEEE International Conference on Rehabilitation Robotics, 2011, doi:10.1109/ICORR.2011.5975476.