Princess Sumaya University for Technology
جامعة الأميرة سميّة للتكنولوجيا

صندوق دعم البحث العلمي والابتكار
Scientific Research and Innovation Support Fund

Jordanian Journal of Computers and Information Technology

June 2024 | VOLUME 10 | NUMBER 02 | ISSN 2415 - 1076 (Online) ISSN 2413 - 9351 (Print)

JJCIT

www.jjcit.org

jjcit@psut.edu.jo

An International Peer-Reviewed Scientific Journal Financed by the Scientific Research and Innovation Support Fund

Jordanian Journal of Computers and Information Technology (JJCIT)

The Jordanian Journal of Computers and Information Technology (JJCIT) is an international journal that publishes original, high-quality and cutting edge research papers on all aspects and technologies in ICT fields.

JJCIT is hosted and published by Princess Sumaya University for Technology (PSUT) and supported by the Scientific Research Support Fund in Jordan. Researchers have the right to read, print, distribute, search, download, copy or link to the full text of articles. JJCIT permits reproduction as long as the source is acknowledged.

AIMS AND SCOPE

The JJCIT aims to publish the most current developments in the form of original articles as well as review articles in all areas of Telecommunications, Computer Engineering and Information Technology and make them available to researchers worldwide. The JJCIT focuses on topics including, but not limited to: Computer Engineering & Communication Networks, Computer Science & Information Systems and Information Technology and Applications.

INDEXING

JJCIT is indexed in:

EDITORIAL BOARD SUPPORT TEAM

JJCIT ADDRESS

WEBSITE: www.jjcit.org
EMAIL: jjcit@psut.edu.jo
ADDRESS: Princess Sumaya University for Technology, Khalil Saket Street, Al-Jubaiha
B.O. BOX: 1438 Amman 11941 Jordan
TELEPHONE: +962-6-5359949
FAX: +962-6-7295534

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 02, June 2024.

A FUSION OF A DISCRETE WAVELET TRANSFORM-BASED AND TIME-DOMAIN FEATURE EXTRACTION FOR MOTOR IMAGERY CLASSIFICATION

Fouziah Md Yassin[1], Norita Md Norwawi[2], Nor Azila Noh[3], Afishah Alias[4] and Sofina Tamam[5]

(Received: 20-Nov.-2023, Revised: 1-Feb.-2024, Accepted: 16-Feb.-2024)

ABSTRACT

A motor imagery (MI)-based brain-computer interface (BCI) has performed successfully as a control mechanism with multiple electroencephalogram (EEG) channels. For practicality, fewer EEG channels are preferable. This paper investigates a single-channel EEG signal for MI. However, there are insufficient features that can be extracted due to a single-channel EEG signal being used in one region of the brain. An effective feature extraction technique plays a critical role in overcoming this limitation. Therefore, this study proposes a fusion of discrete wavelet transform (DWT)-based and time-domain feature extraction to provide more relevant information for classification. The highest accuracy obtained on the BCI Competition III (IVa) dataset is 87.5% with logistic regression (LR) while the OpenBMI dataset attained the highest accuracy of 93% with support vector machine (SVM) as the classifier. Addressing the potential of enhancing the performance of a single EEG channel located on the forehead, the achieved result is relatively promising.

KEYWORDS

Motor imagery, Feature extraction, Electroencephalogram (EEG), Discrete wavelet transform, Brain-computer interface.

1. INTRODUCTION

There are two techniques for measuring brain activity: invasive measurement and non-invasive measurement. The non-invasive design was ranked as a high-priority design. It is risk-free and does not require surgery as an invasive necessity, even though the invasive technique is more accurate [1]. An electroencephalogram (EEG) is widely used for non-invasive measurement that records the brain's electrical fields through metal electrodes placed on the scalp with the standard international 10–20 electrode site placement [2]. Besides that, it is relatively inexpensive, has a good temporal resolution, enabling it to accurately capture fluctuations in brain activity throughout time and requires little setup [3]. Additionally, it is highly portable, making it suitable for usage in diverse locations, such as hospital environments, research laboratories and even mobile applications. The EEG is a useful diagnostic tool that is particularly effective in identifying and monitoring neurological illnesses, like Alzheimer's and epilepsy, involving analysis of the EEG recordings to detect abnormal brain activity linked with seizures [4][5][6]. Besides that, it is being used in various non-clinical settings, such as education, emotion detection and control mechanisms, to explore new potentials and applications [7]-[8]. The EEG is utilized in education to investigate and improve cognitive processes, providing valuable information about attention, focus and learning patterns. Furthermore, the EEG plays a crucial role in the advancement of brain-computer interfaces (BCIs) which function as control mechanisms.

BCI links the human brain's electrical activity to an external device, such as a wheelchair or computer system. Neuronal electrical signals in the human brain are detected, interpreted and converted into machine language that corresponds to the user's desires [1]. This technology has significant potential to offer alternative communication channels for those with physical limitations. In BCI, users' comfort is

1. F. M. Yassin is with Faculty of Sci. and Tech., Uni. Sains Islam Malaysia and with Faculty of Science and Natural Resources, Universiti Malaysia Sabah, Sabah, Malaysia. Emails: fouziahy@raudah.usim.edu.my and fouziahy@ums.edu.my
2. N. M. Norwawi is with Cyber Security and System Research Unit, Faculty of Science and Technology, Universiti Sains Islam Malaysia, Negeri Sembilan, Malaysia. Email: norita@usim.edu.my
3. N. Noh is with the Brain and Behaviour Research Group, and Faculty of Medicine and Health Sciences, Universiti Sains Islam Malaysia, Negeri Sembilan, Malaysia. Email: azila@usim.edu.my
4. A. Alias is with the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn, Johor, Malaysia. Email: afishah@uthm.edu.my
5. S. Tamam is with the Brain and Behaviour Research Group and Faculty of Science and Technology, Universiti Sains Islam Malaysia, Negeri Sembilan, Malaysia. Email: sofinatamam@usim.edu.my

not only sitting on a comfortable chair, but also allowing them to have mobility [9]. The portability and adaptability of the EEG make it well-suited for investigating diverse elements of brain activity and connectivity in BCI, including motor imagery (MI). BCI can be classified into active, passive and reactive. MI is an active BCI, whereby the user intentionally generates specific brain signals to interact with an external device by executing imaginary movements without physically performing them [10].

MI-BCI is a beneficial technique that has been applied successfully for rehabilitation, gaming and device control. The most typical movements or commands used in MI tasks as control mechanisms for wheelchairs and cursors are left, right, forward, up and down [11]. MI-BCI activity is generated in two different rhythms (8–13 Hz and 13–30 Hz) [12]-[13]. When developing a BCI system, it is important to consider the number of sensors that can accurately record and resolve the signal's reliability [14]. It is also associated with user comfort. Multi-channel EEG data could achieve high classification accuracy (CA). However, it has increased the complexity and setup time of the experimental procedure [5]. For daily-time usage, a smaller number of EEG channels is more practical, but there are still limitations with reliability and low accuracy [15]. Gaur et al. [16] employed two public datasets to investigate the performance of reducing the number of channels. They discovered that by reducing the number of channels from 118 to 13, they were able to reach an accuracy of 80.56% in the BCI implementation. The person, as well as the well-designed experimental settings and classification algorithm, have a significant impact on the number of channels required for a high accuracy rate [17]. Thus, if high-accuracy classification can be obtained with only one EEG channel, it will be easier and more comfortable to use BCIs on a regular basis [18]. Extracting meaningful and relevant features from a single-channel EEG signal could be more challenging than from a multi-channel system. Due to limited dimensionality and information content across various brain regions, this may result in low accuracy and interpretability. Therefore, the feature extraction method is very important for getting sufficient CA for the EEG signals that come from one channel.

Therefore, this study proposes the fusion of a discrete wavelet transform (DWT)-based and time-domain feature extraction to improve the interpretation of movement tasks in single-channel EEG signals. The DWT decomposes the signal into different frequency components, enabling the analysis of a wide range of temporal and frequency characteristics within the same signal. Selected specific frequency components are used for relevant features extraction. Fusion of features integrates different sets of features captured from a single-channel EEG signal, providing a more comprehensive representation of the signal with useful data. The study investigates the impact of this feature-extraction approach on the performance of specific EEG channels.

The EEG signal from the following channels: Fp1, Fp2 and AF3 presented sufficient CA in the previous studies [18][19][20]. Besides that, it was reported that AF3 and AF4 are among the most informative channels found in two benchmark datasets [21]. According to previous research, it was found that there was significant activation in the prefrontal region in implementing MI tasks, making it plays an important role in MI tasks, including those related to gait and lower limb movement [22]-[23]. The region is involved in various cognitive and executive functions. Based on the findings of [24], it was suggested that MI depends greatly on executive resources, because tasks involving executive processes, such as calculations, have a significant impact on MI, but have a lesser effect on overt actions. Therefore, the study explores the potential of four channels located at the frontal right and left hemispheres of the forehead (Fp1, Fp2, AF3 and AF4) for implementing the MI. The position of the channels was considered to have the potential to enhance the practicality of the MI-BCI system [18]. Furthermore, the most employed channels for MI involving the hands and feet, specifically C3, C4 and Cz, are examined as well for the purpose of comparison. Foot movements should be observed around the Cz channel [25].

To achieve the aim of the study, three feature-extraction approaches are applied to two benchmark datasets, resulting in three feature sets: time-frequency features (DWT-based), time-domain features and fusion of DWT-based and time-domain features. Three classifiers are utilized to classify the selected features. They are Support Vector Machine (SVM), Logistic Regression (LR) and Naïve Bayes (NB).

2. RELATED WORKS

For multichannel feature extraction, Common Spatial Pattern (CSP) is commonly used [26]. Among all feature extraction techniques studied by Selim et al. [27], CSP produced excellent results when measuring accuracy and execution time. It is frequency-domain feature extraction that requires more

channels for MI signal processing. Therefore, the time-frequency domain decomposition method is introduced for single-channel EEG signal execution. Short-time Fourier transforms (STFTs) are one of the time-frequency domain methods that were used in previous studies [21], [28][29][30]. However, STFT is not suitable for non-stationary applications because of the fixed window size [31]-[32]. Tiwari [21] introduced a novel Logistic S-shaped Binary Jaya Optimization Algorithm (LS-BJOA) for MI classification in BCI, while the Regularized Common Spatial Pattern (RCSP) was applied for feature extraction. The study validated its method on three public EEG datasets, achieving the CA of 83.59%, 82.09% and 89.02% on these datasets, respectively, with a reduced number of channels compared to baseline methods. In the single EEG channel research conducted by Chen et al. [30], it was reported that the FitzHugh-Nagumo (FHN)-PSD system achieved an average CA of 67.06% ± 8.73%, specifically on the C4 channel. The FHN-PSD system exhibited a 2.29% improvement over the FHN-STFTCSP approach, indicating its greater effectiveness in classifying EEG signals.

Through most of the datasets utilized for the comparison study, Wavelet Transform (WT) showed more robustness than CSP and power spectral density (PSD), as reported by Moumgiakmas and Papakostas [26]. This could be seen from the previous works implementing decomposition method in their studies. Three different signal-decomposition algorithms for MI-BCI systems were tested and compared. It was found that wavelet packet decomposition (WPD) was the most accurate method (92.8%). It was followed by the discrete wavelet transform (DWT) and empirical mode decomposition (EMD) [33]. The research indicates that the efficacy of their methodology can be improved by carefully choosing a suitable decomposition method and features, even with the small number of EEG channels (C3, C4 and Cz). WPD was used for decomposing the EEG signal to apply a hybrid feature set that combined it with the time-domain feature set [34]. However, results showed that the smallest number of channels (C3, Cz and C4) obtained the lowest CA, while using eighteen channels resulted in the highest mean CA of 91.1% for the BCI Competition III dataset IVa. The hybrid feature selection played an important role in the research to select the relevant features to be classified by SVM. In the studies of Ji et al. [35], two stages of the decomposition approach were applied. The EEG signal was decomposed using DWT to generate a narrow-band signal. The second decomposition method, EMD, was used to obtain a more concentrated signal for frequency-band signals. In order to improve the classification, an approximate entropy was determined. Rather than using DWT alone, this approach provides an additional method for extracting movement imagining signal features from two EEG channels (C3 and C4). The accuracy was 95.1% using SVM. However, the statistics of the approximate entropy were inconsistent. The tunable Q wavelet transform (TQWT) is a discrete-time WT that has been parametrized and has a tunable quality factor. The quality factor (Q), the redundancy rate (r) and the level of decomposition (L) are required as the basis functions for the decomposition. It was employed in the research conducted by Khare et al. [20]. The highest accuracy of 99.78% was achieved with the least squares support vector machine (LS-SVM) model. The selection of the most informative channels from a total of 118 was a critical aspect of the study, aimed at reducing computational load. However, despite the optimization, the multichannel setup is still required, which is too time-consuming to feasibly be deployed on the scalp every day.

In different applications, DWT was employed to extract the features, since it is an effective approach in terms of accuracy compared to other methods for categorizing epileptic cases based on the EEG. The combination of DWT with differential evaluation (DE) enhanced the classification result [36]. A promising result was achieved in the seizure identification by the concatenation of a feature matrix derived from DWT and EMD across multi and single-channel EEG recordings [37]. Instead of employing only DWT, the concatenation increased the accuracy of the single-channel EEG signals from 85% to 100%. Some EEG channels in the MI-BCI system worked better when DWT was combined with other feature extraction methods, as in the earlier research. This approach also enhanced the performance of single-channel EEG signals across different applications, which may benefit MI application.

3. METHODOLOGY

The methodology is divided into four phases: dataset acquisition, data preprocessing, feature extraction and selection and classification. The block diagram of the proposed approach is shown in Figure 1.

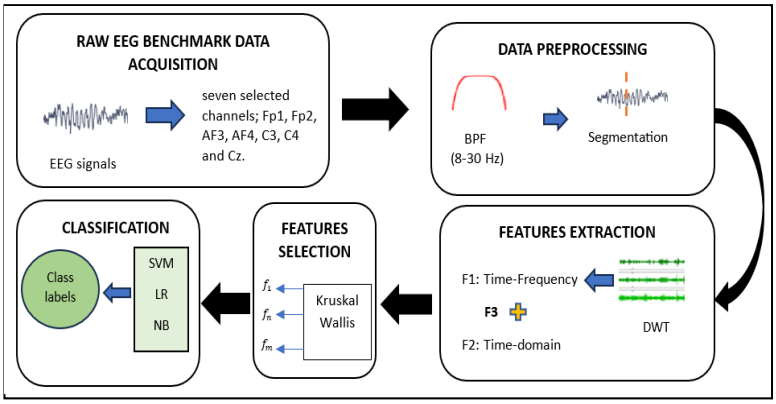


Figure 1. A block diagram of the proposed approach.

In the feature extraction phase, let the features in the first experiment be F1= $\{f_1,\ldots\ldots\ldots, f_n\}$, while the features in the second experiment is F2= $\{f_1,\ldots\ldots\ldots, f_m\}$. The features for the third experiment are therefore formally represented by:

$$F3= \{f_1,\ldots\ldots\ldots, f_n, f_{n+1},\ldots\ldots\ldots, f_m\} \quad (1)$$

where f_n represents different features in F1, f_m represents different features in F2 and F3 is the fusion features of F1 and F2. All phases are described in detail in the following sub-sections.

3.1 Motor Imagery (MI) Dataset Acquisition

Dataset 1 is the BCI Competition III (IVa) dataset [38]. It is the cued MI data with two tasks that was recorded using the extended 10-20 international system's 118 EEG channels. It was from five healthy participants (*aa*, *al*, *av*, *aw* and *ay*) who comfortably sat during the experiment. The MI tasks were completed by conducting 140 trials of MI tasks with the right-hand (*rh*) movement and 140 trials with the right-foot (*rf*) movement. The total number of trials for each participant was 280. They were each provided with a 3.5-second visual stimulus. The dataset was collected at a sampling rate of 1000 Hz. However, they were down-sampled to 100 Hz for analysis purposes.

Dataset 2 is the OpenBMI dataset which comprises data collected from 54 healthy individuals [39]. The experiment recorded binary MI tasks of the right (*rh*) and left (*lh*) hands. Each trial started with a preparation phase, followed by MI tasks for a duration of 4 seconds once a visual cue was displayed. The experiment had four different stages with 100 trials each, equally split between right- and left-hand imagery. For this study, 19 participants (S1, S2, S3, S9, S18, S19, S21, S22, S28, S29, S30, S32, S33, S36, S37, S43, S44, S45 and S52) with proven MI-BCI literacy were selected. The data was analyzed using offline EEG data from the first session [39]. EEG signals were recorded at 1000 Hz over 62 channels, but down-sampled to 100 Hz for analysis.

C3, Cz and C4 channels are extensively employed for analysis in MI studies, regardless of whether they are multi-EEG channel or single-EEG channel analysis. The channels' location is over the primary motor cortex areas of the brain for controlling voluntary movements. C3 and C4 are located over the left and right hemispheres, respectively, and are known to capture important MI properties [30], [40]. The Fp2 channel, which is located on the forehead, was reported to have equivalent high CA to C4 by Ge et al. [18]. Furthermore, in the previous study, the AF3 and AF4 channels, located near the forehead, were identified as the channels providing the most relevant information [19], [21]. The Laplacian score for channel selection demonstrated that Fp1 was also identified as the most informative channel [20]. Thus, among the 118 channels in Dataset 1 and the 62 channels in Dataset 2, only the EEG signals from the following channels are selected for subsequent processing and analysis: C3, Cz, C4, Fp1, Fp2, AF3 and AF4.

3.2 Data Preprocessing

The preprocessing phase is essential for obtaining reliable data that is ready for meaningful interpretation about brain activity. It filters out any artifacts and noise in the signal, allowing useful features to be extracted from the raw data. Two EEG frequency sub-bands, α (8-13 Hz) and β (13-30 Hz), needed to

be isolated from other ranges of the raw EEG signal. The frequency bands are associated with the motor activation, preparation and planning of imagery movement [41]. Therefore, the 5th-order Butterworth filter with a pass band of 8- 30 Hz was employed in the first step, as implemented in previous works [19], [39]. After that, the signals were segmented to extract the dataset's epochs with a window length of 3 seconds (0.5 s to 3.5 s) for Dataset 1 and 2.5 seconds (1s to 3.5s) for Dataset 2. They were prepared to extract relevant features that can distinguish between two MI tasks.

3.3 Feature Extraction

In the feature-extraction phase, three experiments were conducted. F1 involved the DWT decomposition technique to process single-channel EEG. The technique enables the separation of a single-component signal into multiple sub-signals. Each component corresponds to a different part in a specific frequency band of the signal that is essential for feature extraction. It is possible to efficiently extract relevant information or features from a complex signal that are beneficial for task classification. Down samplers and consecutive high- pass (g_n) and low-pass (h_n) filtering of the time series are used in the DWT decomposition of the input signal (x_n), as shown in Figure 2. It split the signal into high-frequency and low-frequency content in the form of detail and an approximate coefficient [42]. For a further level of decomposition, only the approximations are passed again through the low-pass and high-pass filters. In this study, signals were decomposed using three levels of DWT with the Daubechies 4 wavelet (db4) [35], resulting in three detailed components (D1, D2 and D3) and one approximation component (A3), as shown in Figure 2. The frequency range for each selected band is shown in Table 1.

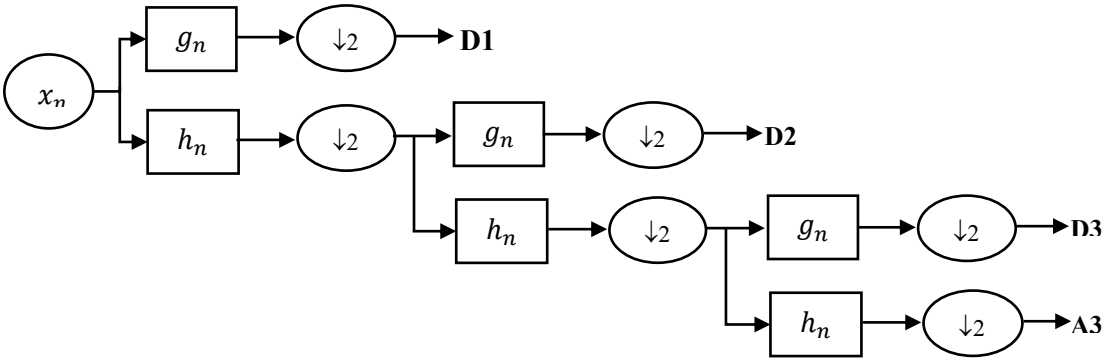


Figure 2. DWT decomposition of the input signal.

Table 1. Decomposition levels, coefficient vectors and their frequency ranges.

Level	Coefficient Vector	Frequency Range (Hz)
1	D1	25-50
2	D2	12.5-25
3	D3	6.25-12.5

As A3's frequency range (0- 6.25 Hz) is outside the required frequency band, it was excluded for feature extraction. Moreover, the frequency of a signal that is less than 5 Hz may have artifacts [35]. D1 was included in the analysis, because it might still offer a useful filtered signal that is relevant to the specific frequency of interest, which is 25 to 30 Hz. As a result, five features were extracted from each detail component, yielding a total of fifteen (15) features in F1. The feature in each sub-band is represented by $F1_{D1} = \{\mu_1, \sigma_1, skewness_1, kurtosis_1, P_{av\ 1}\}$ for D1.
D2 is represented by $F1_{D2} = \{\mu_2, \sigma_2, skewness_2, kurtosis_2, P_{av\ 2}\}$, while D3 is represented by $F1_{D3} = \{\mu_3, \sigma_3, skewness_3, kurtosis_3, P_{av\ 3}\}$. The equations of the features in F1 are as follows:

The absolute mean (μ)is defined as [33]:

$$\mu = \frac{1}{N}\sum_{n=1}^{N} |y_n| \tag{2}$$

The standard deviation (σ) is defined as [37]:

$$\sigma = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \bar{y})^2} \tag{3}$$

The skewness is defined as [37]:

$$skewness = \frac{1}{N}\frac{\sum_{n=1}^{N}(y_n - \bar{y})^3}{\sigma^3} \quad (4)$$

The kurtosis is defined as [37]:

$$kurtosis = \frac{1}{N}\frac{\sum_{n=1}^{N}(y_n - \bar{y})^4}{\sigma^4} \quad (5)$$

The average power (P_{av}) is defined as [33]:

$$P_{av} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} {y_n}^2} \quad (6)$$

where N is the number of samples, y_n is the signal in each sub-band, $\bar{y}$ is the mean of the signal in each sub-band and n is an integer that belongs to 1 to N.

Eight time-domain features were directly extracted from the processed signal in F2. These include the mean absolute value, Root Mean Square (RMS), Hjorth parameters (activity, mobility and complexity), waveform duration, skewness and kurtosis represented by F2 $= \{\mu, RMS, Activity, Mobility, Complexity, WL, skewness, kurtosis\}$. The equations for the parameters are as follows:

The mean absolute (μ) value is written as [33] :

$$\mu = \frac{1}{T}\sum_{t=1}^{T} |x_t| \quad (7)$$

RMS is the square root of the average of the signal's squared value in the time domain. The RMS is written as [43]:

$$RMS = \sqrt{\frac{1}{T}\sum_{t=1}^{T} x_t^2} \quad (8)$$

The amplitude variance of signal samples is used to calculate the Hjorth activity that is written as [44]-[45]:

$$Activity = var\,(x_t) \quad (9)$$

The frequency's mean approximation is determined by the Hjorth mobility that is written as [44]-[45]:

$$Mobility = \sqrt{\frac{var(x_t')}{var\,(x_t)}} \quad (10)$$

The power spectrum's standard deviation is determined by Hjorth complexity that is written as [44], [45]:

$$Complexity = \frac{Mobility(x_t')}{Mobility\,(x_t)} \quad (11)$$

Waveform length (WL) represents the cumulative absolute difference between adjacent samples and provides a measure of the signal's overall variation. It can be written as follows [46]:

$$WL = \sum_{t}^{T} |x_t - x_{t-1}| \quad (12)$$

The skewness is written as [37]:

$$skewness = \frac{1}{T}\frac{\sum_{n=1}^{N}(x_t - \bar{x})^3}{\sigma^3} \quad (13)$$

The kurtosis is written as [37]:

$$kurtosis = \frac{1}{T}\frac{\sum_{n=1}^{N}(x_t - \bar{x})^4}{\sigma^4} \quad (14)$$

where T is the number of samples, $x_t$is the processed signal, x_t' is the first derivative of the signal sample x_t, $var\,(x_t)$ is the variance of the signal sample x_t, $\bar{x}$ is the mean of the signal and t is an integer that belongs to 1 to T.

In F3, all features in F1 and F2 were combined and represented as F3 = $F1_{D1} \cup F1_{D2} \cup F1_{D3} \cup F2$ with twenty-three (23) features in total. The Kruskal-Wallis test was applied to select the features of F3 that have a p-value of less than 0.05. F3 in combination with the feature selection is represented as FS. The approach enables us to consider the unique features of both sets, providing a more comprehensive analysis. The features were evaluated and analyzed to investigate the enhancement to the classification performance.

3.4 Classification

The SVM classifier is one of the most frequently used methods for MI task classification. The SVM aims to accomplish both accurate classification and robust generalization by maximizing machine performance while minimizing the complexity of the learned model [47]. The SVM identifies the hyperplane that maximizes the margin between different classes of data points. The margin is the distance between the hyperplane and the nearest data points from each class. This is also known as support vectors. SVMs sometimes give a better fit and are computationally more efficient [48].

The LR has a low risk of overfitting, because the model complexity is minimal [49]. It determines a relationship between a single or a set of independent variables (features) and the likelihood that the dependent variable will fall into a specific class. In the LR, the result always falls between 0 and 1 by using the logistic function representing the estimated probability of the positive class. Based on the probability, the LR model creates a decision boundary that separates the two classes. During training, the method changes the model's parameters to reduce the difference between the predicted probability and the actual binary labels in the training data. The data points near the margin have significantly less influence due to the logit transform [48].

Naïve Bayes (NB) is useful when most or all the predictor variables are also binary or categorical. Given the class label, the assumption in the NB is that all features are conditionally independent and this simplifies the modeling process [50]. The NB can also be applied in situations where there are three or more possible outcomes. The strategy in this case is to determine the probabilities of each possible outcome before selecting the one with the highest probability. However, rather than being highly precise values, the estimated probabilities should be considered approximation figures when the assumption of conditional independence is violated [48]. The NB classification algorithm has demonstrated high CA when applied to a limited sample dataset utilizing the Poisson distribution model [50].

Performance was evaluated by the CA and F1-score. The accuracy is defined as the ratio of correctly identified samples or observations to the total number of input samples in the same class. It describes the classifier's effectiveness in performing its tasks successfully. The F1-score is a metric that balances both precision and recall. The equations are written as [51]:

$$Accuracy = \frac{TP+TN}{TN+TP+FN+FP} \times 100\% \quad (15)$$

$$F1_{score} = \frac{2TP}{2TP+FP+FN} \quad (16)$$

where true positive (TP) refers to the trials in the experiment that are correctly labelled as positive. The term "TN" refers to true negatives, which represent the number of trials correctly classified as negative, while “FP” is false positive representing the trials incorrectly classified as positive and “FN” is false negative when the trials are incorrectly classified as negative.

By preventing over-fitting, the cross-validation approach enhances model efficiency. The performance parameters were evaluated using a ten-fold cross-validation approach. A ten-fold cross-validation of results divides features into ten segments or folds of a similar scale. It consists of nine training sets and one testing set, whereby the model is trained in each round with the training sets and evaluated using one testing set. The average accuracy is computed with ten rounds of the process. The classification is implemented using MATLAB R2022a.

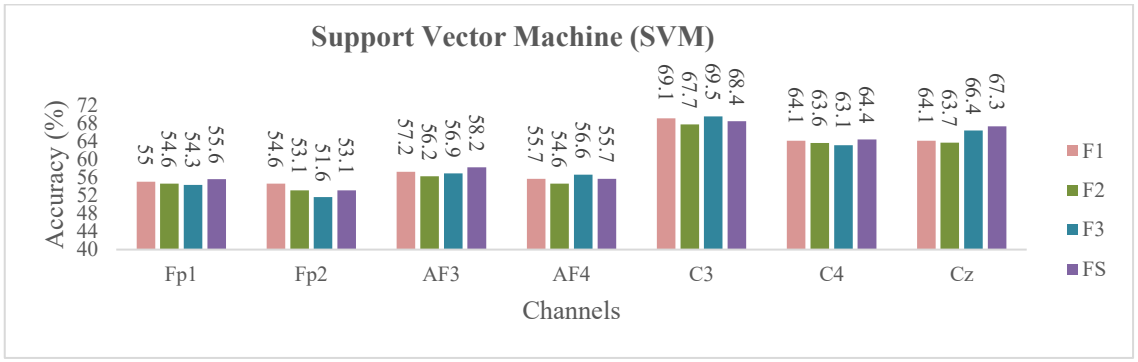
4. RESULTS

This section consists of two sub-sections. The first sub-section discusses the classification-performance result obtained by combining the feature vectors of all participants. The second sub-section focuses on the classification performance of the FS across all the participants and channels.

4.1 Performance Evaluation of All Participants

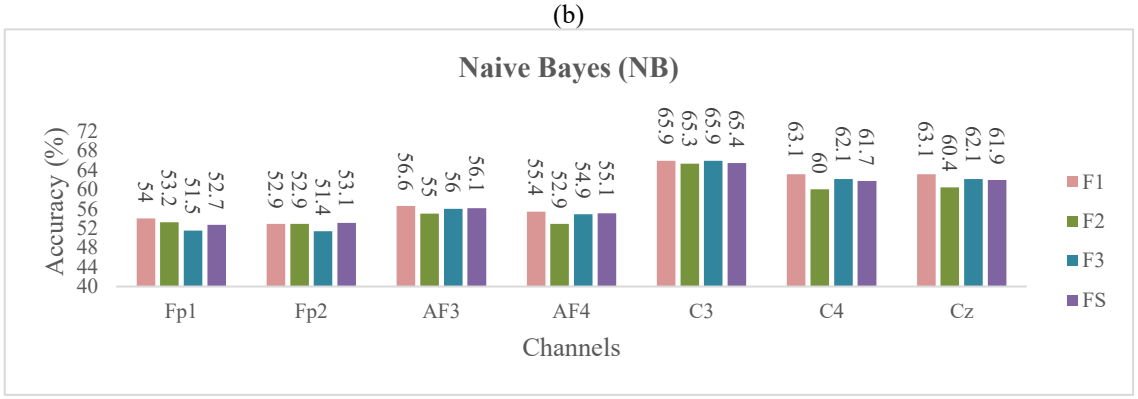
Figure 3 shows the CA of MI tasks when the features vectors of all participants in the Dataset 1 were combined. The accuracy of all channels is less than 70%, possibly due to significant inter-individual variability in brain signals, including cognitive ability and activity patterns, that has an impact on overall performance [16]. Employing F3 without FS along with Support Vector Machines (SVMs) has been shown to improve the CA of EEG signals, specifically from AF4, C3 and Cz, as shown in Figure 3(a).

On the other hand, the implementation of FS led to a decrease in the CA for AF4 and C3, while Cz, Fp1, AF3 and C4 showed an improvement in the CA. This implies that FS had a different impact on the various channels, potentially due to the specific features that were chosen and their significance according to the classification task. Figure 3(b) shows that F3 with LR also resulted in positive results for enhancing the CA of Cz. Furthermore, FS resulted in further improvement in the CA of Cz. It indicates that applying FS, along with LR, can be a beneficial approach for getting a higher CA of Cz. Furthermore, FS led to higher CA, not only for Cz, but also for other channels, such as Fp1, AF3 and C4, as shown in Figure 3(c). The figure also illustrates the lack of CA improvement when combining F3 with the NB classifier. This indicates that the interaction between F3 and NB, as well as FS and NB, did not enhance the CA. It emphasizes the incompatibility of using F3 and FS together with NB. In overall, C3 has higher CA across all classifiers.



(a)

(b)



(c)

Figure 3. Dataset 1 classification accuracy (%) comparison across experiments and channels using a) SVM, b) LR and c) NB.

A statistical analysis was applied to determine whether there are statistically significant differences in the CA between three different classifiers in separate experiments using Dataset 1. Due to the small sample size, the Friedman test was conducted. The findings demonstrated a statistically significant difference in the accuracy of the classifiers when assessed using feature sets F1, F2 and F3. This was shown by the p-value of 0.004, 0.008 and 0.018, respectively. This suggests that the selection of a

classifier, when used in conjunction with the features from F1, F2 or F3, significantly affects the CA. In contrast, the use of classifiers based on the feature set of FS did not have a significant effect on the CA. This is supported by a p-value of 0.104, which is higher than 0.05.

Figure 4 shows the CA of MI tasks when the feature vectors of nineteen participants in Dataset 2 were combined. The accuracy of all channels is also less than 70%, possibly due to significant inter-individual variability in brain signals that have an impact on the overall performance [16]. Figure 4(a) depicts that employing F3 in the absence of FS in conjunction with Support Vector Machines (SVM) has demonstrated its effectiveness in improving the CA of EEG signals, specifically from Fp2, C4 and Cz. On the other hand, the implementation of FS led to an increase in the CA for C3. Figure 4(b) shows that F3 with LR also resulted in positive results for enhancing the CA of C4, Cz and AF4. Furthermore, FS resulted in higher CA, not only for Fp2, but also for other channels; AF4 and C3. FS resulted in further improvement in the CA of AF4. Figure 4(c) illustrates the positive results in terms of CA improvement of Fp1, C3, C4 and Cz resulting from the combination of F3 and NB. The interaction between F3 and NB, as well as FS and NB, did not result in an improvement of AF3. It shows that F2 resulted in higher CA for AF3 compared to F3 and FS.

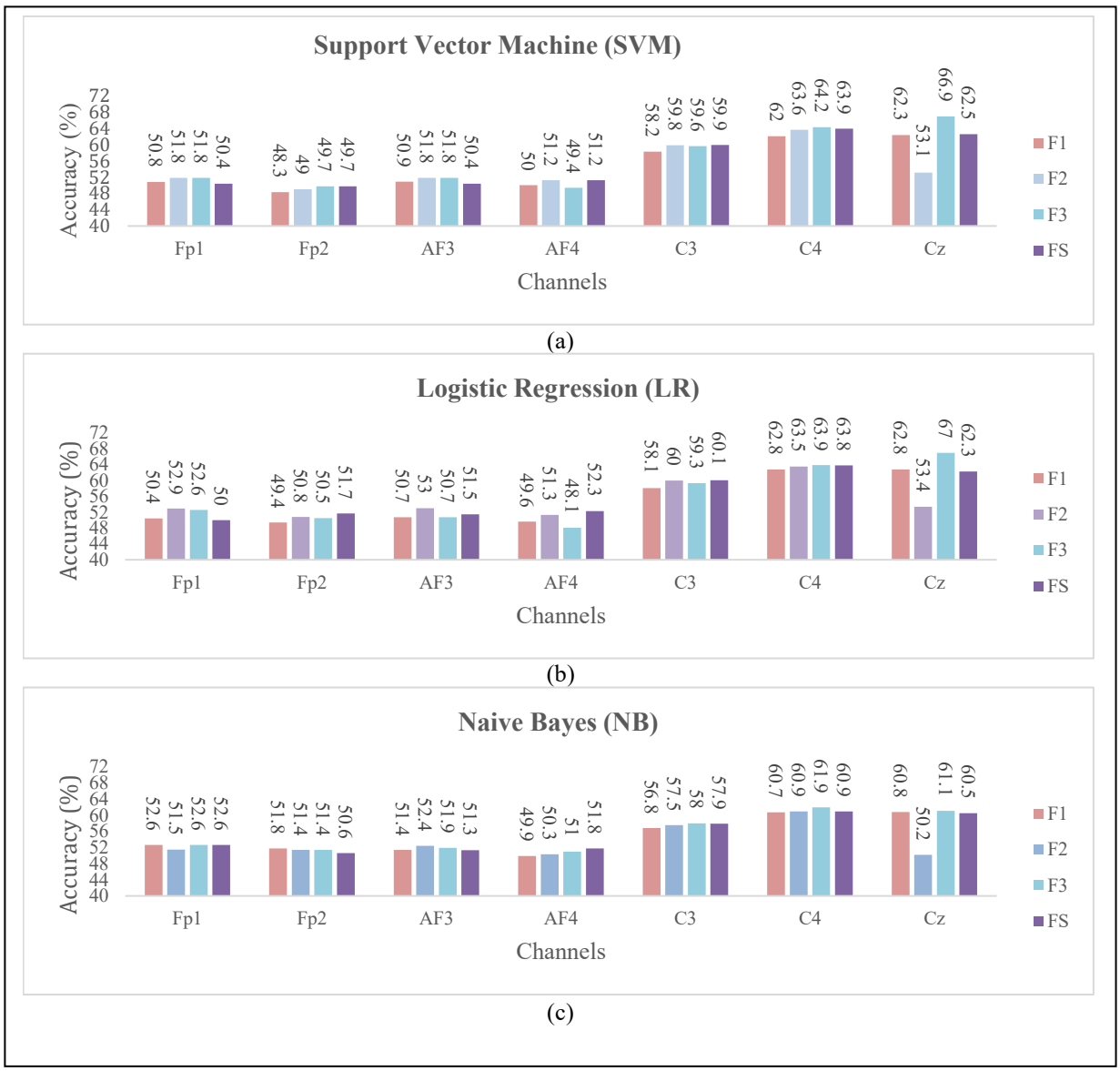


Figure 4. Dataset 2 classification accuracy (%) comparison across experiments and channels using a) SVM, b) LR and c) NB.

A statistical analysis was applied to determine whether there are statistically significant differences in the CA between three different classifiers in separate experiments using Dataset 2. The Friedman test

revealed no significant difference in the accuracy of the classifiers across feature sets F1, F3 and FS. This was shown by a p-value exceeding 0.05. This suggests that the selection of a classifier, when applied with F1, F3 and FS, does not significantly affect the CA. In contrast, classifiers using the F2 feature set significantly impact the CA as indicated by a p-value of 0.05.

Table 2 presents F1-scores for *rh* and *rf* in Dataset 1 and *rh* and *lh* in Dataset 2 across the channels and classifiers. The results show that *rh* generally has higher scores than *rf* and *lh*, indicating better classifier performance for *rh* tasks. This suggests that the features are more distinct and easier to identify. It is important to note that the F1-score obtained from the combination of feature vectors for all participants can vary due to the variation in participants' abilities in the MI tasks. There are challenges in classifying either *rf* or *lh* tasks, as seen in varied F1-scores except for C4 and Cz in Dataset 2. It can be associated with features and data patterns of *lh* that are more dominant and easier to recognize. This highlights the importance optimizing the feature extraction or selection approach for improving the CA. In the previous study utilizing the same dataset as Dataset 1, it was also observed that *rh* exhibited a higher accuracy compared to *rf* when employing the WPD- k-NN method [33].

Table 2. $F1_{score}$ (%) of FS for Dataset 1 and Dataset 2.

CHN	Dataset 1						Dataset 2					
	SVM		LR		NB		SVM		LR		NB	
	rh	***rf***	***rh***	***rf***	***rh***	***rf***	***rh***	***lh***	***rh***	***lh***	***rh***	***lh***
Fp1	59.1	51.6	57.5	54.17	60.6	40.9	63.1	24.3	57.8	38.6	63.4	32.8
Fp2	57.9	47.3	55.2	50.33	60.8	41.5	64.7	12.5	58.8	41.7	62.3	28.1
AF3	58.9	57.5	59.1	57.70	55.6	56.5	62.3	27.7	56.9	44.6	61.8	62.4
AF4	58.3	52.8	56.5	53.23	58.6	51.1	55.2	46.4	54.0	50.5	57.4	44.4
C3	68.2	68.5	67.9	68.49	67.4	63.3	61.6	58.1	60.6	59.6	64.2	48.9
C4	64.8	63.9	65.6	64.43	63.4	59.9	62.3	65.3	62.9	64.6	53.2	66.4
Cz	67.8	66.4	68.2	67.08	63.1	60.5	58.8	65.7	59.8	64.5	52.7	66.1

4.2 Performance Evaluation of Individual Participants

The CA for each participant was determined by classifying features that were selected through the feature-selection process (FS). They were analyzed to get a more in-depth study of the proposed technique, as shown in Table 3. The CA exceeding 70% is presented in boldface.

In Dataset 1, noteworthy results were achieved with different classifiers. When employing the SVM, *aw* achieved the highest CA at 81.4% by utilizing commonly employed channels, demonstrating the robustness of the SVM in capturing neural patterns related to MI. In addition, the channel on the forehead had the highest CA at 63.6%. It was performed by *aw* with AF4. Switching to the LR as the classifier, *al* achieved the CA of 87.5% via C3 which represents the highest CA obtained for the dataset. *al* also demonstrated the CA of 62.1%, specifically from AF4. The NB achieves a maximum CA of 79.3% when using the C3 channel. The highest achievable accuracy for the forehead channel is 60%, specifically from the AF3 channel. The results are in line with the finding in [21], where AF4 was identified as one of the most informative channels for *aw* and both AF3 and AF4 for *al*.

In Dataset 2, S36 gets the maximum CA of 93% by classifying the features of C4 using the SVM. S36 outperformed other participants with features from both C4 and Cz channels across all classifiers. When examining the channels on the forehead (Fp1 and AF3), S29 demonstrated great performance compared to other participants. The CA for S29 exceeded 70% and reached up to 86% across all classifiers, which is sufficient for BCI. There are participants getting higher CA on the channels that are located on the forehead compared to the channels that are commonly used for MI. This is demonstrated by S28 and S29 with SVM. For example, the CA of AF4 (64%) is higher than that of C3 (51%) and C4 (58%), while S29 has a higher CA of Fp1 (86%) and AF3 (77%) compared to the CA of C3 (70%), C4 (60%) and Cz (59%).

Table 3. The minimum and maximum values of CA across classifiers and channels.

Dataset	Classifier	Level	Channels						
			Fp1	**Fp2**	**AF3**	**AF4**	**C3**	**C4**	**Cz**
Dataset 1	SVM	Min	50.7 (*av*)	52.1 (*aa*)	48.2 (*aa*)	56.1 (*aw*)	52.1 (*ay*)	52.1 (*ay*)	53.7 (*ay*)
		Max	61.8 (*al*)	58.6 (*al*)	60.7 (*al*)	63.6 (*aw*)	**81.4 (*aw*)**	**81.4 (*aw*)**	**79.3 (*aw*)**
	LR	Min	51.1 (*av*)	50.4 (*av*)	52.1 (*av*)	53.2 (*aw*)	55.4 (av)	54.3 (ay)	55.7 (ay)
		Max	57.5 (al)	58.2 (*aw*)	59.6 (*aw*)	62.1 (*al*)	**87.5 (*al*)**	**80.7 (*aw*)**	**79.3 (*aw*)**
	NB	Min	52.1 (*ay*)	50.4 (*aa*)	53.2 (*aa*)	55.4 (*aa*)	55.7 (aw)	57.9 (ay)	55 (ay)
		Max	56.1 (*al*)	56.1 (*ay*)	60.0 (*al*)	58.9 (*al*)	**79.3 (*al*)**	**73.2 (*aw*)**	**74.3 (*aw*)**
Dataset 2	SVM	Min	45.00 (S32,S33 &S44)	38.00 (S32)	42.00 (S9)	33 (S32)	40.00 (S1)	40.00 (S2)	41.00 (S2)
		Max	**86 (S29)**	63.00 (S3&S22)	**77 (S29)**	64 (S28)	**75 (S44)**	**93 (S36)**	**91.00 (S36)**
	LR	Min	41 (S44)	37 (S5)	37 (S2)	42 (S1&S32)	45 (S19)	45 (S5)	40 (S1)
		Max	**78 (S29)**	66 (S3)	**73 (S29)**	67 (S29)	**73 (S37)**	**86 (S36)**	**85 (S36)**
	NB	Min	44 (S36)	39 (S5)	38 (S2)	36 (S52)	43 (S1)	44 (S2)	38 (S2)
		Max	**81 (S29)**	64 (S18)	**72 (S29)**	63 (S18)	**75 (S44)**	**88 (S36)**	**85 (S36)**

5. DISCUSSION

Based on the result of combining the feature vectors of all participants, C3 and C4 exhibit the highest CA across all classifiers for Dataset 1 and Dataset 2, respectively. C3 and C4 were used a lot in past studies, because they are placed over the motor cortex of the brain and can record unique patterns during MI activities [30], [35], [40]. The SVM and LR produce comparable results, because each model uses all data points, with points closer to the margin having considerably less impact [48]. The F3 features, extracted from EEG signals across both datasets, could provide relevant information for classification. Further improvement in the CA is also possible with feature selection. This suggests that by selecting relevant features with an appropriate classifier, the CA could be greatly improved. Further exploration might also expand the potential of Fp1, AF3 and AF4 channels for a single-channel execution in MI. Moreover, the positioning of the AF3 and AF4 channels on the forehead, away from the eyes, could offer practical advantages by minimizing the direct impact of eye blinks compared to Fp1 and Fp2.

Regarding the performance of individual participants, the study demonstrated that the commonly employed channels consistently achieved accuracy levels exceeding 70% across different classifiers using the proposed approach in both datasets. The threshold of 70% is generally recognized as meeting the requirements needed for effective BCI implementation. This indicates the practicality and reliability of the proposed approach for BCI applications [39]. AF3 and AF4 showed the potential for use in the single-channel BCI execution, as they frequently exhibited accuracies higher than 60% with the proposed approach. This is also considered as BCI literacy threshold determined in the previous MI study [52]. Further exploration could strengthen the practicality and reliability of a single-channel BCI systems for broader implementation in real-world applications. Moreover, S29 offers the CA that is comparable to the common channels. Despite the recognition of C3 and C4 as the channels that offer superior MI features [40], there are participants achieving higher CA using channels other than C3 or C4 [18][19][20]. The example is shown by S28 and S29 by using SVM as the classifier. It is strongly affected by the involvement of participant, as well as the well-designed experimental conditions and the implementation of an effective classification algorithm [17], [21].

Table 4 presents several existing techniques employed on the Dataset 1. For any of the five subjects, the proposed approach does not provide the best CA. Most of the past research used more than two channels to get more relevant information or features, resulting in high CA. The study conducted by Khare and

Bajaj [19] successfully obtained very high CA. However, they have different approaches to select the best single channel for further processing. They were using multi-cluster unsupervised learning channel selection (MCCS) to rank or find the best single channel from the 118 channels in the BCI system. Even though they are using the same dataset, the most informative channel was different when they used different methods for channel selection [20]. For Dataset 2, the CA is compared with the CA from the first session of the previous study [39] with the same participants. The technique compared include the 10-fold cross validation with the CSP (CSP-cv) and Linear Discriminant Analysis (LDA) as the classifier. They employed 20 channels that are in the motor cortex region for the analysis. The average of the CA for 19 participants is 87.14±9.14 which is 33.85% greater than the average of the CA in AF4 (53.29±9.12) and 22.43% higher than the average of the CA in C4 (64.71±11.49). S33 achieves the highest CA of 98.1%. For this study, the highest CA is 93% which is 4.9% lower than S33. Even though the overall performance is not comparable to the previous work, the CA for certain participants is relatively promising when consider the number of channels.

Table 4. A comparison of the CA for Dataset 1 across different participants.

Author(s)	**Approach**	**No. of channels**	***aa***	***al***	***av***	***aw***	***ay***	**Average ± std.**
Selim et al. [27]	CSP\AM-BA-SVM	18	86.10	100	66.84	90.63	80.95	85.00± 12.29
Khare and Bajaj [19]	F-VMD\ F-ELM	1 (AF3)	100	100	100	100	100	100
Roy et al. [53]	HWE/ Decision Tree	16	100	87.50	100	71.87	100	91.87± 12.42
Tiwari [21]	LS-BJOA/ RCSP	In the bracket below the CA	89.34 (29)	94.08 (18)	80.54 (37)	93.5 (31)	87.68 (23)	89.02± 4.88 (27.6)
Gao et al. [54]	SR-TT/ SVM-RBF	118	84.64	91.07	82.50	87.50	81.07	85.36
Proposed approach	DWT and time domain features/ LR	1(C3)	60	**87.5**	55.4	60.4	76.4	67.94±13.52
		1 (AF4)	59.3	**62.1**	57.5	53.2	54.3	57.28±3.64

There are a few limitations to the proposed approach. Each participant's brain activity during MI might exhibit a unique pattern and variation. The approach might not be able to capture specific patterns from certain participants, including those who might be BCI illiterate. The applicability of Kruskal-Wallis for feature selection may be restricted for some individual participants. There is a possibility that the accuracy of the findings could decrease after selecting specific features, indicating the inconsistency of the selector in improving the performance of the classification. The selection could lead to the elimination of valuable information for certain participants, hence making the MI tasks difficult to interpret. The study is limited to the datasets of healthy participants. The results may differ when applied to a specific group of individuals with relevant health conditions.

6. CONCLUSION

In this study, the proposed approach was applied to Dataset 1 (BCI Competition III (IVa)) and Dataset 2 (OpenBMI) to evaluate the classification performance. The approach slightly increased the CA on certain channels with F3 and FS, compared to relying only on the DWT decomposition. While not all channels showed an increase in the CA for all participants, the CA of individual performance improved notably. Particularly, *al* reached up to 87.5% of CA on C3 by using LR and S36 achieved 93% of CA on the C4 channel. Additionally, participant S29 achieved sufficient CA on Fp1 and AF3 channels that is comparable to those of commonly used channels for MI. This suggests that the proposed approach, when combined with relevant features and appropriate classifiers, has the potential to improve overall classification performance. This extends its applicability to the forehead channels, necessitating further investigation of the channels in the context of our study. Although the proposed approach encounters difficulties in achieving high CA across participants, there is a room for improvement through comprehensive evaluation. Evaluating the effectiveness of the approach in different MI tasks and participant groups can aid in determining its strengths and limitations in various contexts. To optimize the classification performance, other feature-selection techniques, such as hybrid or wrapper methods,

could be explored. Besides that, it would be valuable to conduct a comprehensive comparison between the proposed approach and the standard methods, such as the CSP technique, to identify their respective strengths and weaknesses.

REFERENCES

[1] J. L. Collinger et al., "Functional Priorities, Assistive Technology and Brain-Computer Interfaces After Spinal Cord Injury," J. of Rehabilitation Research and Development, vol. 50, no. 2, pp. 145–160, 2013.

[2] E. K. St. Louis, L. C. Frey and J. W. Britton, Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children and Infants, Chicago: American Epilepsy Society; PMID: 27748095, 2016.

[3] L. Kauhanen et al., "EEG and MEG Brain-Computer Interface for Tetraplegic Patients," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 14, no. 2, pp. 190–193, Jun. 2006.

[4] N. Kulkarni and V. Bairagi, "Electroencephalogram and Its Use in Clinical Neuroscience," in Book: EEG-based Diagnosis of Alzheimer Disease, pp. 25–35, DOI: 10.1016/B978-0-12-815392-5.00002-2, 2018.

[5] J. Liao, J. Wang, C. A. Zhan and F. Yang, "Parameterized Aperiodic and Periodic Components of Single-channel EEG Enables Reliable Seizure Detection," Physical and Engineering Sciences in Medicine, DOI: 10.1007/s13246-023-01340-6, Sep. 2023.

[6] G. Kaushik, P. Gaur, R. R. Sharma and R. B. Pachori, "EEG Signal Based Seizure Detection Focused on Hjorth Parameters from Tunable-Q Wavelet Sub-bands," Biomed. Signal Process. Control, vol. 76, p. 103645, DOI: 10.1016/j.bspc.2022.103645, Jul. 2022.

[7] A. Babiker and I. Faye, "A Hybrid EMD-Wavelet EEG Feature Extraction Method for the Classification of Students' Interest in the Mathematics Classroom," Applied Computational Intelligence and Soft Computing, vol. 2021, pp. 1–8, DOI: 10.1155/2021/6617462, Jan. 2021.

[8] M. Zhong, Q. Yang, Y. Liu, B. Zhen, F. Zhao and B. Xie, "EEG Emotion Recognition Based on TQWT-features and Hybrid Convolutional Recurrent Neural Network," Biomed. Signal Process. Control, vol. 79, p. 104211, DOI: 10.1016/j.bspc.2022.104211, Jan. 2023.

[9] M. F. Mridha et al., "Brain-Computer Interface: Advancement and Challenges," Sensors, vol. 21, no. 17, p. 5746, DOI: 10.3390/s21175746, Aug. 2021.

[10] X. Gu et al., "EEG-based Brain-Computer Interfaces (BCIs): A Survey of Recent Studies on Signal Sensing Technologies and Computational Intelligence Approaches and Their Applications," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 5, pp. 1645–1666, 2021.

[11] M. Rashid et al., "Current Status, Challenges and Possible Solutions of EEG-based Brain-computer Interface: A Comprehensive Review," Frontiers in Neurorobotics, vol. 14. Frontiers Media S.A., DOI: 10.3389/fnbot.2020.00025, Jun. 03, 2020.

[12] Jusas and Samuvel, "Classification of Motor Imagery Using a Combination of User-specific Band and Subject-specific Band for Brain-Computer Interface," Applied Sciences, vol. 9, no. 23, p. 4990, 2019.

[13] C. Neuper, M. Wörtz and G. Pfurtscheller, "ERD/ERS Patterns Reflecting Sensorimotor Activation and Deactivation," Progress in Brain Research, pp. 211–222, DOI: 10.1016/S0079-6123(06)59014-4, 2006.

[14] S. Saha et al., "Progress in Brain Computer Interface: Challenges and Opportunities," Frontiers in Systems Neuroscience, vol. 15, DOI: 10.3389/fnsys.2021.578875, Feb. 25, 2021.

[15] L. Brusini, F. Stival, F. Setti, E. Menegatti, G. Menegaz and S. F. Storti, "A Systematic Review on Motor-imagery Brain Connectivity-based Computer Interfaces," IEEE Transactions on Human-Machine Systems, vol. 51, no. 6, pp. 725–733, DOI: 10.1109/THMS.2021.3115094, Dec. 2021.

[16] P. Gaur, K. McCreadie, R. B. Pachori, H. Wang and G. Prasad, "An Automatic Subject Specific Channel Selection Method for Enhancing Motor Imagery Classification in EEG-BCI Using Correlation," Biomed Signal Process Control, vol. 68, p. 102574, DOI: 10.1016/j.bspc.2021.102574, Jul. 2021.

[17] L. Zhang and Q. Wei, "Channel Selection in Motor Imaginary-based Brain-Computer Interfaces: A Particle Swarm Optimization Algorithm," J. of Integrative Neuroscience, vol. 18, no. 2, pp. 141–152, DOI: 10.31083/j.jin.2019.02.17, 2019.

[18] S. Ge, R. Wang and D. Yu, "Classification of Four-class Motor Imagery Employing Single-channel Electroencephalography," PLoS One, vol. 9, no. 6, p. e98019, Jun. 2014.

[19] S. K. Khare and V. Bajaj, "A Facile and Flexible Motor Imagery Classification Using Electroencephalogram Signals," Computer Methods and Programs in Biomedicine, vol. 197, p. 105722, DOI: 10.1016/j.cmpb.2020.105722, Dec. 2020.

[20] S. K. Khare, N. Gaikwad and N. D. Bokde, "An Intelligent Motor Imagery Detection System Using Electroencephalography with Adaptive Wavelets," Sensors, vol. 22, no. 21, p. 8128, Oct. 2022.

[21] A. Tiwari, "A Logistic Binary Jaya Optimization-based Channel Selection Scheme for Motor-imagery Classification in Brain-Computer Interface," Expert Systems with Applications, vol. 223, p. 119921, DOI: 10.1016/j.eswa.2023.119921, Aug. 2023.

[22] K. Kotegawa, A. Yasumura and W. Teramoto, "Activity in the Prefrontal Cortex during Motor Imagery of Precision Gait: An fNIRS Study," Experimental Brain Research, vol. 238, no. 1, pp. 221–228, 2020.

[23] L. Almulla, I. Al-Naib, I. S. Ateeq and M. Althobaiti, "Observation and Motor Imagery Balance Tasks Evaluation: An fNIRS Feasibility Study," PLoS One, vol. 17, no. 3, p. e0265898, Mar. 2022.

[24] S. Glover, E. Bibby and E. Tuomi, "Executive Functions in Motor Imagery: Support for the Motor-cognitive Model over the Functional Equivalence Model," Experimental Brain Research, vol. 238, no. 4, pp. 931–944, DOI: 10.1007/s00221-020-05756-4, Apr. 2020.

[25] J. A. Wilson, G. Schalk, L. M. Walton and J. C. Williams, "Using an EEG-based Brain-Computer Interface for Virtual Cursor Movement with BCI2000," Journal of Visualized Experiments, no. 29, DOI: 10.3791/1319, Jul. 2009.

[26] S. S. Moumgiakmas and G. A. Papakostas, "Robustly Effective Approaches on Motor Imagery-based Brain Computer Interfaces," Computers, vol. 11, no. 5, p. 61, DOI: 10.3390/computers11050061, 2022.

[27] S. Selim, M. Tantawi, H. Shedeed and A. Badr, "A Comparative Analysis of Different Feature Extraction Techniques for Motor Imagery Based BCI System," Proc. of the Int. Conf. on Artificial Intelligence and Computer Vision (AICV2020), pp. 740–749, DOI: 10.1007/978-3-030-44289-7_69, 2020.

[28] J. Camacho and V. Manian, "Real-time Single Channel EEG Motor Imagery Based Brain Computer Interface," Proc. of the IEEE 2016 World Automation Congress (WAC), pp. 1–6, DOI: 10.1109/WAC.2016.7582973, Rio Grande, PR, USA, Jul. 2016.

[29] L.-W. Ko, S. S. K. Ranga, O. Komarov and C.-C. Chen, "Development of Single-channel Hybrid BCI System Using Motor Imagery and SSVEP," J. of Healthcare Eng., vol. 2017, pp. 1–7, DOI: 10.1155/2017/3789386, 2017.

[30] R. Chen et al., "Enhancement of Time-frequency Energy for the Classification of Motor Imagery Electroencephalogram Based on an Improved FitzHugh–Nagumo Neuron System," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 282–293, 2023.

[31] G. Rodríguez-Bermúdez and P. J. García-Laencina, "Automatic and Adaptive Classification of Electroencephalographic Signals for Brain Computer Interfaces," J. of Medical Systems, vol. 36, no. S1, pp. 51–63, DOI: 10.1007/s10916-012-9893-4, Nov. 2012.

[32] R. R. Sharma and R. B. Pachori, "A New Method for Non-stationary Signal Analysis Using Eigenvalue Decomposition of the Hankel Matrix and Hilbert Transform," Proc. of the 2017 4th IEEE Int. Conf. on Signal Processing and Integrated Networks (SPIN), pp. 484–488, DOI: 10.1109/SPIN.2017.8049998, Feb. 2017.

[33] J. Kevric and A. Subasi, "Comparison of Signal Decomposition Methods in Classification of EEG Signals for Motor-imagery BCI System," Biomed. Signal Process. Control, vol. 31, pp. 398–406, DOI: 10.1016/j.bspc.2016.09.007, Jan. 2017.

[34] O. Attallah, J. Abougharbia, M. Tamazin and A. A. Nasser, "A BCI System Based on Motor Imagery for Assisting People with Motor Deficiencies in the Limbs," Brain Sciences, vol. 10, no. 11, p. 864, DOI: 10.3390/brainsci10110864, Nov. 2020.

[35] Ji, Ma, Dong and Zhang, "EEG Signals Feature Extraction Based on DWT and EMD Combined with Approximate Entropy," Brain Sciences, vol. 9, no. 8, p. 201, DOI: 10.3390/brainsci9080201, Aug. 2019.

[36] A. al-Qerem, F. Kharbat, S. Nashwan, S. Ashraf and K. Blaou, "General Model for Best Feature Extraction of EEG Using Discrete Wavelet Transform Wavelet Family and Differential Evolution," Int. J. of Distributed Sensor Networks, vol. 16, no. 3, p. 155014772091100, Mar. 2020.

[37] G. C. Jana, A. Agrawal, P. K. Pattnaik and M. Sain, "DWT-EMD Feature Level Fusion Based Approach over Multi and Single Channel EEG Signals for Seizure Detection," Diagnostics, vol. 12, no. 2, p. 324, DOI: 10.3390/diagnostics12020324, Jan. 2022.

[38] B. Blankertz et al., "The BCI Competition III: Validating Alternative Approaches to Actual BCI Problems," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 14, no. 2, pp. 153–159, DOI: 10.1109/TNSRE.2006.875642, Jun. 2006.

[39] M.-H. Lee et al., "EEG Dataset and OpenBMI Toolbox for Three BCI Paradigms: An Investigation into BCI Illiteracy," Gigascience, vol. 8, no. 5, DOI: 10.1093/gigascience/giz002, May 2019.

[40] S. Kanoga, A. Kanemura and H. Asoh, "A Comparative Study of Features and Classifiers in Single-channel EEG-based Motor Imagery BCI," Proc. of the 2018 IEEE Global Conf. on Signal and Information Processing (GlobalSIP), pp. 474–478, DOI: 10.1109/GlobalSIP.2018.8646636, Nov. 2018.

[41] M. Al-Quraishi, I. Elamvazuthi, S. Daud, S. Parasuraman and A. Borboni, "EEG-based Control for Upper and Lower Limb Exoskeletons and Prostheses: A Systematic Review," Sensors, vol. 18, no. 10, p. 3342, DOI: 10.3390/s18103342, Oct. 2018.

[42] H. U. Amin et al., "Feature Extraction and Classification for EEG Signals Using Wavelet Transform and Machine Learning Techniques," Australasian Physical and Engineering Sciences in Medicine, vol. 38, no. 1, pp. 139–149, DOI: 10.1007/s13246-015-0333-x, Mar. 2015.

[43] A. A. Abdul-latif, I. Cosic, D. K. Kumar, B. Polus and C. da_Costa, "Power Changes of EEG Signals Associated with Muscle Fatigue: The Root Mean Square Analysis of EEG Bands," Proc. of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conf., pp. 531–534, DOI: 10.1109/ISSNIP.2004.1417517, 2004.

[44] B. Hjorth, "EEG Analysis Based on Time Domain Properties," Electroencephalography and Clinical Neurophysiology, vol. 29, no. 3, pp. 306–310, DOI: 10.1016/0013-4694(70)90143-4, Sep. 1970.
[45] M. S. Safi and S. M. M. Safi, "Early Detection of Alzheimer's Disease from EEG Signals Using Hjorth Parameters," Biomed. Signal Process. Control, vol. 65, p. 102338, DOI: 10.1016/j.bspc.2020.102338, 2021.
[46] F. Lotte, "A New Feature and Associated Optimal Spatial Filter for EEG Signal Classification: Waveform Length," Proc. of the 21st Int. Conf. on Pattern Recognition (ICPR2012), pp. 1302–1305, Tsukuba, Japan, 2012.
[47] D. Garrett, D. A. Peterson, C. W. Anderson and M. H. Thaut, "Comparison of Linear, Nonlinear and Feature Selection Methods for EEG Signal Classification," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 11, no. 2, pp. 141–144, DOI: 10.1109/TNSRE.2003.814441, Jun. 2003.
[48] P. Nadkarni, "Core Technologies: Machine Learning and Natural Language Processing," Clinical Research Computing, pp. 85–114, DOI: 10.1016/B978-0-12-803130-8.00004-X, Elsevier, 2016.
[49] S. Dreiseitl and L. Ohno-Machado, "Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review," J. of Biomedical Informatics, vol. 35, no. 5–6, pp. 352–359, Oct. 2002.
[50] Y. Huang and L. Li, "Naive Bayes Classification Algorithm Based on Small Sample Set," Proc. of the 2011 IEEE Int. Conf. on Cloud Computing and Intelligence Systems, pp. 34–39, DOI: 10.1109/CCIS.2011.6045027, Sep. 2011.
[51] Pawan and R. Dhiman, "Motor Imagery Signal Classification Using Wavelet Packet Decomposition and Modified Binary Grey Wolf Optimization," Measurement: Sensors, vol. 24, p. 100553, DOI: 10.1016/j.measen.2022.100553, Dec. 2022.
[52] M. Ahn, H. Cho, S. Ahn and S. C. Jun, "High Theta and Low Alpha Powers May Be Indicative of BCI-illiteracy in Motor Imagery," PLoS One, vol. 8, no. 11, p. e80886, Nov. 2013.
[53] G. Roy, A. K. Bhoi and S. Bhaumik, "A Comparative Approach for MI-Based EEG Signals Classification Using Energy, Power and Entropy," IRBM, vol. 43, no. 5, pp. 434–446, Oct. 2022.

ملخص البحث:

لقد ثبت نجاح استخدام بينيّةٍ تربط الدّماغ بالحاسوب مبنيّة على التّصوير الحركي كآلية تحكّم مع قنوات متعدّدة لتصوير الدّماغ. ولأغراضٍ عمليةٍ، فإنّ من المفضّل تقليل عدد قنوات صور تخطيط الدّماغ.

هذه الورقة تبحث في قناةٍ مفردةٍ لصور تخطيط الدّماغ في سياق التّصوير الحركي. علاوة على ذلك، فإنّ عدد السِّمات الّتي يمكن استخلاصها من نظامٍ ذي قناةٍ واحدة يكون غير كافٍ، لذا فإنّ استخدام تقنيةٍ فعّالةٍ لاستخلاص السِّمات يلعب دوراً حاسماً في التّغلُّب على هذا المحدِّد.

من هنا، تقترح هذه الورقة دمج استخلاص السِّمات بواسطة الانتقال المجرَّد للمويجات وبواسطة المجال الزّمني من أجل توفير معلوماتٍ أدقّ لأغراض التّصنيف. وقد تمّ تجريب التّقنية المقترحة على مجموعتي بياناتٍ؛ إذ تمّ الحصول على دقّةٍ وصلت إلى 87.5% باستخدام الانحدار اللّوجستي، بينما بلغت الدّقّة 93% باستخدام آلة متّجهات الدّعم (SVM) للتّصنيف. وعند تناول احتمالية تحسين الأداء لقناةٍ مفردة توضع على جبهة الشّخص المفحوص، وكانت النتائج واعِدة.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 02, June 2024.

DDoS ATTACK-DETECTION APPROACH BASED ON ENSEMBLE MODELS USING SPARK

Yasmeen Alslman, Ashwaq Khalil, Remah Younisse, Eman Alnagi,
Jaafer Al-Saraireh and Rawan Ghnemat

(Received: 17-Sep.-2023, Revised: 2-Dec.-2023 and 7-Feb.-2024, Accepted: 20-Feb.-2024)

ABSTRACT

We live in an era where time is the most precious resource. Thus, dealing with the vast amount of data collected from different resources for various purposes requires creating systems that can process the data correctly to make it worthwhile. Using big data in machine-learning (ML) and artificial-intelligence (AI) models enhances the efficiency and robustness of such models. This work proposes a DDoS attack detection model using Apache-spark to deal with the CIC-DDOS2019 dataset, a significant public dataset used to train this model. The model is trained to predict the type of DDoS attack among multiclass attacks: SYN, UDP and MSSQL. Two state-of-the-art algorithms, Random Forest (RF) and eXtreme Gradient Boosting (XGBoost), have been chosen as the base of our proposed model. These two algorithms inherit their robustness and efficiency from the ensemble nature of their architecture, where each is constructed of several decision trees with different parameters. To contribute to this work, a stacked ensemble model has been built using both RF and XGBoost to enhance the accuracy of the DDoS attack-detection task. It has been found that using such a combination guarantees the best results. The prolonged execution time that resulted from training such a large dataset, on the other hand, is another issue that should be handled. To tackle the speed problem, the Apache-spark platform has been used. Apache-spark divides the large dataset, distributes the divisions and trains them in parallel using the proposed model. Thus, it enhances the execution time while preserving the accuracy of training the same dataset without Apache-Spark. The proposed model has achieved a high accuracy of (99.94%) while reducing the execution time to almost half of the time when applied without Apache-spark. Using Apache-Spark increases the demand on RAMs; using Spark to build the proposed DDoS attack-detection model urged us to improve the hardware used to run the code on Spark. Other relevant research works focus on accuracy measures and need more suitable time analysis, which is crucial in DDoS attack-detection applications; some other models provide less accuracy than the accuracy provided in this study.

KEYWORDS

Ensemble model, Random forest (RF), XGBoost (XGB), Apache-spark, PySpark, Big data, CIC-DDoS2019, DDoS attacks

1. INTRODUCTION

The term big data emerged in 2011. Over the years, it has been massively used in industry, media, commerce and research [1]. Big data is being created rapidly, and different types of data are generated and used: tabular data, text, images, videos, and others.

Machine learning (ML) is a science that is proposed to handle, analyze, and study data [2]. Classification, regression, and clustering are examples of ML tasks that are conducted on data. The bigger the data used to train such models, the better their yield. Nevertheless, feeding extensive (big) data into such models would increase the computation time needed. So, the state-of-the-art machine-learning models are not adequate anymore to deal with big data [3], and thus, researchers are stuck in the dilemma of the compromise between accuracy and performance.

Unique platforms, such as Apache-Spark, have been proposed in recent years to deal with big data in a way that preserves the accuracy as much as possible and, on the other hand, increases the model's performance [4]. Apache-Spark is a unified platform for large-scale data analytics, usually used with machine-learning models to distribute the data load on multiple machines running the same models. It is designed to enhance the scalability and computational speed needed for big data. The architecture of Apache-Spark is a hierarchical master/slave one, where a master node, running the cluster manager, is used to manage the distribution of data between the slave nodes, thus aggregating the results from

Y. Alslman, A. Khalil, R. Younisse, E. Alnagi, J. Al-Saraireh and R. Ghnemat are with Department of Computer Science, Princess Sumaya University for Technology, Amman, Jordan. Emails: yas20219006@std.psut.edu.jo, ash20219002@std.psut.edu.jo, r.baniyounisse@psut.edu.jo, ema20219005@std.psut.edu.jo, j.saraireh@psut.edu.jo and r.ghnemat@psut.edu.jo

these nodes. Such data splitting may compromise the accuracy of models conducted on this platform. Thus, further enhancement of the ML models is required.

An ensemble model is one proposed solution that enhances the results of several ML models working together, where the best of each is highlighted. Several methods are applied to merge the results of several ML models, such as averaging, voting, maximizing, and others. The literature review summarizes and discusses several papers that applied such models. Ensemble learning was mentioned in [5] as a powerful tool for intrusion-detection AI applications.

Distributed Denial of Service (DDoS) attacks are critical network attacks that can harm the whole network system if applied by attackers. The authors of [6] have created a dataset named CIC-DDoS2019 that resembles a real-time network flow during several DDoS attacks. Figure 1 illustrates the classifications of DDoS attacks tackled and studied in [5] using their generated dataset. DDoS attacks are widely studied and analyzed using different AI methods and scenarios. For example, the study in [7] has addressed the attacks in the context of IPv6 datasets and created a dataset containing the traffic information when a DDoS attack is applied on an IPv6 network.

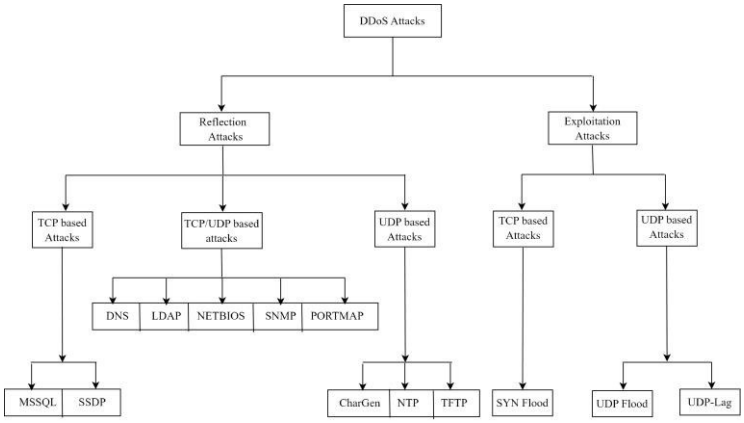


Figure 1. DDoS attack classifications [5].

The proposed work here investigates the effectiveness of ensemble DDoS attack-detection models by analyzing the accuracy, the F1-score and the training of XGBoost and RF models. The work then expands to using Apache-Spark to distribute the used dataset on different slaves to speed up the training process. The time reduction is then analyzed to detect how efficiently Apache-Spark can reduce the training time to build highly robust models.

An ensemble of ensemble models is proposed using Apache-Spark. Using an ensemble of ensemble models is often referred to as a stacking ensemble, an ML technique where multiple ensemble models are combined to make predictions. The essential advantage of using such a model is improving the performance of ML algorithms. In addition, the diversity of base ensemble models contributes to the model's robustness, mitigating the risk of overfitting, particularly in scenarios where individual models may underperform on specific subsets of data. Moreover, using stacked ensemble models can enhance adaptability, as it allows for incorporating various models. Figure 2 illustrates the proposed model architecture. Building the proposed model using Apache-Spark enables the efficient handling of large datasets. In other words, the proposed model combines the strengths of RF and XGBoost with Spark's distributed processing capabilities.

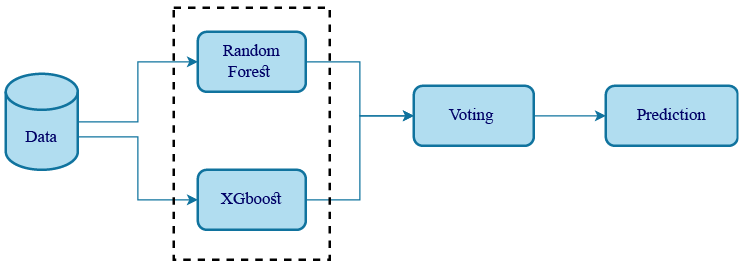


Figure 2. Proposed-model architecture.

The main contributions of this paper can be described in the following points:

1) Using a stacked ensemble model to detect DDoS attacks with high accuracy. The stacked ensemble model used is based on machine-learning algorithms (Random Forest and eXtreme Gradient Boosting), which are also considered ensemble algorithms of their own.
2) Reducing the time needed in the classification process by utilizing Apache-Spark to distribute the big dataset on several slaves to perform the targeted task.
3) The proposed model detects specific DDoS attacks, namely, DDoS-UDP, DDoS-MSSQL, and DDoS-SYN.
4) Due to the short training time needed to build the proposed model, the model can be trained and used to detect new or unknown attacks quickly compared to other approaches.

The rest of this paper is organized as follows: the relevant literature is surveyed and summarized in Section 2, and the research methodology is described in detail in Section 3. Then, in Section 4, the experiments that were conducted are discussed along with their results and evaluation. Finally, the last section presents a conclusion and avenues for future work.

2. LITERATURE REVIEW

The literature has extensively used machine-learning algorithms to perform classification and detection tasks in different research scopes. Random forest and XGBoost are standard algorithms that yield higher accuracy than others. This section concentrates on literature that uses either of these two algorithms separately or in ensemble models. The scope of the selected research has been concentrated on security, especially DDoS attack detection since it is the main scope of the current research.

Public datasets that resemble DDoS attacks are available and used in literature. CIC_DoS and CIC_IDS are noticeably used as valid and big datasets with their versions 2017, 2018, and 2019. Authors of [19][24][28][33] have used the 2017 version, while authors of [8] have used the 2018 version. As for the 2019 version, researchers in [13][28][30] used it along with other datasets. Other public datasets, such as the UNWS_np-15 [18] and NSL_KDD [24], have also been used as training datasets for DDoS attack detection.

Machine learning (ML) and deep learning (DL) models have been used to train these datasets to tackle the problem of DDoS attack detection. Several ML algorithms, such as Naive Bayes (NB) [13][20][29], Decision Tree (DT) [8][13][16][19][20][29], Random Forest (RF) [8][13], [17]-[20], [24], [27]-[29], K-Nearest Neighbors (KNN) [13][19][28], Support Vector Machine (SVM) [13][17], Gradient Boosting (GB) [16][21], Extreme Gradient Boosting (XGB) [13]-[14], [18]-[19], among others, are extensively used in literature. Neural Networks (NN) [22][27][29] and Long Short-term Memory (LSTM) [30]-[31] are examples of DL models that were also used in literature.

Some works have proposed ensemble models using various algorithms, either ML or DL [3][28]. Apache-Spark has been noticeably used in literature as an efficient way to split large datasets and apply parallel training [3], [20]-[22], [24]-[27], [29].

The following sub-sections will categorize literature according to the algorithms used for DDoS attack detection. So, the ones that focused on ML models are discussed in sub-section 2.1, those that used DL models in sub-section 2.2, and finally, those that used spark and/or ensemble models are described in sub-section 2.3. A final sub-subsection, 2.4, is added to address the research gap and limitations in previous literature.

2.1 Literature That Used Machine-learning Algorithms

Starting with the work proposed in [8], the ML models were trained with CIC-DDOS2018 and tested with CC-DDOS2018 to investigate how training the model with specific data and testing the model with another data version can affect the model's accuracy. The Decision Tree and the Random Forest model yielded testing accuracy results over 99.7%. Evaluating ML models with the CIC-DDOS datasets was presented in different works, such as [9]-[11]. Only some works focused on applying big-data approaches for the Intrusion Detection System (IDS) models to speed up the training process so they can be used and re-trained easily during a crisis. On the other hand, the work in [12] emphasized the usefulness of using Apache-Spark with the CIC-DDOS datasets and showed how it can reduce the training time of different models.

XGBoost with DDoS attacks has been discussed in [14], but with software-defined networks; the work presented an adaptive bandwidth profile-based threshold for attack detection and packet drop ratio reduction. It also presented a trigger-based detection and classification method that efficiently uses the XGBoost algorithm for traffic classification into normal or abnormal. Meanwhile, in [13], the CIC-DDoS 2019 dataset was evaluated using different machine-learning models: Naïve Bayes (NB), k-Nearest Neighbors (k-NN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF) and Extreme Gradient Boosting (XGBoost). The XGBoost achieved higher performance than the other used ML models, while the NB model achieved minimal performance.

On the other hand, the XGBoost model was integrated with the adaptive bandwidth profile-based threshold [13] to detect attacks while minimizing the packet drop ratio. Hyperbolic functions, such as the fuzzy function and entropy, are used in the Euclidean distance-based multi-scale fuzzy entropy (EDM-Fuzzy) model in [15] to find the similarities between two vectors and handle the issues of the Heaviside function based on comparison between vectors. However, these models are not considered distributed systems. In contrast, the FEDFOREST [16] combines federated learning and Gradient Boosting Decision Tree (GBDT), an efficient framework for attack detection while achieving privacy.

In [17], the authors proposed a hybrid machine-learning system consisting of two algorithms: Support Vector Classifier (SVC) and Random Forest. Their goal was to detect benign traffic from DDoS attacks. In their work, they built the model to enhance the classification output from SVC, with further training using Random Forest. When using SVC, some instances may not be precise enough that they belong to a specific class; they may reside on the line that separates these classes and thus, the probability of belonging to the classes is equal. Here comes the role of the Random Forest classifier to give more precise decisions about the questionable instances. The authors created their dataset within a Software-defined Network (SDN) by monitoring the flow of incoming and outgoing switches in the network and gathering statistical information about the flow and ports in a CSV file. They have generated a large dataset of about 100,000 records and 23 features. They have compared their results with those of other machine-learning models and theirs have outperformed the others with an accuracy of 98.8%.

In [18], the authors also tackled the problem of DDoS attack classification. They have applied their experiments on the UNWS-np-15 dataset, with about 80,000 records and 45 features. Their proposed model used both the Random Forest classifier and the XGBoost classifier separately. The chosen classifiers yielded 89% and 90% accuracy, respectively. They have compared their results with those of other algorithms applied to different datasets. Their results have outperformed the ones working on the same dataset, UNWS-np-15.

As for [19], the authors examined several machine-learning algorithms to detect DoS and DDoS attacks using the CIC-IDS-2017 dataset. Decision Tree (DT), Random Forest (RF), K-Nearest Neighbors (KNN) and XGBoost (XGB) have been selected for their experiments. They evaluated the four algorithms using precision and recall as the primary evaluation metrics and they found that all four algorithms yielded high Recall, over 99%, with slight differences between them. The maximum Recall was achieved by XGB (99.87%), while KNN had the minimum Recall (99.52%). XGB, on the other hand, yielded the worst precision of 97.6%, while RF achieved 99.76% precision.

Enhancing the performance of the DDoS detection (normal and abnormal) system by reducing the misclassification error was the primary motivation for Alduailij et al. [28]. Two datasets were used in the proposed model (CICIDS2017 and CICDoS2019) for feature selection, mutual information and random forest feature importance. These features have been used in different machine-learning algorithms (Linear Regression LR, Random Forest RF, k-Nearest Neighbor KNN and Weighted Voting Ensemble WVE). It has been concluded that RF achieved the highest accuracy in all cases (when the number of the selected features was 19, 16, and 23).

2.2 Literature That Used Deep-learning Algorithms

A distributed learning environment deploying a neural network learning model was used [22]. The work distribution was achieved by using Spark. The specialty that [22] has gained is that the authors used their DDoS detection model to analyze the live traffic over the network and obtain the results, where they gained an accuracy result of 94%.

Deep-learning algorithms are also used in literature for DDoS attack detection. The authors of [30] have used a hybrid system of both BI-LSTM (Bi-directional) and Gaussian Mixture Model (GMM) to detect DDoS attacks using two datasets, CIC-IDS2017 and CIC-DDoS2019. Their experiments yielded up to a 94% accuracy score. As for [31], the authors have also used BI-LSTM and CNN to conform to a hybrid system of two deep-learning algorithms. They applied training and testing using the CIC-DDoS2019 dataset and yielded an accuracy of up to 94.52%.

2.3 Literature That Used Apache-Spark and Ensemble Models

In [20], the authors proposed a distributed system that takes advantage of the distribution to overcome the latency of multiple classification models. The time factor is an essential measure reflecting the strength of DDoS detection systems; the later the attack is discovered, the worse the consequences are. The classification models are Random Forest, Naive Bayes and Decision Tree. Spark was used to run the three models in parallel to fasten the flow packet classification process with the aim of DDoS attack detection. In contrast, fuzzy-logic rules were used to direct the packets to an algorithm based on traffic.

In the context of DDoS attacks, the Gradient boosting algorithm performance was taken to a new level [21] by parallelizing the gradient boosting algorithm, which was applied to classify packets based on potential DDoS attack threats carried with them. At the same time, the gradient boosting algorithm was supposed to build the machine-learning model out of multiple smaller decision trees and pass the packets through the tree branches to perform the classification process. Spark was used to accomplish the task in parallel, aiming to speed up the operation. The gradient boosting algorithm has enhanced the performance of the classification model. At the same time, Spark has distributed the operation over three slave machines and one master-machine architecture to enhance the time factor.

An unsupervised machine-learning algorithm was developed by [23] to create an analysis system to detect possible DDoS attacks over a vast amount of traffic. Through the analysis of 14 PCAP files recording the traffic on a network under a DDoS attack, the study aimed to reduce the required time to train and run the model for detection.

In [3], the authors compared the performance of two ensemble-learning models in the Spark environment. Through the study, they used the CIDDS dataset, which is used to train machine-learning models of intrusion-detection systems. The two models that were compared are the logistic regression-based blending ensemble and SVM-based blending ensemble and the study concluded that the latter outperformed logistic regression in terms of accuracy by 5%. The SVM-based blending ensemble model accuracy was recorded at 95%.

Distributed machine learning was used in [24] to detect the presence of concept drift in network traffic and detect network-based attacks. Spark archived the distribution. The study uses K-means clustering for detecting drift happening to the network traffic. Random Forest and Linear Regression models were used for intrusion detection. The datasets used in the study are the NSL-KDD dataset, the CIDDS-2017 dataset, and the generated Testbed dataset.

A real-time detection of DDoS attacks using machine-learning classifiers on a distributed-processing platform was proposed in [25]. The authors generated a traffic-simulating DDoS attack and fed the data features into different classifiers distributed over multiple Spark slave machines.

Online distributed denial-of-service attack detection using Spark streaming was studied in [26] and [27]. In the latter, the authors applied two machine-learning models, RF and MLP, for training and testing. Both models were applied with and without big-data approaches. Apache-Spark was deployed for the big-data approach.

Patil et al. [29] proposed a real-time network flow classification using a novel Spark streaming and Kafka-based classification system. The proposed model successfully classified the network flow into seven categories (Normal, DDoS-DNS, DDoS-LDAP, DDoS-MSSQL, DDoS-UDP, DDoS-SYN and DDoS-NetBIOS).

The CICDDoS dataset was used in training the model using four different machine-learning algorithms (RF, MLP, DT and NB), with RF being the best method in classifying the network flow with an accuracy of 89%. Table 1 displays a summary of most related literature reviewed.

Table 1. Literature-review summary.

Ref.#	Model	Advantages	Limitations
Manickam et al., 2022 [6]	Decision Tree, KNN, SVM, Naïve Bayes and CNN.	The evaluation of DDoS attack detection in the context of IPV6 networks.	The accuracy was not within satisfying levels.
De Araujo et al., 2021[12]	XGBoost.	Presented an efficient feature-selection method with the XGBoost model when used for DDoS attack detection applications.	Multi-class classifier accuracy was low.
Alamri et al., 2021 [14]	Several traditional machine-learning models: XGBoost, NB, k-NN, SVM, DT and RF.	Utilizing different types of machine-learning techniques to evaluate the performance of DDOS detection.	The authors did not consider the ensemble, distributed system. Training overhead time was mentioned.
Alamri et al., 2020 [13]	Bandwidth-control mechanism and XGBoost model.	High accuracy.	Using a bandwidth-control mechanism reduces the accuracy of the multi-classification task using XGBoost compared with previous works. Furthermore, the distributed system was not considered.
Zhou et al., 2021 [15]	Euclidean Distance-based Multi-scale Fuzzy Entropy (EDM-Fuzzy).	Achieved training stability by handling the traditional distance-computation issues, such as bouncing between 0 and 1 produced by the Heaviside function.	The authors did not use distributed systems and self-learning classifiers, such as deep learning to train the large dataset.
Dong et al., 2022 [16]	Federated Learning and Gradient Boosting Decision Tree.	Achieved data privacy and utilized a distributed system.	It is sensitive to hyper-parameter tuning.
Ahuja et al., 2021 [17]	Hybrid system consisting of SVC and RF.	Created their dataset or about 100,000 records and 23 features. They have enhanced the results of SVC by using RF as a first decision layer.	The authors have only tested their dataset over binary classes to detect benign traffic.
Mohmand et al., 2022 [18]	RF and XGBoost.	Used UNWS-np-15 Dataset with 80,000 records and 45 features.	Although the results have outperformed those of others who used the same datasets, but the percentage of 90% is still low compared with other research in the same scope.
Zewdie and Girma, 2022 [19]	DT, RF, KNN and XG-Boost.	Used CIC-IDS-2017 dataset, each algorithm has yielded a high Recall score separately.	Some algorithms have yielded better Recall than others; thus, merging them into an ensemble model could yield higher scores.
Alsirhani et al., 2018 [20]	RF, NB and DT.	Using Spark to run the three models in parallel, a fast classification process, using Fuzzy logic to decide the best algorithm to be used.	There is a trade-off between accuracy and speed, which lowers the accuracy of the models.
Alsirhani et al., 2018 [21]	Gradient Boosting Algorithm GBT.	Using Spark to enhance the performance of the model, two datasets are used for experiments.	Datasets used contains only two classes.
Hsieh and Chan, 2016 [22]	Neural Network.	Analysis of live traffic over the network using Spark.	Accuracy is comparatively low to other research 49%
Jain and Kaur, 2021 [24]	RF, LR and SVM.	Two datasets are used to generate Testbed dataset.	97% accuracy for the ensemble model.
Alduailij et al., 2022 [28]	Several machine-learning algorithms: LR, RF, WVE and KNN.	High accuracy with a minimum number of features.	Using the proposed model as a detection model (normal and abnormal) rather than classification.
Patil et al., 2022 [29]	Using four different machine-learning algorithms (RF, MLP, DT and NB).	Provided a real-time net-workflow classification using a novel Spark streaming and Kafka-based classification system.	Low accuracy (89%) when comparing it to the literature.
Chartuni et al., 2021[32]	Multi-class CNN classifier composed of seven layers.	Multi-class classifier with high accuracy.	The model was not evaluated with a computer network the flow of which was not previously seen by the model.

In [33], the authors investigated DDoS attack detection in the context of Internet of Things (IoT). The work uses an ensemble model of different ML models to enhance the detectability of the attacks.

In the work presented in [34], the authors highlighted the importance of using deep-learning models in IDSs and suggested using Apache-Kafka stream and distributing the data on multiple workstations. The work analysis is conducted on online streamed data. It shows that the proposed technique surpasses the baseline strategies by 11% in the F1-measure when the number of workers is two and by 25% when the number of workers is equal to 32.

In their research paper, the authors of [35] developed a new model named Stacked Convolutional Neural Network and Bidirectional Long Short-Term Memory (SCNN-Bi-LSTM) for the purpose of intrusion detection in wireless sensor networks. The work utilizes Spark to distribute the workload among multiple nodes. The model was able to achieve an impressive classification accuracy of 99.9%. However, the paper does not delve into the details of how the use of Spark has improved the model's accuracy or the time required for the work.

2.4 Research Gaps and Limitations

The previously discussed literature that focused on creating IDSs against DDoS attacks did consider time. Many works presented binary IDS models rather than multi-class models. Additionally, the multi-class classification models achieved lower accuracy than the accuracy achieved in this work.

Hence, in this work, we focus on building an ensemble model of two ensemble models, the Random Forest and the XGBoost, to prepare an intrusion-detection system that can efficiently operate with big datasets in a reasonable amount of time. No feature selection was used on the datasets, since there is no need for further pre-processing techniques as long as the proposed model accuracy is already sufficiently high. Compared to other works, this work can be considered a simple approach to building an efficient ensemble multi-class DDoS attack-detection model, which requires simple pre-processing steps and provides the classification results in an easily understandable form. The presented model can easily be used with similar datasets concerning the multi-class classification in the context of DDoS attacks. The methodology followed in this work is discussed in the next section.

Most of the mentioned works focus on analyzing the accuracy results or the false-positive measures of the proposed DDoS attack-detection models while the required time was not analyzed or mentioned. In this work we believe that decreasing the time required for detecting such attacks is very critical, since these attacks are based on paralyzing network systems by establishing unnecessary communications with the systems. Hence, detecting these attacks early makes them much less harmful. The accuracy is critical, yes, but with a very large dataset such as CIC-DDoS2019, building an accurate detection system is not as challenging as decreasing the detection time.

3. METHODOLOGY

This section discusses the methodology used to build the proposed model. It starts from the dataset selection, goes through its pre-processing and ends with implementing the proposed model using Apache-Spark. This paper proposes an ensemble model of random forest and XGB regressor for multiclass classification using Apache-Spark.

As mentioned before, the main reason for choosing such algorithms is to build a stacked ensemble model, which is an ensemble model built using ensemble-based models, such as random forest and XGB regressor.

Spark typically results in enhanced performance, scalability and the ability to handle larger datasets efficiently. Its impact on the model's development and deployment can be significant, providing a competitive edge in handling big data and complex analytics tasks. Spark's ability to distribute data processing across a cluster of machines is crucial for handling large datasets. Its in-memory processing capability accelerates computations. Spark was also used due to its adept at stream processing *via* Spark streaming, enabling real-time analytics on data streams. Furthermore, Spark supports various programming languages (like Scala, Java and Python), making it accessible and convenient. Figure 2 shows the proposed stacked ensemble model used on the CIC-DDOS2019 dataset.

It can be noticed that after selecting that data, it will be passed to two ensemble models; namely,

random forest and XGB regressor. Afterwards, the resulting predictions from both models will be passed to a voting mechanism, which in turn produces the final result. The steps below summarize the basic steps of building the proposed model.

1. Read the dataset into a DataFrame df_pyspark.
2. Pre-process the dataset.
3. Split data into training and testing subsets using the specified train_test_split ratio.
4. Train a Random Forest classifier model on the training subset.
5. Train the XGB regressor classifier model on the training subset.
6. Evaluate the Random Forest classifier model on the testing subset.
7. Evaluate the XGB regressor classifier model on the testing subset.
8. Combine predictions from both models to form the ensemble prediction.
9. Evaluate the final result using the evaluation metrics.

3.1 Dataset Selection

The Distributed Denial of Service (DDoS) attack uses malicious traffic to exhaust the target networks, where the CIC-DDoS2019 dataset is the latest released version of the DDoS datasets. Each record consists of features indicating the traffic status, whether it is an attack or not. The CIC-DDoS2019 collects many malicious and normal traffic cases collected in two days. In 2019, the Canadian Institute for Cybersecurity (CIC) at the University of New Brunswick created this dataset. The dataset was created to have a more professionally engineered and diversified set of DDoS detection attacks. These attacks are used to evaluate the proposed ensemble model. In the SYN attack, the attackers aim to send a large number of SYN (synchronization) requests; TCP packets are used to connect to a server to overwhelm it with an open connection or a half-open connection, which aims to overwhelm the targeted server with fake connections. On the other hand, the UDP attack overwhelms random ports on the targeted host with IP packets containing UDP datagrams, making the targeted server unable to process the flood of arriving packets and serve legitimate users. MSSQL is a web-application attack that uses a bad application design that does not sanitize inputs, exposing application vulnerabilities.

The dataset is one of the most comprehensive and used datasets in the scope of building DDoS attack-detection models. The main limitation in this dataset is the existence of NaN and infinity values among many tuples. This problem was solved in this work by replacing these values with zeros. The other limitation is its big size, which means that building an IDS system using the dataset will require a long time. This limitation is solved by using Spark to reduce the training and testing times of the proposed IDS system.

The dataset has been organized in several CSV files that consist of millions of records that present different types of attacks. As shown in Figure 3, two files have been selected: Syn-file and UDP-file, which consist of about eight million records, considered significant for the proposed model. The files for these attacks were chosen to be analyzed through this study, because the study focuses on multi-class-classification problems when massive datasets are used. These attacks have around eight million records and they are famous, repeatedly discussed in the literature and can cause serious harm to the attacked systems.

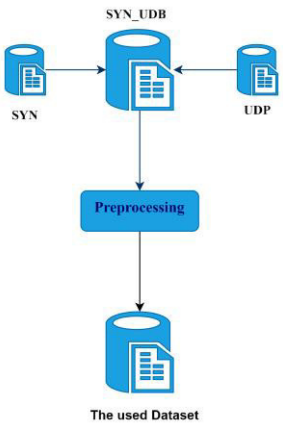


Figure 3. The data-selection method.

This dataset is chosen for this research, because it is one of the most recent and essential datasets in the IDS field. It is widely used in other studies; hence, the comparison can be valid. Moreover, since the research focuses on using IDS with large datasets, this dataset is a convenient choice.

3.2 Random Forest

The first supervised machine-learning technique used in the proposed ensemble model is the random forest (RF). The RF is a robust machine-learning model that reduces overfitting and performs efficiently on large datasets. The RF randomly divides the dataset's features into sub-sets of features, where each sub-set is trained using the decision-tree model separately and independently of other sub-trees. In the training process, each dataset sample is trained in a particular sub-tree, while in the testing, the entire test data is trained in each sub-tree. The final result is aggregated by producing an average of these results. The decision-tree model consists of nodes and branches. At each node, evaluate the sub-set of features to generate and divide the observations into other nodes in the training set or to flow to a specific path when making a prediction.

3.3 eXtreme Gradient Boosting (XGBoost)

The second robust algorithm that handles the bias-variance trade-offs and provides a parallel tree boosting for classification is the XGBoost model. Like the RF algorithm, the XGBoost model uses gradient boosting, providing adaptation and generalization. It is considered an ensemble of multiple learners, such as decision trees, where the final decision is an ensemble of the subtrees' outputs. Consequently, the XGBoost prevents poor performance. XGBoost uses the gradient descent algorithm to optimize the model by updating weight and reducing cost value and the discrepancy between the expected and actual values. The mean squared error (MSE) cost function is used as an evaluation metric for classification tasks.

In the next section, the experiments are discussed by starting with the data pre-processing phase, then concluding the results of experiments with and without Spark, along with an evaluation and discussion of these results.

4. EXPERIMENT RESULTS

4.1 Data Pre-processing and Experiment Setup

Data pre-processing is crucial in the data analysis and machine-learning pipeline. It contributes to data quality, integrity and suitability for modeling, leading to more accurate and reliable results.

As aforementioned, the CSV files selected from the CIC-DDoS2019 dataset for evaluation in the proposed model were UDP.csv and SYN.csv. The UDP file consists of three classes: UDP attack, MSSQL attack and benign. At the same time, the SYN file consists of two classes: SYN attack and benign. These two files are under exploitation attacks, as shown in Figure 1. Table 2 represents the number of samples for each class before and after the pre-processing phase. The data pre-possessing consists of the following steps:

1) Eliminating duplicate records using drop_doplicate function provided by Python, where the number of samples for SYN and Benign classes is reduced from 4,284,751 to 3,806,356 and from 35,790 to 31,386, respectively, as shown in Table 2. However, there were no duplicate records for UDP and MSSQL. This step is essential, as it helps maintain data quality and consistency. Duplicate records can skew analysis results, leading to biased models or overfitting problems.
2) Eliminating attributes with summation and variances of zero can lead to more efficient, generalizable models and it is a common data pre-processing step. These attributes are: Bwd Packet Length Std, Bwd PSH Flags, Fwd URG Flags, Bwd URG Flags, FIN Flag Count, PSH Flag Count, ECE Flag Count, Fwd Avg Bytes/Bulk, Fwd Avg Packets/Bulk, Fwd Avg Bulk Rate, Bwd Avg Bytes/Bulk, Bwd Avg Packets/Bulk, Bwd Avg Bulk Rate and Similar HTTP. This step reduces the number of features from 88 to 74.
3) Replacing infinity and null values with zeros using replace(nan, 0) function provided by numpy module in Python.

4) Encoding categories including source IP, flow ID, destination IP and timestamp using LabelEncoder function provided by sklearn module in Python. This step is crucial, since many machine-learning algorithms work with numerical data; so, categorical variables must be encoded into a suitable numerical format.

Table 2. The number of records for each class.

Class	Samples count before pre-possessing	Samples count after pre-possessing
UDP attack	3754680	3754680
MSSQL attack	24392	24392
SYN attack	4284751	3806356
Benign (no attack)	38924	34520

The dataset was split randomly into 80% for training and 20% for testing. Consequently, the number of records in training and testing sets are 6095958 and 1523990, respectively. Furthermore, the default parameters of the Random Forest and XGBoost models were determined as follows: The number of sub-trees of both models is 100 and the weak learner in the XGBoost is the decision trees. Table 3 and Table 4 show the hyper-parameters of RF and XGBoost used to train the model.

Table 3. Random forest hyper-parameters.

RF hyper-parameter	Value
Number of trees.	100
Function to measure the quality of a split.	gini
The minimum number of samples required to split an internal node.	2
The minimum number of samples required to be at a leaf node.	1

Table 4. XGBoost hyper-parameter.

XGboost hyper-parameter	Value
Number of boosting rounds.	100
Loss function	Squared error
Boosting model (weak learner)	Decision trees (gbtree)
The feature-importance type	Information gain
Maximum depth	6

4.2 Performance Metrics

To evaluate the proposed model, four metrics were utilized: accuracy, precision, recall and time. The assessment was conducted both with and without the Spark platform. Accuracy indicates the percentage of correctly classified instances out of the total number of cases evaluated, as per Equation 1, while time denotes the training time in minutes. Meanwhile, the ensemble model employing Apache-Spark was evaluated through a classification report that includes precision, recall and F1-Score. The formula for calculating the measures are presented in Equations 1-3.

$$\text{Accuracy} = \text{TP+TN/(TP+TN+FP+FN)} \quad (1)$$

$$\text{Precision} = \text{TP/(TP+FP)} \quad (2)$$

$$\text{Recall} = \text{Precision.Recall/Precision+Recall} \quad (3)$$

TN stands for True Negative, FP for False Positive, TP for True Positive and FN for False Negative. These measures were utilized due to data imbalance, with F1-Score being commonly used to evaluate IDSs, as it combines precision and recall and provides valuable insights into the study outcomes.

The value of each measurement ranges from 0 to 1, with higher values indicating a more robust model that is better suited for detecting possible intrusions. Time measures were utilized to compare the performance of the proposed model with and without Spark, highlighting the value added by Spark usage. With Spark, the required times for building and training the model, as well as for running the detection model, are significantly reduced.

4.3 Experiments and Results

Two experiments have been conducted. Apache-Spark's distributed computing capabilities make it a valuable tool for handling large datasets. Its scalability, in-memory processing and distributed model enable it to process vast amounts of data efficiently. However, one of its potential limitations is the resource requirements. As a result, both experiments have been implemented on Colab Pro with (25 GB) RAM; the first experiment was done without Apache-Spark by applying the RF and XGBoost ensemble model.

The same model with the same environment has been applied again but with the use of Apache-Spark using the PySpark library, considered an open-source interface for Apache-Spark. It allows SQL-like analysis on large amounts of structured or semi-structured data. Figure 4 and Table 5 illustrate the training time needed to conduct Random Forest and XGBoost separately, once under the Apache-Spark environment and the other without using it.

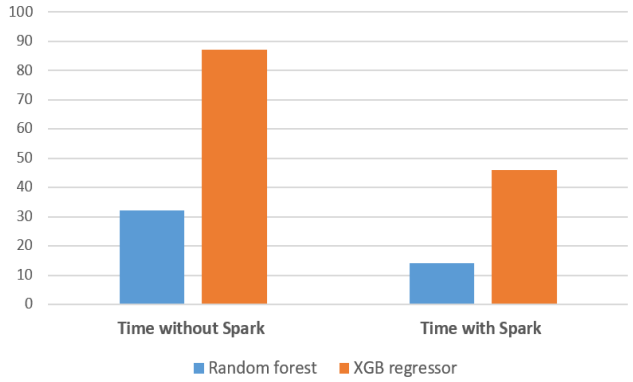


Figure 4. The training time required to train random forest and XGBoost models with/without Spark.

Table 5. The training time using the CIC-DDOS 2019 dataset.

Model	Training time without Spark in minutes	Training time with Spark in minutes
Random forest	32	14
XGB regressor	87	46

Table 6 illustrates the accuracies from conducting Random Forest, XGBoost and the proposed ensemble model, either with Spark or without it. It can be noticed how the results are close in both cases. As for Figure 5, the Recall, precision and F1-score measures are compared for classes SYN, MSSQL, UDP and Normal when conducting the proposed ensemble model.

Table 6. The accuracy using the CIC-DDOS 2019 dataset.

Model	Accuracy without Spark (%)	Accuracy with Spark (%)
Random forest	99.9995	99.9155
XGB regressor	99.9762	99.9942
Ensemble	99.94	99.9419

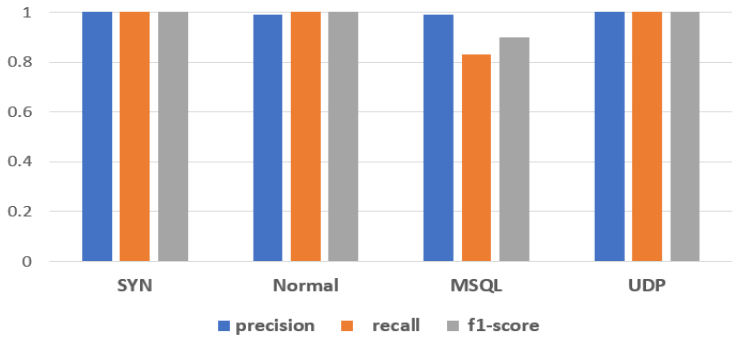


Figure 5. The performance of the proposed ensemble model for each class.

Table 7 conducts a comparison between selected proposed models in the literature. The comparison has been conducted regarding accuracy, F1-score, number of classes predicted and whether Spark is used or not. In reference [29], the F1-measure was calculated by averaging the F1-scores for the seven classes mentioned in that work.

Table 7. Comparison between the proposed model and literature.

Model	Accuracy (%)	F1-score	No. of classes	Spark
Extreme Gradient Boosting (XGBoost) [12]	99.7	99.79	two classes	-
XGBoost [13]	99.7	100	two classes	-
XGBoost [13]	91.26	92	multi-class	-
FEDFOREST (L+L) [16]	67.03	94.60	two classes	-
Random forest [28]	99.99	-	two classes	-
Parallel NB, DT and RF [20]	61.4, 97.9 and 97.3	51.3, 97.9 and 9.73	two classes	✓
Random forest [29]	89	89	seven classes	✓
CNN [32]	94.21	94.12	seven classes	-
Proposed (ensemble -RF and XGBoost-)	99.94	97.5	four classes	✓

The code that we have written and used throughout this study is publicly available in [36].

4.4 Discussion

The accuracy of the proposed model has been one of many concerns during the extraction of experiment results. The training time has been considered one vital criterion to concentrate on to prove our work's high performance and validity.

Table 3 and Figure 4 show that the time needed to train either a Random Forest or an XGBoost model has been reduced to about a half when using Spark. The required time to train Random Forest and XGBoost using Spark is reduced by 18 and 41 minutes, respectively. This time reduction makes the model more efficient, since detecting DDoS attacks is usually deployed in sensitive applications where time matters. Expanding the required time to train and use the model makes it less reliable and less usable in such applications.

From Table 4, it can be noted that the proposed ensemble model achieved a high accuracy regardless of whether or not Spark was used. Table 4 proves that even splitting the dataset in a way that Spark can use did not affect the accuracy of the proposed model. Thus, the model's performance has been enhanced while preserving its accuracy.

Figure 5 illustrates that both RF and XGBoost, working as a stacked ensemble model, can successfully distinguish the SYN, UDP, MSQL and benign. However, Recall and F1-score have been slightly reduced when using Spark in MSSQL attack. This is because of the number of its samples, which is considered small compared with the rest of the classes, as represented in Table 2. Thus, splitting and training small samples may yield less F1-score for such classes.

The proposed model achieves the best accuracy when compared with other models in the literature. Table 5 shows that the proposed model exceeds the existing models in terms of accuracy when the model is trained in more than two classes. It also outperformed the accuracy of models that classified only two classes, considering that binary classification is usually more accessible than multi-class classification.

Overall, the combination of stacked ensemble models under the Apache-Spark environment outperformed other models described in the literature.

5. CONCLUSION

This work proposed a distributed ML model for DDoS attack detection using Apache-Spark. Ensemble learning was used to build a robust and efficient IDS. The system can detect three DDoS attacks: UDP, MSSQL and SYN. The model comprises two trusted ML models: Random Forest and

XGB regressor. Hence, it can be considered a stacked ensemble model. Apache-Spark was used to train data distribution in parallel using the proposed model. Data distribution using Spark has enhanced the required time to train the model, as the time was reduced to around a half. At the same time, the accuracy was preserved at a level of over 99%. The required training time for the XGBoost regression model was reduced from 87 minutes to 46 minutes when Spark was used. The required training time for the Random Forest model was reduced from 32 to 14 minutes when Spark was used. The time reduction comes at the cost of increasing the used RAMs, which is considered the main limitation of the proposed approach in this work. This limitation can be avoided by using computers with large RAM capacity.

The presented methodology can be considered an efficient IDS approach that can be used with DDoS attacks, such as the attacks the data of which is recorded in the CIC-DDOS2019 dataset. The presented distributed model is also robust and fast. Hence, it can be used when online intrusion-detection models are not affordable, complicated or down. Using an ensemble IDS of XGBoost and Random Forest with Apache-Spark has proved to be an easily built and trained model. The approach guarantees short training time and robustness against failure; when one distributed node in the Apache-Spark platform is down, the other nodes are available and a replacement node can take over the failed one.

In the future, more ML models will be added to the distributed ensemble model for the DDoS IDS by editing the used code and adding more slaves (nodes) to the ensemble model. Furthermore, the work will be expanded to detect other intrusions by training the model with datasets related to other attacks.

REFERENCES

[1] J. S. Ward and A. Barker, "Undefined by Data: A Survey of Big Data Definitions," arXiv preprint, arXiv: 1309.5821, 2013.

[2] M. I. Jordan and T. M. Mitchell, "Machine Learning: Trends, Perspectives and Prospects," Science, vol. 349, no. 6245, pp. 255-260, 2015.

[3] G. Kaur and M. Jain, (2020), "A Comparison of Two Blending-based Ensemble Techniques for Network Anomaly Detection in Spark Distributed Environment," International Journal of Ad Hoc and Ubiquitous Computing, vol. 35, no. 2, pp. 71-83, 2020.

[4] M. Zaharia et al., "Apache Spark: A Unified Engine for Big Data Processing," Communications of the ACM, vol. 59, no. 11, pp. 56-65, 2016.

[5] I. Sharafaldin, A. H. Lashkari, S. Hakak and A. A. Ghorbani, "Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy," Proc. of the 2019 IEEE Int. Carnahan Conf. on Security Technology (ICCST), pp. 1-8, Chennai, India, 2019.

[6] S. Manickam et al., "Labelled Dataset on Distributed Denial-of-Service (DDoS) Attacks Based on Internet Control Message Protocol Version 6 (ICMPv6)," Wireless Communications and Mobile Computing, vol. 2022, Article ID 8060333, DOI: 10.1155/2022/8060333, 2022.

[7] T. H. Chua and I. Salam, "Evaluation of Machine Learning Algorithms in Network-based Intrusion Detection Using Progressive Dataset," Symmetry, vol. 15, no. 6, p. 1251, 2023.

[8] B. I. Farhan and A. D. Jasim, "Performance Analysis of Intrusion Detection for Deep Learning Model Based on CSE-CIC-IDS2018 Dataset," Indonesian Journal of Electrical Engineering and Computer Science, vol. 26, no. 2, pp. 1165-1172, 2022.

[9] A. Elhanashi, K. Gasmi, A. Begni, P. Dini, Q. Zheng and S. Saponara, "Machine Learning Techniques for Anomaly-based Detection System on CSE-CIC-IDS2018 Dataset," Proc. of the Int. Conf. on Applications in Electronics Pervading Industry, Environment and Society (ApplePies 2022), Part of the Lecture Notes in Electrical Engineering Book Series, vol. 1036, pp. 131-140, Springer, 2022.

[10] I. F. Kilincer, F. Ertam and A. Sengur, "A Comprehensive Intrusion Detection Framework Using Boosting Algorithms," Computers and Electrical Engineering, vol. 100, p. 107869, 2022.

[11] R. Atefinia and M. Ahmadi, "Performance Evaluation of Apache Spark MLlib Algorithms on an Intrusion Detection Dataset," arXiv preprint, arXiv: 2212.05269, 2022.

[12] P. H. H. N. de Araujo et al., "Impact of Feature Selection Methods on the Classification of DDoS Attacks using XGBoost," Journal of Communication and Information Systems, vol. 36, no. 1, pp. 200-214, 2021.

[13] H. A. Alamri and V. Thayananthan, "Bandwidth Control Mechanism and Extreme Gradient Boosting Algorithm for Protecting Software-defined Networks against DDoS Attacks," IEEE Access, vol. 8, pp. 194269-194288, 2022.

[14] H. A. Alamri and V. Thayananthan, "Analysis of Machine Learning for Securing Software-defined Networking," Procedia Computer Science, vol. 194, pp. 229-236, 2021.

[15] R. Zhou, X. Wang, J. Yang, W. Zhang and S. Zhang, "Characterizing Network Anomaly Traffic with

Euclidean Distance-based Multiscale Fuzzy Entropy," Security and Communication Networks, vol. 2021, Article ID 5560185, DOI: 10.1155/2021/5560185, 2021.

[16] T. Dong, S. Li, H. Qiu and J. Lu, "An Interpretable Federated Learning-based Network Intrusion Detection Framework," arXiv preprint, arXiv: 2201.03134, 2022.

[17] N. Ahuja, G. Singal, D. Mukhopadhyay and N. Kumar, "Automated DDOS Attack Detection in Software Defined Networking," Journal of Network and Computer Applications, vol. 187, p. 103108, DOI: 10.1016/j.jnca.2021.103108, 2021.

[18] M. I. Mohmand et al., "A Machine Learning-based Classification and Prediction Technique for DDoS Attacks," IEEE Access, vol. 10, pp. 21443-21454, 2022.

[19] T. G. Zewdie and A. Girma, (2022), "An Evaluation Framework for Machine Learning Methods in Detection of DoS and DDoS Intrusion," Proc. of the 2022 IEEE Int. Conf. on Artificial Intelligence in Information and Communication (ICAIIC), pp. 115-121, Jeju Island, Korea, 2022.

[20] A. Alsirhani, S. Sampalli and P. Bodorik, (2018), "DDoS Attack-detection System: Utilizing Classification Algorithms with Apache Spark," Proc. of the 2018 9th IEEE IFIP Int. Conf. on New Technologies, Mobility and Security (NTMS), pp. 1-7, Paris, France, 2018.

[21] A. Alsirhani, S. Sampalli and P. Bodorik, "DDoS Detection System: Utilizing Gradient Boosting Algorithm and Apache Spark," Proc. of the 2018 IEEE Canadian Conf. on Electrical & Computer Engineering (CCECE), pp. 1-6, Quebec, Canada, 2018.

[22] C. J. Hsieh and T. Y. Chan, (2016), "Detection DDoS Attacks Based on Neural-Network Using Apache Spark," Proc. of the 2016 IEEE Int. Conf. on Applied System Innovation (ICASI), pp. 1-4, Okinawa, Japan, 2016.

[23] K. Kato and V. Klyuev, "Development of a Network Intrusion Detection System Using Apache Hadoop and Spark," Proc. of the 2017 IEEE Conf. on Dependable and Secure Computing, pp. 416-423, Taipei, Taiwan, 2017.

[24] M. Jain and G. Kaur, "Distributed Anomaly Detection Using Concept Drift Detection Based Hybrid Ensemble Techniques in Streamed Network Data," Cluster Computing, vol. 24, pp. 2099-2114, 2021.

[25] S. Gumaste, D. G. Narayan, S. Shinde and K. Amit, "Detection of DDoS Attacks in OpenStack-based Private Cloud Using Apache Spark," Journal of Telecommunications and Information Technology, vol. 2020, no. 4, pp. 62-71, 2020.

[26] B. Zhou, J. Li, J. Wu, S. Guo, Y. Gu and Z. Li, "Machine-learning-based Online Distributed Denial-of-Service Attack Detection Using Spark Streaming," Proc. of the 2018 IEEE Int. Conf. on Communications (ICC), pp. 1-6, Kansas City, USA, 2018.

[27] M. J. Awan et al., "Real-time DdoS Attack Detection System Using Big Data Approach," Sustainability, vol. 13, no. 19, p. 10743, 2021.

[28] M. Alduailij, Q. W. Khan, M. Tahir, M. Sardaraz, M. Alduailij and F. Malik, "Machine-learning-based DdoS Attack Detection Using Mutual Information and Random Forest Feature Importance Method," Symmetry, vol. 14, no. 6, p. 1095, 2022.

[29] N. V. Patil, C. R. Krishna and K. Kumar, "SSK-DdoS: Distributed Stream Processing Framework Based Classification System for DdoS Attacks," Cluster Computing, vol. 25, no. 2, pp. 1355-1372, 2022.

[30] C. S. Shieh et al., "Detection of Unknown DdoS Attacks with Deep Learning and Gaussian Mixture Model," Applied Sciences, vol. 11, pp. 11, p. 5213, 2021.

[31] D. Alghazzawi, O. Bamasag, H. Ullah and M. Z. Asghar, "Efficient Detection of DdoS Attacks Using a Hybrid Deep Learning Model with Improved Feature Selection," Applied Sciences, vol. 11, no. 24, p. 11634, 2021.

[32] A. Chartuni and J. Márquez, "Multi-classifier of DdoS Attacks in Computer Networks Built on Neural Networks," Applied Sciences, vol. 11, no. 22, p. 10609, 2021.

[33] Y. Yilmaz and S. Buyrukoglu, "Development and Evaluation of Ensemble Learning Models for Detection of Distributed Denial-of-Service Attacks in Internet of Things," Hittite Journal of Science & Engineering, vol. 9, no. 2, pp. 73-82, 2022.

[34] M. Seydali, F. Khunjush and J. Dogani, "Streaming Traffic Classification: A Hybrid Deep Learning and Big Data Approach," Cluster Computing, DOI: 10.1007/s10586-023-04234-0, 2024.

[35] S. M. S. Bukhari et al., "Secure and Privacy-preserving Intrusion Detection in Wireless Sensor Networks: Federated Learning with SCNN-Bi-LSTM for Enhanced Reliability," Ad Hoc Networks, vol. 155, p. 103407, 2024.

[36] "ColabCode," [Online], Available: Https://Colab.Research.Google.Com/Drive/1oZu2czCK9tJSwcjEfyvLiZnrI0JqYW62?Usp=Sharing, December 22, 2023.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 02, June 2024.

ملخص البحث:

إنّنا نعيش في عَصْرٍ يُعَدّ فيه الوقت المورِد الأغلى. لذا، فإنّ التّعاملَ مع الكمّ الهائل من البيانات التي تُجمّع من مصادر مختلفة لأغراض مختلفة يتطلّب إيجاد أنظمةٍ يمكنها معالجة البيانات بشكلٍ صحيح يجعلها ذات معنى. وإنّ استخدام البيانات الضّخمة في نماذج تعلُّم الآلة والذّكاء الاصطناعي من شأنه أن يُحسِّن فعالية تلك النّماذج ومتانتها.

تقترح هذه الورقة نموذجاً لكشْف الهجمات الموزّعة المتعلّقة برفض الخدمة (DDoS) باستخدام مجموعة بيانات عامّة معروفة لتدريب النّموذج. ويتمّ تدريب النّموذج بحيث يُمكنه توقُّع نوع الهجمة من بين أنواع متعدّدة من الهجمات يستخدمها المخترقون. وتُستخدم اثنتان من الخوارزميات الواردة في أدبيات الموضوع بوصفهما أساسَ النّموذج المقترح. وتستمدّ هاتان الخوارزميتان فعاليتهما ومتانتهما من الطبيعة المجمَّعة لتركيبهما، حيث تتكون كلٌّ منهما من عددٍ من شجرات القَرار (decision trees) بمتغيرات مختلفة. وقد جرى بناء نظام مجمَّع باستخدام الخوارزميتين من أجل تحسين دقّة كشْف الهجمات؛ واتّضح أنّ استخدام تلك التركيبة المجمَّعة يعطي أفضل النتائج.

من ناحيةٍ أخرى، فإنّ طول زمن التّنفيذ المترتِّب على تدريب النّموذج باستخدام مجموعة بيانات ضخمة يُعدّ مسألةً أخرى لا بدّ من أخذها بعين الاعتبار. ولتسريع عمل النّموذج، فقد تمّ استخدام "الشّرارة" (Apache-Spark) الّتي يؤدي استخدامها إلى تجزئة مجموعة البيانات ومعالجة تلك الأجزاء بالتّوازي، الأمر الّذي يقلّل زمن التّنفيذ مع المحافظة على دقّة كشْف الهجمات.

لقد حقق النّموذج المقترح دقّةً وصلت إلى 99.94%، وقلّل استخدام الشّرارة زمن التّنفيذ إلى ما يقرب من النّصف مقارنة بعدم استخدام الشّرارة. وبالمقارنة مع عددٍ من نماذج كشْف الهجمات الواردة في أدبيات الموضوع، تبيّن أنّ تلك النّماذج حققّت دقّة كشفٍ أقلّ من النّموذج المقترح في هذه الدّراسة.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 02, June 2024.

ILLUMINATION ENHANCEMENT OF NIGHTTIME IMAGES USING A REGULATED SINGLE SCALE RETINEX ALGORITHM

Ola A. Basheer[1] and Zohair Al-Ameen[2]

(Received: 14-Jan.-2024, Revised: 11-Mar.-2024, Accepted: 13-Mar.-2024)

ABSTRACT

Nowadays, people are active during the nighttime and take many photos to record their activities. Due to the low-light nature of the environment at nighttime, captured images tend to appear with dimmed and imbalanced illumination, limited contrast, covert noise and diminished colors. Thus, this paper presents a practical algorithm to improve the illumination of nighttime images based on the single-scale retinex model, image processing methods and certain statistical functions. The developed algorithm initiates by converting the image from the RGB into the HSV model. Then, it enhances only the value (V) channel while preserving the H and S channels. Next, estimating the illumination version of the image and calculating the logarithms of both the illumination and original image are performed. Afterward, a logarithmic subtraction occurs and a modified cumulative distribution function of Gumble probability is applied and the result is further enhanced using a logarithmic transform method. These operations produce the processed V channel and a conversion to the RGB format occurs to generate the final output. The proposed algorithm is experimented with by using two datasets, compared to ten different contemporary algorithms and outcomes are evaluated via three sophisticated metrics. Based on the attained results, promising performances by the developed algorithm have been recorded, surpassing the performance of many existing algorithms in various objective, subjective and runtime terms.

KEYWORDS

Nighttime, Image enhancement, Single-scale retinex, Statistical methods.

1. INTRODUCTION

Nighttime is the period from dusk to dawn [37]. Images taken at nighttime are of defective quality and characterized by unbalanced illumination, unpleasant colors, limited contrast and undesirable noise [1]. Due to the significant increase in nighttime photography to visualize large-scale events, such as personal activities, surveillance and speed cams, there has been an urgent need for efficient nighttime image-enhancement algorithms. Thus, this topic has attracted widespread attention by various beneficiaries. Since hardware is constantly improving, most modern devices and computer-vision applications are required to deliver high-quality images [2]. Image enhancement (IE) refers to the operations applied to an image to improve its perceived quality and make it more visually pleasing to the recipient. The primary goal of IE is to change the characteristics of an image to enhance its suitability for a particular activity and viewer without introducing errors [3]. IE techniques must seek to consider two crucial factors: 1) There may be hidden noise in dark areas of nighttime images, so it must be ensured that the noise is suppressed or kept from being amplified when improving the illumination [4]. 2) Preserving the brightness in the already bright areas from being amplified to avoid the state of over-enhancement [5].

Different algorithms have been introduced to help improve the quality of digital images. One concept of interest is the Retinex theory, which is commonly used for image enhancement and owns many versions, such as the single-scale retinex (SSR), multi-scale retinex (MSR) and multi-scale retinex with color restoration (MSRCR) [6]. This research introduces a well-developed SSR algorithm using statistical and image-processing methods for nighttime-image enhancement. The performance of the proposed algorithm has been tested on two datasets, compared to ten contemporary algorithms explained in the related-work section, in addition to evaluating and discussing the results thoroughly. The paper is organized as follows: the 2nd section explains a literature review of recent years' research work, the 3rd section explains the developed algorithm in depth, the 4th section presents the attained results and the 5th section gives a brief conclusion.

1. O. A. Basheer is with the Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, Mosul, Nineveh, Iraq. Email: olaalh9@gmail.com
2. Z. Al-Ameen is with the ICT Research Unit, Computer Center, University of Mosul Presidency, University of Mosul, Mosul, Nineveh, Iraq. Email: qizohair@uomosul.edu.iq

2. RELATED WORK

In recent years, numerous studies have been introduced on improving nighttime images due to their high importance in different real-world applications. The selected studies are reviewed in a newer to older style. In 2024, a method that utilizes gamma correction and merged color spaces is introduced [35], in that the algorithm starts by determining a transmission map (TM) that includes the saturation information of the degraded image in two different color spaces. Next, the calculated TM is transformed into a function that contains the max and mean values and these values are approximated from a poor illumination image by utilizing a gamma-correction approach. After that, an adaptive value-determination algorithm is applied to enhance the image, prevent the over-enhancement phenomenon and generate the output. In 2023, a Gaussian-based model (GM) was developed [34] and this algorithm starts by creating the GM to get the estimated reflectance and illumination information based on the retinex theory. Then, based on the retinex theory, a decomposition in the GM-based operation is applied to the illumination layer and a gradient descent-based approach is implemented to enhance the image's illumination. Lastly, a denoising process based on the total-variation concept is executed on the reflectance layer to reduce the noise and generate the output.

In 2023, a triangle similarity-based algorithm (TS) was presented [33], in that it begins by transforming the image into the HSI color domain and maintaining the hue channel while processing the saturation and intensity channels. Next, a translation-based operation is applied to the saturation channel to improve the color representation. After that, various scaling operations are implemented in the intensity channel to improve the illumination and visual information. Lastly, a transformation to the RGB domain is applied to create the output. In 2022, a structure preservation-based variation model (SPV) was provided [32] and it started by utilizing a variation-coefficient-based concept to improve the illumination information. Next, a total-variation concept is implemented to reduce the noise information in the image. Lastly, these two images are mixed using the retinex concept in an iterative way to generate the output. In 2021, a progressive-recursive network-based algorithm was established [36], in that the method begins by getting the degraded image and sending it to a dual-attention approach to extract the global features. After that, a mixture of residual blocks and recurrent layers is utilized to extract the local features. Based on the extracted local and global features, several recursive operations are applied to enhance the image and create the output.

In 2020, a semi-decoupled decomposition (SDD) algorithm was proposed [7], in that it decomposes the image using the retinex model into reflectance and illumination components in a semi-decoupled manner. The illumination layer is enhanced progressively and the reflectance layer is improved jointly using a specialized total-variation concept. These components are united to create the output. In 2020, a retinex-based multi-phase (RBMP) algorithm was proposed [8], which is initiated by computing the illumination image in a manner akin to the standard SSR algorithm, subtracts the log of the illumination image from the log of the original image using a modified method and then processes the output through a gamma-corrected sigmoid and normalization approaches to generate the output. In 2019, an adaptive image-enhancement (AIE) algorithm was presented [9], where it first transforms the image to the HSV domain and the V channel is processed to isolate the illumination component of the scene through a multi-scale Gaussian function. Afterward, a correction function is implemented *via* the Weber-Fechner law and two outputs are generated by adaptively adjusting the parameters according to the distribution profiles of the illumination components. The output is created by combining both images using a specially-developed approach. Similarly, in 2019, an algorithm named LECARM was developed [10], which began by utilizing illumination-estimating algorithms to calculate the exposure ratio for every pixel. After that, the chosen camera-response model is employed to modify each pixel to achieve the required exposure based on the estimated exposure-ratio map. Lastly, the output is obtained using a specific mapping method.

Moreover, in 2018, a robust retinex model (RRM) algorithm was presented [11], which starts by applying advanced regularization terms for illumination and reflectance approximation. More precisely, it employs one norm to limit the smoothness of the illumination in different regions, joining a fidelity term to highlight the structural details in low-light areas with the gradients of the reflectance to estimate the noise map using a robust Retinex concept. Next, the enhancement is applied using a Lagrange multiplier-based approach to build the output image. In 2016, a fusion-based enhancement (FBE) algorithm was developed [12], which utilizes an illumination-estimation algorithm based on

morphological closure to separate an observed image into reflectance and illumination components. This algorithm generates two images from the illumination image, one with brightness enhancement and the other with contrast enhancement, by applying sigmoid transform and adaptive-histogram equalization. Moreover, two weights are created using a multi-scale process. Lastly, the two images are fused using the determined weights to create the output. In 2016, an algorithm called LIME was proposed [13], which starts by determining the max values of the RGB image, followed by the determination of reflectance and illumination information *via* the retinex model. The illumination information is enhanced using a structure-processing concept, followed by implementing different maps to further boost illumination. The output image is generated by joining the improved-illumination and reflectance components.

As seen from the studied algorithms, different notions were used and the computational cost of each algorithm varies. Most of the proposed algorithms in this field did not reach the required level of enhancement. Thus, the chance remains to develop a new method that can improve the illumination of nighttime images more efficiently. The proposed algorithm differs from existing algorithms in several aspects. First, low computational developments are utilized to make the proposed algorithm efficient and particularly fast in filtering different nighttime images. Second, the utilized developments improve the illumination in a direct and non-iterative way while considering minimal noise augmentation, which is needed as many existing algorithms utilize the iterative feature and their utilized processing steps may amplify the noise, which leads to the requirement of another major step for image denoising, making such algorithms slow and require high computational cost.

3. PROPOSED ALGORITHM

Land and McCann initially proposed the retinex theory [30]. The term "retinex" is derived from the combination of the root terms "retina" and "cortex," which are both essential components of human vision. Retinex is more visually consistent with human vision. This is predicated on the notion that the reflectance and illumination components' collective influence creates the image, as shown in Figure 1.

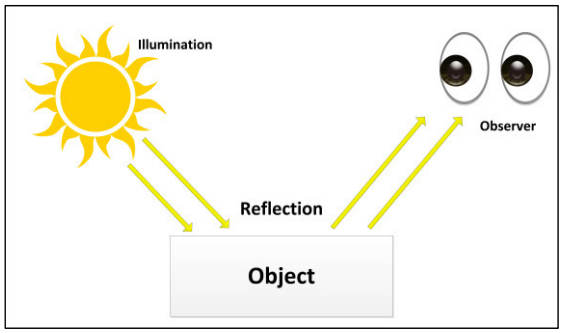


Figure 1. The retinex theory.

Specifically, when light illuminates an object, it creates a reflection that is then seen as an image by the human eye [6]. Various algorithms have been designed based on the retinex theory, such as the SSR [14] and the MSR [15]. Both algorithms utilize a specific Gaussian function to modify a given image. Therefore, the brightness level of the output image is determined by using the natural logarithm of reflectance. Nevertheless, it may exhibit a color-distortion effect, which poses a difficulty in both the SSR and MSR. A potential solution to address this problem is the implementation of a "Multi-scale Retinex with Color Restoration" (MSRCR) technique [16]. Here, the inclusion of MSRCR allows for the handling of color distortion and restoration by utilizing the color ratio of the red, green and blue (RGB) channels. However, due to the universally-applied mapping curve, this method tends to diminish the level of detail in the image, particularly in the areas of high brightness.

The primary motivation of the proposed regulated SSR (RSSR) algorithm is to improve the quality of night images by improving lighting in dark areas without intensifying brighter regions using non-complex concepts. The SSR model and other less-complex statistical concepts and methods were used among these concepts. The original SSR model is used for illumination estimation of degraded images by applying convolution (*) between the input image $I_{(x,y)}$ and the Gaussian function $G_{(x,y)}$, which is calculated in the following manner [17]:

$$G_{(x,y)} = Q \cdot \exp\left(-\frac{\left(U^2 + V^2\right)}{2\sigma^2} \right) \tag{1}$$

$$Q = \frac{1}{\sum_{=1}^{N}\sum_{y=1}^{M} \exp\left(-\frac{\left(U^2 + V^2\right)}{2\sigma^2} \right)} \tag{2}$$

Let Q be a normalizing factor; x and y represent the coordinates of the digital image; U and V denote two grayscale gradients, where U is horizontal and V is vertical. U and V hold the same size as $I_{(x,y)}$. Additionally, M and N represent the dimensions of $I_{(x,y)}$, $(\cdot)$ denotes a multiplication operation and σ is a parameter that addresses the brightness. Then, the logarithm of the illumination image resulting from the previous step and the logarithm of the degraded image are taken to produce an enhanced version of the degraded image called reflectance image $R_{(x,y)}$ by subtracting the log of the illumination image from the log of the degraded image [14], in the following manner:

$$R_{(x,y)} = \log\left(I_{(x,y)}\right) - \log\left(G_{(x,y)} * I_{(x,y)}\right) \tag{3}$$

where $R_{(x,y)}$ is the output of original SSR. Experiments have been conducted on applying the standard SSR on different nighttime images to determine its filtering abilities with this type of image. Some results are demonstrated in Figure 2.

Figure 2. Outputs of the standard SSR model when applied to different nighttime images.

From the conducted experiments, the SSR provided resulting images with defects, including extra dimming for the darkened areas, which led to the loss of visual details, as well as amplification of brightness in the bright areas and the production of unrealistic colors, leading to overall unacceptable results. Regardless of these defects, the standard SSR model is characterized by low computational cost, which is a key aspect and has a high development potential [14].

The proposed RSSR algorithm aims to improve illumination while producing appropriate colors and avoids the over-amplification of the latent noise. The RSSR algorithm begins its first phase by converting the image from the RGB form into the HSV color model [18]. This color model is designed to efficiently separate the color information from the brightness (value) information, making it intuitive to improve brightness by simply modifying the value component. Supposing that the input image $I_{(x,y)}$ has three color channels of red (R), green (G) and blue (B) and R, G, B $\in$ [0, W_m], with W_m being the max. range value (typically 1), assuming the range $\in$ [0,1], the conversion to the HSV color domain can be achieved using the following equations [31]:

$$S = \frac{W_r}{W_h} \quad \text{for } W_h > 0, \quad 0 \text{ otherwise} \tag{4}$$

$$V = \frac{W_h}{W_m} \tag{5}$$

with W_h, W_l and W_r defined as $W_h = \max(R, G, B)$, $W_l = \min(R, G, B)$, $W_r = (W_h - W_l)$, where S is the saturation channel and V is the value channel. What is more needed is to determine the hue (H) channel, wherein if the three RGB channels contain a similar value, then it is the case of a gray pixel. In this situation, $W_r = 0$, $S = 0$ and H is undefined. To calculate H when $W_r > 0$, each channel is normalized in the following manner:

$$\hat{R} = \frac{W_h - R}{W_r}, \qquad \hat{G} = \frac{W_h - G}{W_r}, \qquad B = \frac{W_h - B}{W_r}. \tag{6}$$

Next, the initial hue (Ĥ) is calculated based on the notion of which color channel contains the max. value in the following manner:

$$\hat{H} = \begin{cases} \hat{B} - \hat{G} & \text{if } R = W_h \\ \hat{R} - \hat{B} + 2 & \text{if } G = W_h \\ \hat{G} - \hat{R} + 4 & \text{if } B = W_h \end{cases} \tag{7}$$

The outcome value of Ĥ is in the range of [-1, 5] and the final H channel is obtained in the range of [0,1] as follows:

$$H = \frac{1}{6} \cdot \begin{cases} \left(\quad + 6 \right) & \text{for } \hat{H} < 0 \\ \hat{H} & \text{otherwise.} \end{cases} \tag{8}$$

The operations are performed only on the value channel, because the key requirement here is to improve the illumination, as the HSV color domain separates the color information from the illumination information. Thus, the processing becomes rapid and efficient. In the second phase, the Gaussian function $G_{(x,y)}$ is calculated using Eq. (1) and Eq. (2), where ($\sigma = N \times M$). The third phase includes the computation of the illumination image in the following manner [17]:

$$M_{(x,y)} = G_{(x,y)} \quad I_{(x,y)} \tag{9}$$

where, $M_{(x,y)}$ is the illumination image. To apply the convolution (*) in the frequency domain, first, the Fourier transform is used to convert the inputs from the spatial domain into the frequency domain. Then, the element-wise multiplication between two inputs of the same size is computed. It often needs a frequency shift to return the high frequencies in the middle and the low frequencies in the edges and finally convert the image from the frequency domain into the spatial domain [19]. In the fourth phase, the log of the illumination image $M_{(x,y)}$ and the log of the input image $I_{(x,y)}$ are determined as follows:

$$O_{(x,y)} = \log\left(I_{(x,y)} + \varepsilon \right) \tag{10}$$

$$L_{(x,y)} = \log\left(M_{(x,y)} + \varepsilon \right) \tag{11}$$

Here, ($\varepsilon = 0.001$) represents a minor value added to prevent the computation of the log of zero, which is infinite. The fifth utilized phase includes the application of a logarithmic-subtraction approach [20], as logarithmic image processing has been utilized in dynamic range manipulation, improving the visibility of details in both dark and bright regions, replacing the standard-subtraction method in Eq. (3) to produce the reflectance image, as follows:

$$Z_{(x,y)} = \frac{O_{(x,y)} - L_{(x,y)}}{1 \quad O_{(x,y)} \cdot L_{(x\ y)}} \tag{12}$$

After that, the sixth phase is implemented, which includes the utilization of a slightly modified cumulative distribution function of the Gumble probability (CDF-GP) approach. The standard CDF-GP approach can be mathematically expressed as follows [21]:

$$F_{(x,y)} = \exp\left(-\exp\left(-\frac{(x-\mu)}{\beta} \right) \right) \tag{13}$$

This approach redistributes the values across the image, emphasizing certain brightness levels over others and improves the contrast depending on β. With a slight heuristic modification to simplify the calculation, its equation becomes as follows:

$$F_{(x,y)} = \exp\left(-\exp\left(-\frac{Z_{(x,y)}}{\beta} \right) \right) \tag{14}$$

where $\beta > 0$ is the parameter that controls the image illumination and contrast, in that lower β values compress the range of values, reducing illumination and contrast. In comparison, higher β values spread

out the intensity values, potentially enhancing illumination and contrast. Next, the log transform is applied as the seventh phase to further improve fine details in low-density areas. This transformation is appropriate for an excessively dark image, as it increases the values of dark pixels and decreases the values of highly-illuminated pixels [22], resulting in a well-balanced, visually pleasing outcome. The log transform can be computed as follows [23]:

$$S_{(x,y)} = c \cdot \log\left(F_{(x\ y)} + 1\right) \tag{15}$$

where $S_{(x,y)}$ represents the resulting value channel and c is a luminance parameter that is set to 2.5. In the final eighth phase, a conversion from HSV to RGB is applied. To convert the HSV image, where ∈ [0, 1], to the corresponding RGB image, the following is applied [31]:

$$\hat{H} = (6 \cdot H) \bmod 6 \tag{16}$$

where $(0 \leq \hat{H} < 6)$ is initially obtained, then the intermediate values are calculated as follows:

$$\begin{aligned} c_1 &= \lfloor \hat{H} \rfloor & x &= (1 - S) \cdot v \\ c_2 &= \hat{H} - c_1 & y &= \left(1 - \left(S \cdot c_2\right)\right) \cdot V \\ v &= V & z &= \left(1 - \left(S \cdot (1 - c_2)\right)\right) \cdot V \end{aligned} \tag{17}$$

Using these pre-determined values, the normalized RGB channels are computed as follows:

$$\left(\hat{R}, \hat{G}, \hat{B}\right) = \begin{cases} (v, z, x) & \text{if } c_1 = 0 \\ (y, v, x) & \text{if } c_1 = 1 \\ (x, v, z) & \text{if } c_1 = 2 \\ (x, y, v) & \text{if } c_1 = 3 \\ (z, x, v) & \text{if } c_1 = 4 \\ (v, x, y) & \text{if } c_1 = 5 \end{cases} \tag{18}$$

Finally, scaling the channels to the range of $[0, A\text{-}1]$ (normally $A = 256$) is done in the following manner:

$$\begin{aligned} R &= \min\left(\text{round}\left(A \cdot \hat{R}\right), A - 1\right) \\ G &= \min\left(\text{round}\left(A \cdot \hat{G}\right), A - 1\right) \\ B &= \min\left(\text{round}\left(A \cdot \hat{B}\right), A - 1\right) \end{aligned} \tag{19}$$

where RGB is the final algorithm output. The flowchart of the proposed RSSR algorithm is demonstrated in Figure 3.

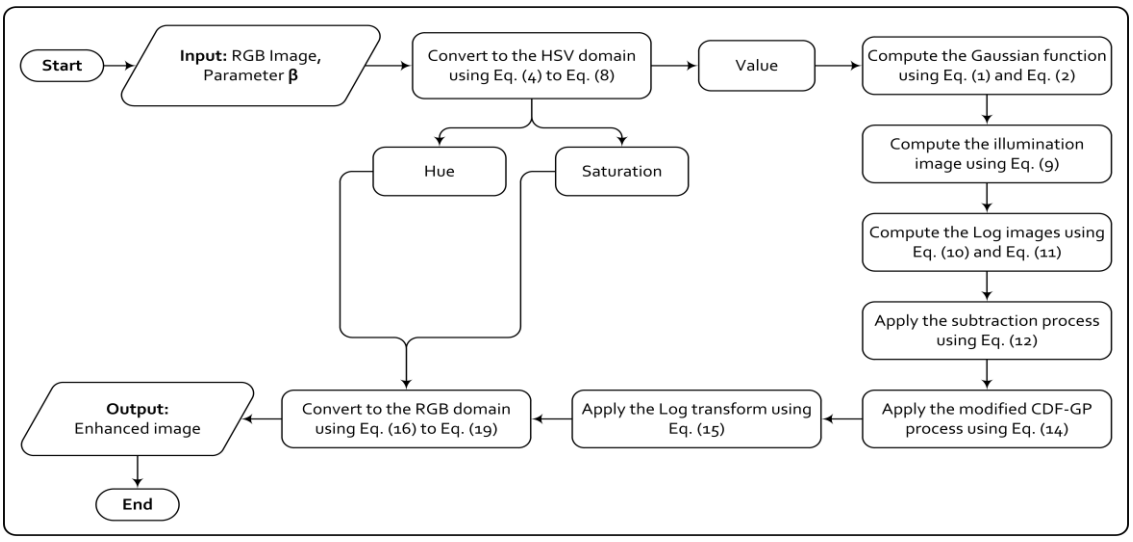


Figure 3. Flowchart of the proposed RSSR algorithm.

4. RESULTS AND DISCUSSION

In this section, the results of the proposed algorithm are presented and its performance is evaluated on low-light nighttime images. These results are also discussed and compared to the results of other algorithms. In this study, two datasets were used. The first is the MIT-Adobe FiveK dataset [24], which contains five thousand images captured using single-lens (SLR) cameras by different photographers, wherein the images are all in RAW format, meaning that all data captured by the camera sensor is pristine. Photoshop was used to convert these images from the DNG format into the JPG format. The second one is the exclusively Dark (ExDARK) dataset [25], containing approximately seven thousand images captured in low-light conditions.

When it comes to the comparison, the proposed algorithm is compared with ten advanced methods; namely, SDD [7], AIE [9], FBE [12], LECARM [10], LIME [13], RBMP [8], RRM [11], SPV [32], TS [33] and GM [34]. Moreover, the results of the proposed method and other methods are evaluated using one reduced-reference (RR) metric, called the lightness order error (LOE) and two no-reference (NR) metrics that are natural image-quality evaluator (NIQE) metric and blind/referenceless image spatial quality evaluator (BRISQUE). The LOE [27] is utilized to measure the error of the lightness order (i.e., illumination quality) between the degraded image and its filtered counterpart. The output of the LOE is a numerical value, where lower scores represent a better illumination quality. The LOE is defined as:

$$LOE = \frac{1}{W \cdot H} \cdot \sum_{=1}^{W} \sum_{y=1}^{H} D_{(x,y)} \tag{20}$$

The variables W and H represent the image dimensions and $D_{(x,y)}$ denotes the relative order difference in luminance between two given images. Moreover, the NIQE [28] measures the naturality and evaluates the quality based on measurable deviations from statistical patterns found in natural images without considering expected distortions or human subjective judgments. The quality of the distorted image is quantified by measuring the difference between the statistical properties of the model and the distorted image. The output NIQE is a numeric value, where lower scores represent better naturality. Likewise, the BRISQUE [29] measures the distortions and perceived quality and utilizes natural scene statistics (NSS) to construct a distortion-agnostic no-reference metric for image quality that functions in the spatial domain. NSS focuses on analyzing the statistical patterns seen in "natural scene" photos and developing metrics to quantify the extent to which the statistical properties of an unfamiliar image differ from those of typical natural scene images. The output BRISQUE is a numeric value, where lower scores represent low distortion and high quality, which is deemed better. In brief, the NIQE measures the naturality, the LOE measures the illumination quality and the BRISQUE measures the existence of distortions.

As for computational complexity, CPU runtime can deliver insights into an algorithm's efficiency and complexity [26]. Let's dissect it: the computational complexity measures the number of resources required by a method to solve a problem. It's usually quantified in terms of space and time complexity. The CPU runtime, on the other hand, denotes the real time needed by a CPU to implement a specific method. It relies on numerous aspects, such as the method's complexity, the input size and the hardware utilized. Comparative analysis (CA) can be applied to this case. CA means that comparing the CPU runtimes of various methods for the same task provides a sense of relative computational complexities, in that a method with a lower runtime for the same input size denotes a lower computational complexity. Thus, CPU runtimes have been considered as a computational complexity measure and provided in this study in Table 4 and Figure 13. The computer on which experiments and evaluations were performed had specifications of 16 GB of RAM, a Core i7-8650U 2.11 GHz processor and MATLAB 2020a. Figures 4-7 show the experimental results of the proposed algorithm with various degraded nighttime images, Figures 8-11 demonstrate the comparison results. Moreover, Tables 1-4 show the recorded scores and implementation times for the compared algorithms. Finally, Figure 12 and Figure 13 display the graphs of the average performance in Tables 1-4.

As in the given samples of the conducted experiments, the proposed algorithm succeeded in improving the quality of nighttime images in that it illuminates dark areas while preserving the illumination of the bright regions from being extremely amplified, in addition to emphasizing the visual details of the filtered images. This balance benefits in maintaining the natural illumination while improving visibility in darker image parts. Moreover, the output images from the proposed RSSR algorithm have vibrant,

eye-comforting colors with acceptable contrast, bearing in mind that the proposed method does not add any distortion or unwanted artifacts during the processing procedure and prevents the noise from being massively augmented. This guarantees that the processed images stay true to the pristine scene without

Figure 4. Experimental results of the MIT-Adobe dataset **(Batch-1)**: (1st row) represents unprocessed nighttime images; (2nd row) represents the version of the enhanced images from the proposed algorithm with β values equal to (5, 6, 7, 5, 6).

Figure 5. Experimental results of the MIT-Adobe dataset **(Batch-2)**: (1st row) represents unprocessed nighttime images; (2nd row) represents the version of the enhanced images from the proposed algorithm with β values equal to (7, 6, 7, 7, 6).

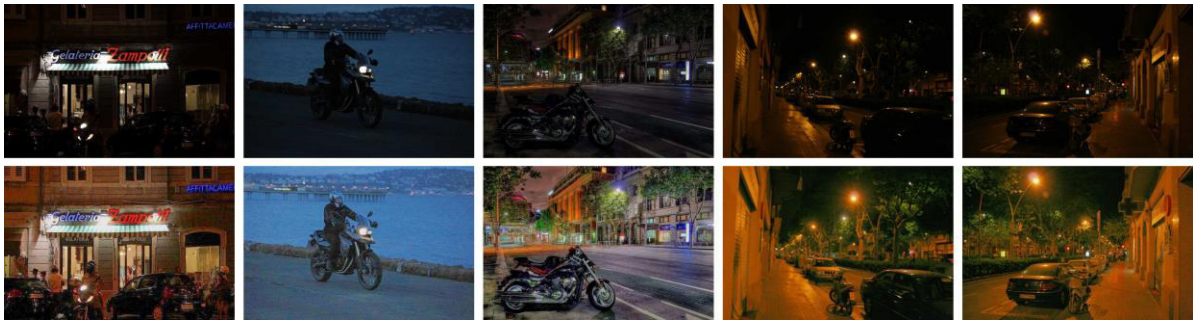


Figure 6. Experimental results of the ExDARK dataset **(Batch-1)**: (1st row) represents unprocessed nighttime images; (2nd row) represents the version of the enhanced images from the proposed algorithm with β values equal to (11, 9, 9, 10, 10).

Figure 7. Experimental results of the ExDARK dataset **(Batch-2)**: (1st row) represents unprocessed nighttime images; (2nd row) represents the version of the enhanced images from the proposed algorithm with β values equal to (12, 11, 10, 9, 9).

presenting any visual irregularities. Moreover, this also indicates that the RSSR algorithm not only enhances visibility, but also improves the images' visual appeal. In addition, the calculations are low and therefore, the proposed method has great potential in night-image processing, making it suitable for resource-constrained applications.

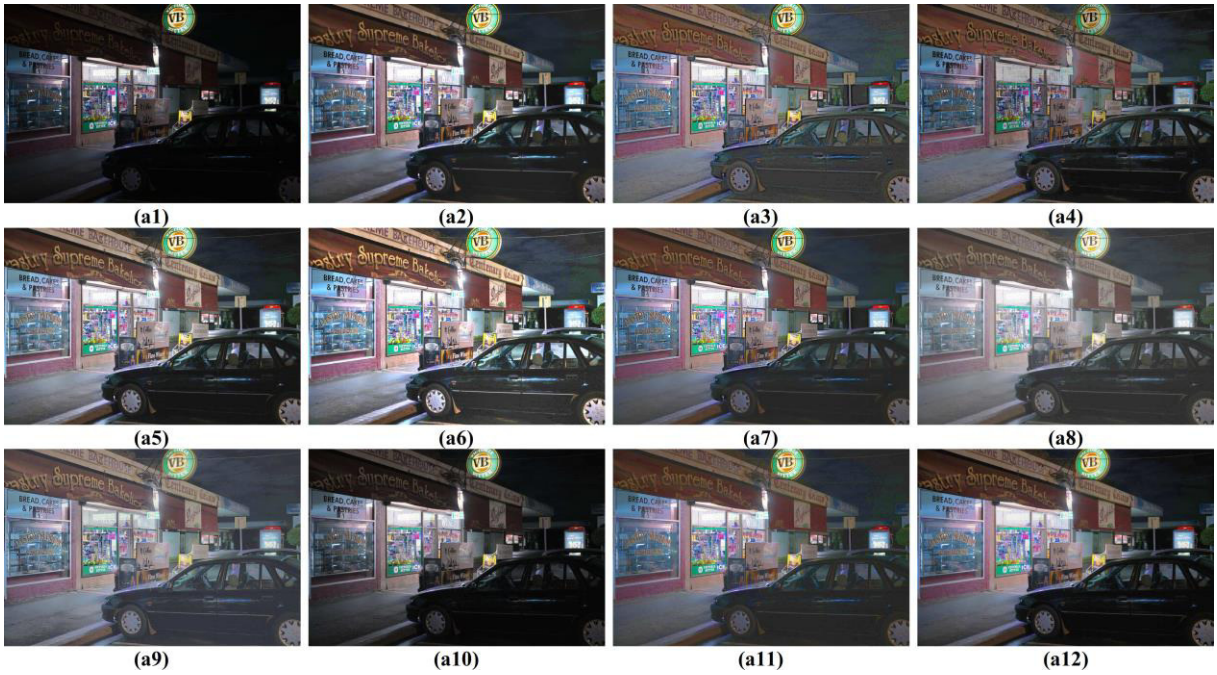


Figure 8. Comparison results **(Batch-1)**. (a1) unprocessed image (1024×683); (a2) SDD; (a3) AIE; (a4) FBE; (a5) LECARM; (a6) LIME; (a7) RRM; (a8) RBMP; (a9) GM; (a10) TS; (a11) SPV; (a12) proposed RSSR.

From the outcomes of the performed comparisons, it is observed that each method provides different enhancement modes due to the different used processing notions, wherein the analysis of each method depends on aspects, such as quality of illumination, contrast, colors, sharpness, in addition to the generation or increase in noise, artifacts or errors. SDD provided insufficient illumination with a smoothed appearance and brightness amplification. It's why the metrics readings are low and the processing speed is slow due to the implementation of the noise-reduction process. AIE delivered the second-best reading in terms of LOE compared to the other methods. However, the unnatural tonality and noise generation led to scoring poorly in NIQE and not good in BRISQUE. Still, it recorded the second fastest method in terms of processing time. FBE recorded low and unusual brightness and contrast but with adequate sharpness. Thus, LOE readings were not good, but BRISQUE and NIQE readings were agreeable and the processing speed was considered acceptable.

Likewise, LECARM produced images with insufficient lighting and had white shadows around the edges. Thus, LOE readings were unacceptable, but the BRISQUE and NIQE readings were reasonable with relatively fast processing speeds. LIME introduced brightness amplification, unusual illumination, processing errors and boosted colors. That is why the LOE readings were the worst among the competitors, yet they averaged in terms of BRISQUE and NIQE with above-average processing speed. In addition, the RRM algorithm provided average illumination with over-smoothness. Due to that, the LOE readings were mediocre, but due to the over-smoothness, the readings of BRISQUE and NIQE were very low. As for the processing time, it was the worst as it took an extremely long processing period. Moreover, RBMP delivered adequate illumination with somewhat pale colors. Thus, the LOE readings were satisfactory as well and the BRISQUE and NIQE metrics recorded the second-best results, considering that it did not generate distortions, provided slightly pale colors and was noticeable fast.

GM delivered results with limited brightness, imbalanced contrast and slightly pale colors, scoring below average in LOE, low in BRISQUE and NIQE and with slow performance according to the average runtime. TS proved to have low illumination and artifacts in the results, leading to low LOE, BRISQUE and NIQE readings with fast runtimes. SPV increased the illumination and surged the difference

between the brightest and darkest regions in the image, leading to somewhat average readings according to the utilized metrics. When it comes to the proposed RSSR, it outperformed all the other comparison algorithms subjectively and objectively, as it recorded the best readings according to LOE, BRISQUE and NIQE metrics with the fastest execution time. It is essential, because it is infrequent to have an algorithm that produces high-quality results rapidly without generating distortion or massive noise presentation. In this context, the proposed algorithm excels and its performance is considered positive and distinctive for the desired purpose, improving the illumination of nighttime images.

Figure 9. Comparison results **(Batch-2)**. (b1) unprocessed image (1024×680); (b2) SDD; (b3) AIE; (b4) FBE; (b5) LECARM; (b6) LIME; (b7) RRM; (b8) RBMP; (b9) GM; (b10) TS; (b11) SPV; (b12) proposed RSSR.

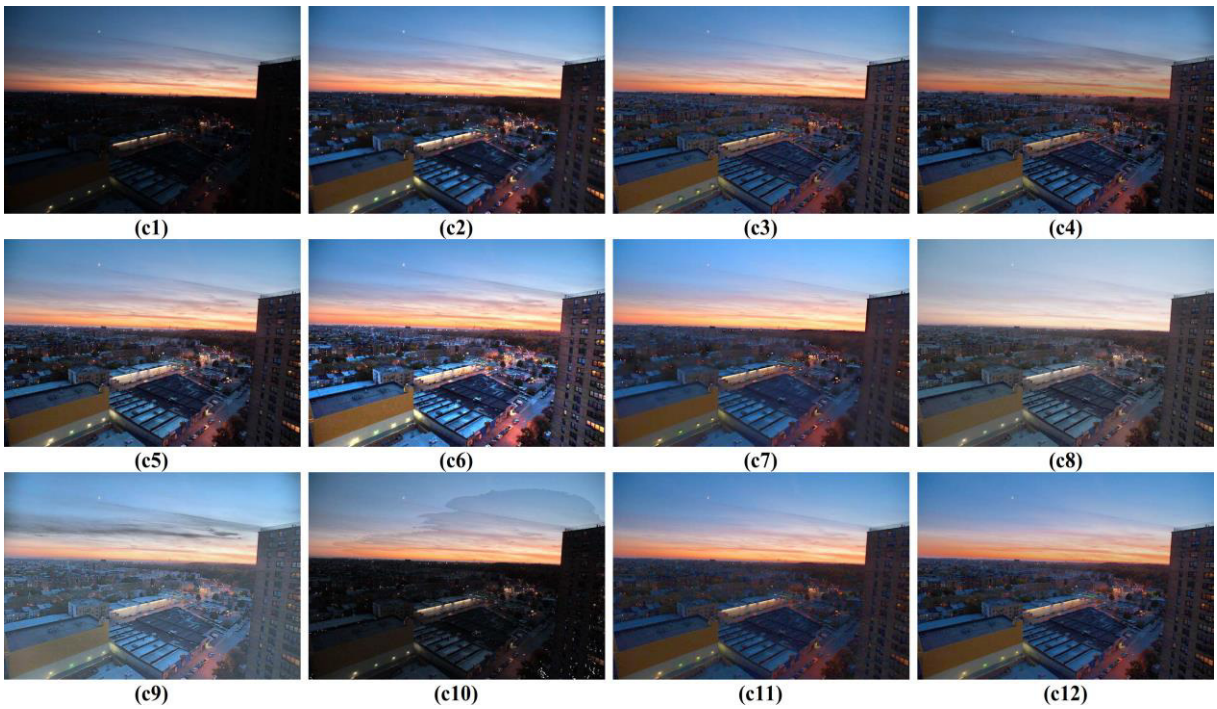


Figure 10. Comparison results **(Batch-3)**. (c1) unprocessed image (800×532); (c2) SDD; (c3) AIE; (c4) FBE; (c5) LECARM; (c6) LIME; (c7) RRM; (c8) RBMP; (c9) GM; (c10) TS; (c11) SPV; (c12) proposed RSSR.

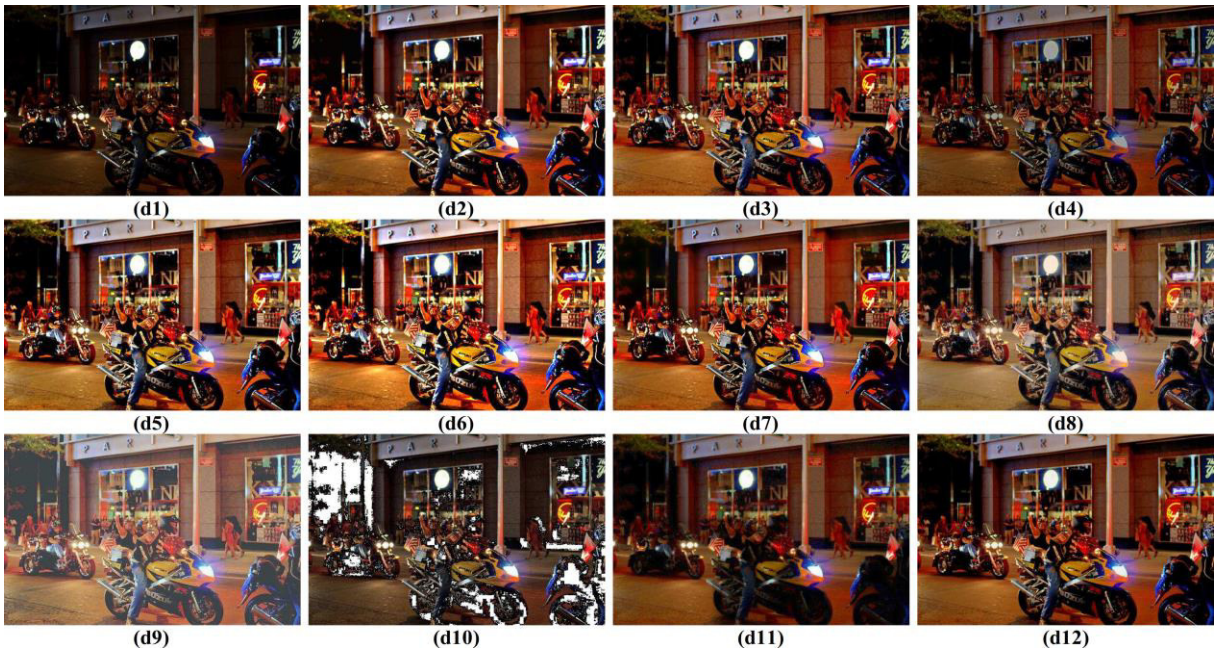


Figure 11. Comparison results **(Batch-4)**. (d1) unprocessed image (640×427); (d2) SDD; (d3) AIE; (d4) FBE; (d5) LECARM; (d6) LIME; (d7) RRM; (d8) RBMP; (d9) GM; (d10) TS; (d11) SPV; (d12) proposed RSSR.

Table 1. The recorded LOE↓ scores.

Image	SDD	AIE	FBE	LECARM	LIME	RRM	RBMP	SPV	TS	GM	Proposed
Fig.8	310.6972	13.3393	285.7465	140.5607	619.3701	175.1018	6.8974	103.5646	100.0498	246.3067	0.9549
Fig.9	283.3439	12.1001	298.9320	172.7044	543.7599	210.1371	35.6361	167.4058	196.5971	286.8417	0.7978
Fig.10	400.8085	10.9486	376.5085	284.0700	978.8198	359.1726	34.6262	197.8955	607.2545	477.0071	7.5333
Fig.11	575.8479	11.4816	427.7972	460.4458	1087.1	571.9542	87.6679	1010.5	1589.9	395.6528	11.3453
Avg	392.6743	11.9674	347.2460	264.4452	807.2624	329.0914	41.2069	369.8414	623.4503	351.4520	**5.1578**

Table 2. The recorded BRISQUE↓ scores.

Image	SDD	AIE	FBE	LECARM	LIME	RRM	RBMP	SPV	TS	GM	Proposed
Fig.8	29.3314	34.9999	26.9860	28.0290	33.1889	31.3011	26.8401	21.8559	21.5089	28.9948	24.7420
Fig.9	28.8798	20.7151	19.1639	19.2840	20.0713	31.2884	19.7321	28.2142	17.9283	31.7083	18.0840
Fig.10	30.0810	13.0398	12.4795	12.6481	12.7708	43.1239	10.2004	23.5718	30.5554	23.8386	8.6066
Fig.11	25.2895	22.4329	18.0028	16.4182	18.8628	25.8702	17.2163	21.2278	39.0027	23.7215	15.5915
Avg	28.3954	22.7969	19.1580	19.0948	21.2234	32.8959	18.4972	23.7174	27.2488	27.0658	**16.7560**

Table 3. The recorded NIQE↓ scores.

Image	SDD	AIE	FBE	LECARM	LIME	RRM	RBMP	SPV	TS	GM	Proposed
Fig.8	2.8009	3.1575	2.3784	2.3234	2.5689	3.1398	2.3259	2.5575	2.2575	3.0323	2.1849
Fig.9	2.3427	2.9661	2.5334	2.5476	2.7440	2.8844	2.2979	2.6727	2.4478	2.7833	2.4188
Fig.10	3.3333	2.0966	2.1190	2.0112	2.1122	3.1282	2.0596	2.5849	3.1453	2.9401	1.9478
Fig.11	3.0179	3.3062	2.6004	2.7839	2.8941	3.4941	2.8951	2.9095	4.4594	3.1705	2.6001
Avg	2.8737	2.8816	2.4078	2.4165	2.5798	3.1616	2.3946	2.68115	3.0775	2.98155	**2.2879**

Table 4. The recorded runtimes↓ (in seconds).

Image	SDD	AIE	FBE	LECARM	LIME	RRM	RBMP	SPV	TS	GM	Proposed
Fig.8	21.40511	0.18263	0.84769	0.68879	2.16288	54.36626	0.45154	18.248078	0.500824	46.933321	0.13474
Fig.9	18.44903	0.18122	0.71538	0.73408	2.42125	105.09780	0.33700	22.879234	0.543656	50.913605	0.17260
Fig.10	13.30351	0.15941	0.71985	0.40115	1.10794	31.33848	0.25359	11.602561	0.310696	33.103307	0.06986
Fig.11	11.11168	0.26329	2.76838	0.37095	3.64605	36.45124	0.18027	18.511313	0.162029	16.652700	0.06114
Avg	16.06733	0.19664	1.26282	0.54874	2.33453	56.81345	0.30560	17.8102	0.37930	36.9007	**0.10958**

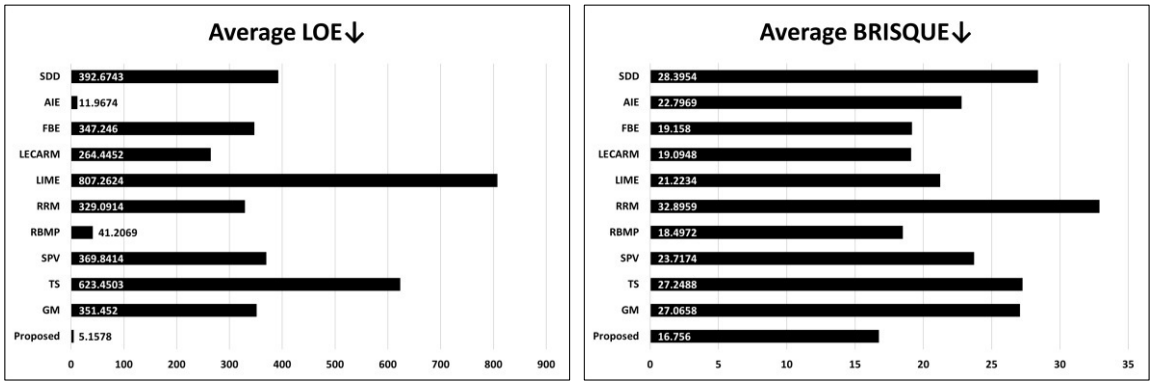


Figure 12. Average readings of LOE and BRISQUE.

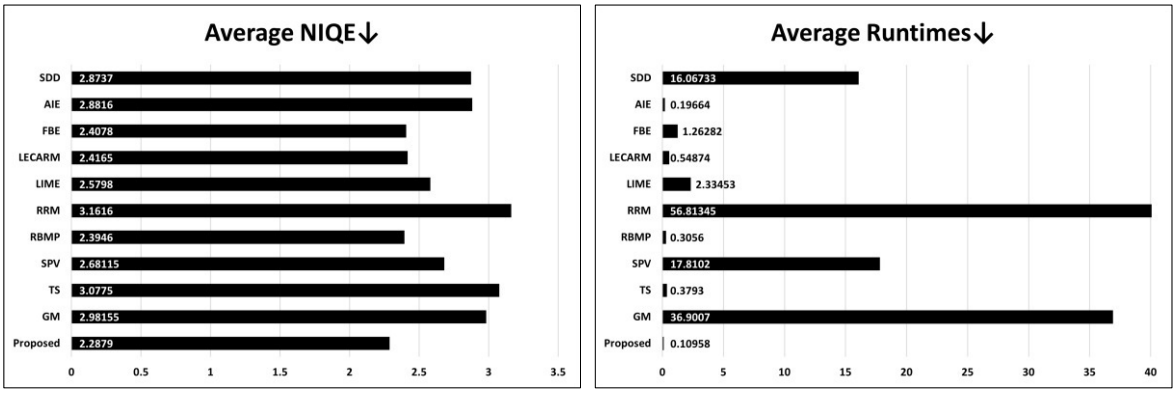


Figure 13. Average readings of NIQE and runtimes.

The proposed RSSR outperformed the other algorithms in the metrics used because of the careful development and attentive analysis of the drawbacks of the related-work methods, knowing what advantages to consider and what disadvantages to avoid. When developing the RSSR, it has been affirmed that proper illumination must be provided with balanced contrast and attractive colors and focused on avoiding the generation of unwanted processing errors in addition to evading noise amplification. Thus, the methods used in the development of the RSSR were added and adapted successfully to introduce a fast and efficient algorithm. Despite the accomplishments of the algorithm, it still has one limitation; that is, it is not fully automatic and the human operator should manually choose the value of β to produce the resulting image with the desired illumination.

5. CONCLUSION

This research proposes an algorithm to improve the illumination of nighttime images. This algorithm works on the HSV color model and estimates the illumination image in a similar way to the standard SSR model. Still, it differs in the subtraction process, as it uses logarithmic subtraction in addition to the utilization of two statistical approaches for further visual enhancement, in that the first is a modified CDF-GP approach, which applies a curvy transform and the other one is a non-complex log transform. The performance of the proposed algorithm is assessed by utilizing two different datasets. By performing a comparison with ten contemporary algorithms, the obtained results are then evaluated using three metrics and recorded CPU runtimes. The study's outcomes showed that the proposed RSSR algorithm improved the quality of nighttime images and properly illuminated the details in dark areas while avoiding over-illumination of bright areas, producing images with natural and balanced brightness, adequate colors and adjusted contrast. As a result, the proposed RSSR algorithm outperformed the other algorithms in the used objective measures and recorded the fastest runtime. This is essential, as it is challenging to find an algorithm that is uncomplicated and fast and, at the same time, generates satisfactory results. In future work, it is likely to embrace developments by including AI for automation.

ACKNOWLEDGMENT

The authors would like to thank the University of Mosul staff for the success of this study.

REFERENCES

[1] Y. F. Wang, H. M. Liu and Z. W. Fu, "Low-light Image Enhancement *via* the Absorption Light Scattering Model," IEEE Transactions on Image Processing, vol. 28, no. 11, pp. 5679–5690, 2019.

[2] X. Guo, Y. Li and H. Ling, "LIME: Low-light Image Enhancement *via* Illumination Map Estimation," IEEE Transactions on Image Processing, vol. 26, no. 2, pp. 982–993, 2017.

[3] Y. Qi et al., "A Comprehensive Overview of Image Enhancement Techniques," Archives of Computational Methods in Engineering, vol. 29, no. 1, pp. 583–607, 2022.

[4] M. Li, J. Liu, W. Yang, X. Sun and Z. Guo, "Structure-revealing Low-light Image Enhancement *via* Robust Retinex Model," IEEE Transactions on Image Processing, vol. 27, no. 6, pp. 2828–2841, 2018.

[5] H. Lee, "Successive Low-light Image Enhancement Using an Image-adaptive Mask," Symmetry (Basel), vol. 14, no. 6, p. 1165, 2022.

[6] R. R. Hussein, Y. I. Hamodi and R. A. Sabri, "Retinex Theory for Color Image Enhancement: A Systematic Review," Int. J. Electr. Comput. Eng. (IJECE), vol. 9, no. 6, p. 5560, 2019.

[7] S. Hao, X. Han, Y. Guo, X. Xu and M. Wang, "Low-light Image Enhancement with Semi-decoupled Decomposition," IEEE Transactions on Multimedia, vol. 22, no. 12, pp. 3025–3038, 2020.

[8] M. A. Al-Hashim and Z. Al-Ameen, "Retinex-based Multiphase Algorithm for Low-light Image Enhancement," Traitement du Signal (TS), vol. 37, no. 5, pp. 733–743, 2020.

[9] W. Wang, Z. Chen, X. Yuan and X. Wu, "Adaptive Image Enhancement Method for Correcting Low-illumination Images," Information Sciences, vol. 496, pp. 25–41, 2019.

[10] Y. Ren, Z. Ying, T. H. Li and G. Li, "LECARM: Low-light Image Enhancement Using the Camera Response Model," IEEE Trans. on Circuits and Sys. for Video Techn., vol. 29, no. 4, pp. 968–981, 2019.

[11] M. Li, J. Liu, W. Yang, X. Sun and Z. Guo, "Structure-revealing Low-light Image Enhancement *via* Robust Retinex Model," IEEE Transactions on Image Processing, vol. 27, no. 6, pp. 2828–2841, 2018.

[12] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding and J. Paisley, "A Fusion-based Enhancing Method for Weakly Illuminated Images," Signal Processing, vol. 129, pp. 82–96, 2016.

[13] X. Guo, "LIME: A Method for Low-light Image Enhancement," Proc. of the 24th ACM Int. Conf. on Multimedia, DOI:10.1145/2964284.2967188, 2016.

[14] D. J. Jobson, Z. Rahman and G. A. Woodell, "Properties and Performance of a Center/Surround Retinex," IEEE Transactions on Image Processing, vol. 6, no. 3, pp. 451-462, 1997.

[15] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Resiliency of the multiscale retinex image enhancement algorithm," in Color Imaging Conference: Color Science, Systems and Applications, 1998.

[16] D. J. Jobson, "Retinex Processing for Automatic Image Enhancement," Journal of Electronic Imaging, vol. 13, no. 1, pp. 100-110, 2004.

[17] M. Ismail and Z. Al-Ameen, "Adapted Single Scale Retinex Algorithm for Nighttime Image Enhancement," AL-Rafidain J. of Computer Sciences and Mathematics, vol. 16, no. 1, pp. 59–69, 2022.

[18] Y. Meng, D. Kong, Z. Zhu and Y. Zhao, "From Night to Day: GANs Based Low Quality Image Enhancement," Neural Processing Letters, vol. 50, no. 1, pp. 799–814, 2019.

[19] R. C. Gonzalez and R. E. Woods, Digital Image Processing: International Edition, 3rd Ed., Upper Saddle River, NJ: Pearson, 2008.

[20] V. Patrascu and V. Buzuloiu, "Color Image Processing Using Logarithmic Operations," Proc. of the IEEE Int. Symposium on Signals, Circuits and Systems (SCS 2003), DOI: 10.1109/SCS.2003.1226966, 2004.

[21] R. J. Oosterbaan, "Software for Generalized and Composite Probability Distributions," International Journal of Mathematical and Computational Methods, vol. 4, no. 1, pp. 1-19, 2019.

[22] W. Wang, X. Wu, X. Yuan and Z. Gao, "An Experiment-based Review of Low-light Image Enhancement Methods," IEEE Access, vol. 8, pp. 87884–87917, 2020.

[23] E. Baidoo and K. Alex, "Implementation of Gray Level Image Transformation Techniques," Int. J. of Modern Education and Computer Science, vol. 10, no. 5, pp. 44–53, 2018.

[24] V. Bychkovsky, S. Paris, E. Chan and F. Durand, "Learning Photographic Global Tonal Adjustment with a Database of Input/Output Image Pairs," Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR), DOI: 10.1109/CVPR.2011.5995413, Colorado Springs, USA, 2011.

[25] Y. P. Loh and C. S. Chan, "Getting to Know Low-light Images with the Exclusively Dark Dataset," Computer Vision and Image Understanding, vol. 178, pp. 30–42, 2019.

[26] M. R. Lone and A. K. Sandhu, "Enhancing Image Quality: A Nearest Neighbor Median Filter Approach for Impulse Noise Reduction," Multimedia Tools and Applications, pp. 1–17, 2023.

[27] S. Bao, S. Ma and C. Yang, "Multi-scale retinex-based contrast enhancement method for preserving the naturalness of color image," Opt. Rev., vol. 27, no. 6, pp. 475–485, 2020.

[28] A. Mittal, R. Soundararajan and A. C. Bovik, "Making a 'Completely Blind' Image Quality Analyzer," IEEE Signal Processing Letters, vol. 20, no. 3, pp. 209–212, 2013.

[29] A. Mittal, A. K. Moorthy and A. C. Bovik, "No-reference Image Quality Assessment in the Spatial Domain," IEEE Transactions on Image Processing, vol. 21, no. 12, pp. 4695–4708, 2012.

[30] E. H. Land and J. J. McCann, "Lightness and Retinex Theory," J. of the Optical Society of America, vol.

61, no. 1, pp. 1–11, 1971.
[31] W. Burger and M. Burge, Principles of Digital Image Processing: Fundamental Techniques, 1st Edn. London, England: Springer, 2009.
[32] X. Wu, B. Wu, J. He, B. Fang, Z. Shang and M. Zhou, "A Structure Preservation and Denoising Low-light Enhancement Model *via* Coefficient of Variation," Int. J. of Pattern Recognition and Artificial Intelligence, vol. 36, no. 13, DOI: 10.1142/S0218001422540180, 2022.
[33] M. F. Hassan, T. Adam, H. Rajagopal and R. Paramesran, "A Hue Preserving Uniform Illumination Image Enhancement *via* Triangle Similarity Criterion in HSI Color Space," Visual Computer, vol. 39, no. 12, pp. 6755–6766, 2023.
[34] X. Yi, C. Min, M. Shao, H. Zheng and Q. Lv, "Low-light Image Enhancement *via* Regularized Gaussian Fields Model," Neural Processing Letters, vol. 55, no. 9, pp. 12017–12037, 2023.
[35] J. J. Jeon, J. Y. Park and I. K. Eom, "Low-light Image Enhancement Using Gamma Correction Prior in Mixed Color Spaces," Pattern Recognition, vol. 146, p. 110001, DOI: 10.1016/j.patcog.2023.110001, 2024.
[36] J. Li, X. Feng and Z. Hua, "Low-light Image Enhancement *via* Progressive-recursive Network," IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 11, pp. 4227–4240, 2021.
[37] D. Dai and L. Van Gool, "Dark Model Adaptation: Semantic Image Segmentation from Daytime to Nighttime," Proc. of the 21st Int. Conf. on Intelligent Transportation Systems (ITSC), DOI: 10.1109/ITSC.2018.8569387, Maui, USA, 2018.

ملخص البحث:

في هذه الأيام، يُمارِس النّاس نشاطاتٍ في اللّيلِ ويلتقطون العديدَ من الصّور لتوثيق نشاطاتهم. ونظراً لضعْف الإضاءة في البيئة اللّيلية، تميل الصّور الملتقطة ليلاً إلى الظُّهور بإضاءةٍ مُعتِمة وغير متوازنة، وتباينٍ محدود، وضجيجٍ كبير، وألوانٍ غير واضحةٍ.

تقترح هذه الورقة خوارزميةً عمليةً لتحسين الإضاءة في الصُّور اللّيلية باستخدام نموذجٍ يرتكز على طُرقٍ لمعالجة الصّور وعددٍ من الدّوالّ الاحصائية. يبدأ النّموذج المستَخدم بتحويل الصّور المراد معالجتها من نموذج (أحمر أخضر أزرق RGB) إلى نموذج (HSV)، ثم يحسّن فقط قناة (V) مُبقياً على قناة (H) و قناة (S). بعد ذلك، يتمّ تقدير نسخة الإضاءة من الصّورة وحساب خوارزميات كلٍّ من الإضاءة والصّورة الأصلية.

بعدئذٍ، تحدث عملية طرح لوغاريثمي، ويتمّ تطبيق دّالة توزيع تراكمية للإحتمالية، وتكون النتيجة مزيداً من التّحسين باستخدام طريقة نقل لوغاريثمية. وتؤدي هذه العمليات إلى إنتاج قناة (V) المعالجة إلى جانب التّحويل إلى نموذج (RGB) للحصول على المخرج النّهائي.

لقد تمّ تجريبُ النّموذج المقترح باستخدام مجموعتي بياناتٍ، ومقارنته مع عشر خوارزميات واردة في أدبيات الموضوع باستخدام ثلاثةٍ من مؤشرات القياس. وبناءً على نتائج المقارنة، تمّ الحصول على مؤشّرات أداءٍ واعدةٍ للنّموذج المقترح تفوق مؤشرات الأداء في كثيرٍ من الخوارزميات القائمة التي تهدف إلى تحسين الإضاءة في الصّور الملتقطة ليلاً.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 02, June 2024.

SMART PROBABILISTIC ROAD MAP (SMART-PRM): FAST ASYMPTOTICALLY OPTIMAL PATH PLANNING USING SMART SAMPLING STRATEGIES

Muhammad Aria Rajasa Pohan and Jana Utama

(Received: 21-Dec.-2023, Revised: 27-Feb.-2024, Accepted: 14-Mar.-2024)

ABSTRACT

An asymptotically optimal path-planning guarantees an optimal solution if given sufficient running time. This research proposes a novel, fast, asymptotically optimal path-planning algorithm. The method uses five smart sampling strategies to improve the probabilistic road map (PRM). First, it generates samples using an informed search procedure. Second, it employs incremental search techniques on increasingly dense samples. Third, samples are generated around the best solution. Fourth, generated around obstacles. Fifth, it repairs the found route. This algorithm is called the Smart PRM (Smart-PRM). The Smart-PRM was compared to PRM, informed PRM and informed rapidly-exploring random tree-connect. Smart-PRM can generate the optimal path for any test case. The shortest distance between the start and goal nodes is the optimal path criterion. Smart-PRM finds the best path faster than competing algorithms. As a result, the Smart-PRM has the potential to be used in a wide variety of applications requiring the best path-planning algorithm.*

KEYWORDS

Probabilistic road map, Fast asymptotically optimal, Path planning, Intelligent sampling, Informed search.

1. INTRODUCTION

An algorithm for path planning is considered asymptotically optimal if it ensures that it will produce an optimal solution given a sufficient number of iterations or time [1]-[2]. The criteria for best solutions may be based on one or more conditions, such as the lowest fuel usage, lowest risk, comfort or shortest distance [3]. The shortest distance between the initial and goal nodes is used as the criterion for an optimal path in this study. Path-planning algorithms that provide optimal solutions are critical in a wide range of robotic applications [4], including automation processes in industries [5], robot navigation [6], driverless autonomous vehicles [7] and robotic surgery procedures [8]. These examples highlight the significance of optimal path-planning algorithms in addressing diverse robotic applications.

Several researchers have proposed asymptotically optimal path-planning algorithms; however, each algorithm exhibits distinct performance characteristics. One common parameter used to evaluate the performance of path-planning algorithms is the computational time required to generate an optimal path [9]-[11]. Karaman and Frazzoli introduced the Rapidly-exploring Random Tree (RRT*) algorithm, providing an asymptotically optimal solution [12]. Nonetheless, Qureshi et al. [13], J. Nasir et al. [14] and I. B. Jeong et al. [15] reported that the computational speed of RRT* in reaching optimal values still needs improvement. A factor contributing to the computational load of the RRT* algorithm is its necessity to sample throughout the entire search space.

To enhance the performance of the RRT* algorithm, Gammel et al. [16] proposed the Informed RRT* algorithm, which constrains the sampling area based on information from the currently known (yet non-optimal) paths. Wang et al. [17] modified the sampling method to enhance the search speed for an initial solution using a bio-inspired algorithm and an RRT algorithm. Mashayekhi et al. [18] combined the RRT-Connect and informed RRT* algorithms to develop a hybrid RRT approach. It is feasible to obtain the initial solution as rapidly as possible by combining the advantages of the two techniques. Informed RRT* has been coupled with the Dynamic Window Approach (DWA) by Dai et al. [19], while Ryu and Park [20] proposed using a grid-map structure in Informed RRT*. Meanwhile, Wu et al. [21] proposed that raising the APF-IRRT* algorithm's computational speed can assist in identifying the optimal solution faster than other algorithms. Aria [22] proposed updating the technique to become informed RRT*-Connect with local search to increase the informed RRT*'s convergence speed. Path-planning research based on informed sampling is still being developed.

M. A. R. Pohan and J. Utama are with Electrical Engineering Department, Universitas Komputer Indonesia, Bandung, Indonesia. Emails: muhammad.aria@email.unikom.ac.id and jana.utama@email.unikom.ac.id

Another asymptotically optimal path-planning algorithm is the Informed Probabilistic Road Map (PRM) algorithm proposed by the author in [23]. Aria reported that by combining informed searching with the PRM algorithm, the performance of the proposed algorithm can be enhanced by up to 25%. Ongoing research continues to improve the performance of the PRM algorithm. Chen et al. [24] proposed a new PRM sampling strategy to generate more suitable configurations for practical applications. Ravankar et al. [25] suggested the use of a Layered Hybrid PRM with an Artificial Potential Field (APF), while Liu et al. [26] proposed combining the PRM and D* algorithm.

This research proposes a new fast, asymptotically optimal path-planning algorithm called the Smart PRM (Smart-PRM) algorithm. The approach enhances the PRM algorithm through five smart sampling strategies. Test results demonstrate the Smart-PRM algorithm's ability to construct optimal paths across all scenarios. The computational time required for Smart-PRM to generate optimal paths surpasses that of PRM, informed RRT*-Connect and informed PRM algorithms. The Smart-PRM algorithm exhibits efficient convergence due to the incorporation of five smart sampling strategies. These include generating samples using an informed search procedure, employing incremental search techniques on increasingly dense samples, samples generated around the best solution, samples generated around obstacles and the algorithm repairing the found route using a wrapping procedure. The efficacy of each strategy is confirmed through testing, showcasing the Smart-PRM algorithm's potential for implementation in diverse robotic systems and autonomous vehicles.

While it is acknowledged that individual components of our proposed Smart-PRM algorithm draw upon existing techniques in motion planning, we contend that the integration and synergy of these strategies represent a novel and significant advancement in the field. Our approach synthesizes five distinct sampling strategies; namely an informed search procedure, incremental search techniques on increasingly dense samples, sample generation around the best solution, sample generation around obstacles and a route repair mechanism using the wrapping procedure. This amalgamation of strategies not only distinguishes our work, but also facilitates enhanced efficiency and performance compared to existing methods. Furthermore, our experimental results demonstrate a notable improvement in computational time and the ability to construct optimal paths across various scenarios when compared against traditional PRM, informed RRT*-Connect and informed PRM algorithms. The efficiency gains achieved by our Smart-PRM algorithm are particularly noteworthy, surpassing existing methods in terms of convergence speed and solution optimality.

This paper is organized as follows: Section 2 describes the design of the suggested Smart-PRM algorithm. This section describes the strategies used to improve PRM's performance. Section 3 contains the findings and discussion. Initially, the effects of each recommended technique on improving PRM performance are investigated. After that, the suggested Smart-PRM algorithm is compared to PRM, informed RRT*-Connect and informed PRM. Finally, Section 4 includes closing remarks.

2. PROPOSED ALGORITHM: SMART-PRM

The proposed algorithm enhances the PRM algorithm through five strategies. First, it generates samples using an informed search procedure. Second, it employs incremental search techniques on increasingly dense samples. Third, samples are generated around the best solution. Fourth, samples are generated around obstacles. Fifth, it repairs the found route using the wrapping procedure. Thus, the PRM algorithm will be repeated for several iterations. In iterations, before a path solution is found, the second and fourth strategies will be employed. However, after finding a path solution, the fifth, first, third and fourth strategies will be used. Sub-section from 2.1 to 2.5 will discuss each of those strategies. Sub-section 2.6 will discuss the complete algorithm of the proposed Smart-PRM.

2.1 First Strategy: Informed Search Procedure for Sample Generation

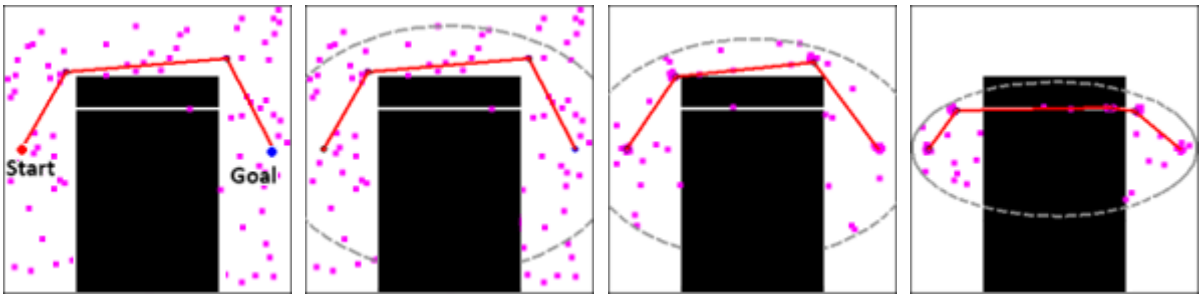
This informed search procedure for sample generation emulates the informed search procedure in the informed RRT* algorithm proposed by Gammel et al. [16]. If a path solution connecting the start and goal nodes is successfully found during an iteration, an area is formed to restrict sample generation. This area takes the shape of an ellipsoid and its eccentricity depends on the length of the shortest-path solution found in that iteration. With the presence of this ellipsoidal area, the sample-generation process in the next iteration will only be carried out within this area. This area enhances the search concentration on

regions with the potential to improve the quality of the path solution. Gammel et al. have demonstrated that once this ellipsoidal area is established, generating samples outside this area does not improve the quality of the path solution.

If a shorter-path solution is found in the next iteration, the size of this ellipsoidal area will decrease and the concentration of the path search will become more focused. Gammel et al. [27] claimed that using this method, the informed RRT* algorithm may obtain an optimal solution approximately 3.4 times faster than the RRT* algorithm.

An illustration of the informed search procedure for sample generation in the PRM algorithm is shown in Figure 1. In the first iteration, sample generation is randomly conducted throughout the area (Figure 1a). Then, using the created-sample nodes, Dijkstra's method [28] is used to find a path connecting the start and finish nodes. An example path successfully created by Dijkstra's algorithm is indicated by the red line in Figure 1a.

Once a path solution is found, an area is established to constrain the sample-generation area, represented by the grey ellipsoid in Figure 1b. Subsequently, the sample generation procedure is applied only within this ellipsoidal area in the next iterations, as shown in Figure 1c. Suppose that a shorter-path solution is found in the following iteration. In that case, the size of this ellipsoidal area will decrease further and the path search will be more concentrated, as depicted in Figure 1d. In the illustration of Figure 1, it can be observed that the optimal solution must pass through a narrow path. Using this first strategy, a solution approaching this optimal path can be achieved by the 10^{th} iteration, as seen in Figure 1d. Therefore, a second strategy for enhancing the PRM algorithm is required to improve the convergence speed, where the search area begins with a small-sized ellipsoidal sub-set.



Iteration: 1
Cost: 182.7
(a)

Iteration: 1
Cost: 182.7
(b)

Iteration: 3
Cost: 173.34
(c)

Iteration: 10
Cost: 147.9
(d)

Figure 1. Illustration of the information-based sample generation process in the Smart-PRM algorithm: (a) Initial random sample generation, (b) Establishment of constraint area based on initial path solution, (c) Subsequent sample generation within the constrained area and (d) Decrease in constraint-area size with successive iterations, leading to a concentrated path search.

2.2 Second Strategy: Incremental Search Techniques on Increasingly Dense Samples

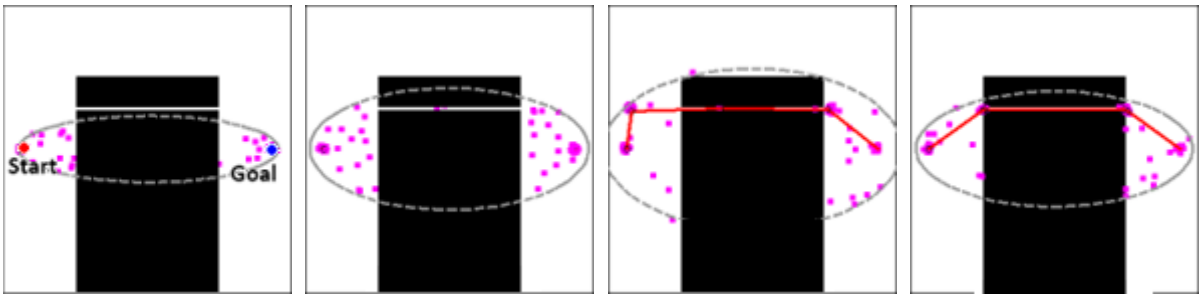
These incremental search techniques on increasingly dense samples emulate the strategies employed in initiating the incremental search techniques on increasingly dense samples within the Batch Informed Tree Star (BIT*) algorithm proposed by Gammell et al. in [27]. This second strategy is distinct from the standard informed RRT* algorithm. During the first iteration of the basic informed RRT* algorithm, no ellipsoidal area constrains the sample-generation area (as illustrated in Figure 1a). However, for the incremental search techniques on increasingly dense samples, initially, sample generation is randomly conducted throughout the entire area. Then, during the first iteration, a small-sized ellipsoidal area is created to restrict only the samples within that ellipsoidal area, which the Dijkstra algorithm will use to find a path connecting the start node with the goal node. If a path solution cannot be obtained by connecting the samples within that small ellipsoidal area, then the ellipsoidal area will be iteratively increased. With the ellipsoidal area growing larger, more dense samples will be within the ellipsoidal area and the Dijkstra algorithm will use more samples to find a path connecting the start node with the goal node.

Once a path solution is found, a new ellipsoidal area, the eccentricity of which depends on the length of that path solution, will be formed. The samples outside this new ellipsoidal area will be removed and

transferred into this new ellipsoidal area, making the number of samples within the new ellipsoidal area denser. This ellipsoidal area will be reduced if a shorter-path solution is obtained and the samples outside the ellipsoidal area will be condensed into the new ellipsoidal area when a shorter-path solution is obtained. Gammell et al. reported that by employing these incremental search techniques on increasingly dense samples, the BIT* algorithm could achieve an optimal solution approximately 6.8 times faster than the RRT* algorithm.

An illustration of this second strategy is depicted in Figure 2. In the first iteration, sample generation is randomly conducted throughout the area. Following that, a small ellipsoidal area is created, as depicted in Figure 2a. The eccentricity of the ellipsoidal area constraining the sample-generation area is determined by a line connecting the start and goal nodes. Since the length of the path connecting the start and goal nodes is unknown in the first iteration, the line determining the eccentricity of the ellipsoid is based on an assumption. An assumption of a straight line connecting the start and goal nodes is used and then, a certain length tolerance is added to that line. This ellipsoidal area will restrict only the samples within it, which the Dijkstra algorithm will use to find a path connecting the start node with the goal node. If a path solution cannot be obtained by connecting the samples within this small ellipsoidal area, then the ellipsoidal area will be iteratively increased, as demonstrated in Figure 2b. With the growing ellipsoidal area, denser samples will be within the ellipsoidal area and the Dijkstra algorithm will utilize more samples to find a path connecting the start node with the goal node.

The procedure of gradually increasing the eccentricity of this ellipsoidal area is repeated until a path connecting the start and target nodes is obtained, as shown in Figure 2c. Once this path solution is discovered, the ellipsoidal area will not be extended in subsequent iterations. Instead, it will be lowered if a shorter-path solution is obtained, as shown in Figure 2d.



Iteration: 1	Iteration: 2	Iteration: 3	Iteration: 5
Cost: -	Cost: -	Cost: 154.8	Cost: 143.57
(a)	(b)	(c)	(d)

Figure 2. Illustration of incremental search techniques on increasingly dense samples in Smart-PRM algorithm: (a) Initial sample generation with a small ellipsoidal area, (b) Iterative expansion of the ellipsoidal area to include denser samples, (c) Finalization of the ellipsoidal area with a path solution and (d) Adjustment of the ellipsoidal area based on path optimization.

2.3 Third Strategy: Sample Generation around the Best Solution

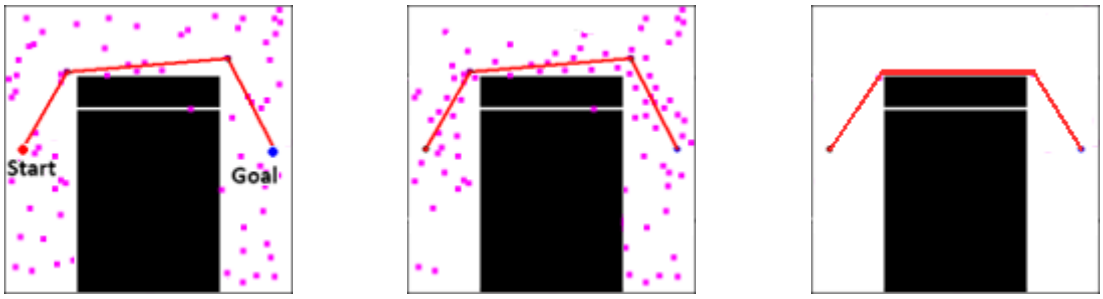
The Smart-PRM algorithm's third strategy focuses on strategically generating samples around the identified best solution during algorithm iterations. This approach aims to refine the obtained path further and leverage the knowledge gained from the informed search.

The Smart-PRM algorithm commences the third strategy once a path solution connecting the start and goal nodes is successfully found. In this strategy, the algorithm utilizes 50% of the sampling points for exploiting the area around this best solution, while the remaining 50% of the sampling points explore the area based on the informed search procedure described in the first strategy.

By concentrating sampling efforts around the best solution, the Smart-PRM algorithm aims to identify alternative paths or variations that may contribute to a more optimal solution. This exploration has the potential to uncover paths that were initially not considered. The approach for exploiting the area around the optimum solution highlights the exploitation process in the RRT-ACS algorithm presented by Pohan et al. in [29]-[30].

An illustration of this third strategy can be seen in Figure 3. Initially, sample generation is conducted

randomly throughout the area. Then, as shown in Figure 3a, Dijkstra's algorithm is used to find a path connecting the start and end nodes using the generated sample nodes. After the path is obtained, some sampling nodes are relocated around the best path. As shown in Figure 3b, there are more sampling nodes around the obtained best path compared to Figure 3a. Therefore, using sampling nodes around the best path has the potential to obtain a more optimal route, as demonstrated in Figure 3c.



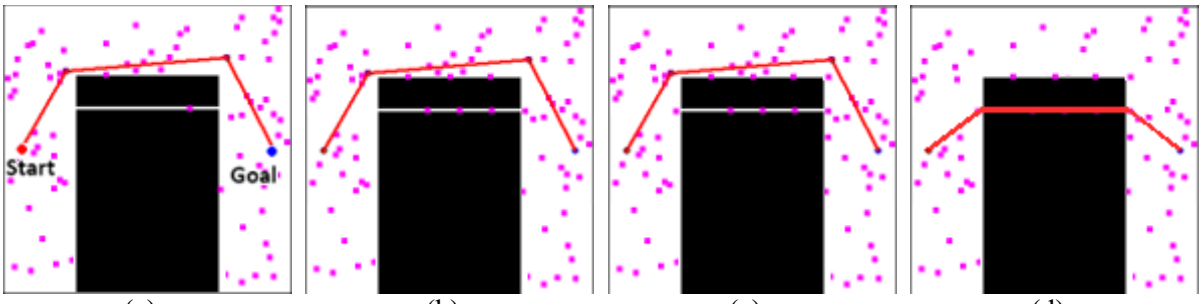
(a) (b) (c)

Figure 3. Illustration of sample generation around the best solution in Smart-PRM: (a) Pathfinding using Dijkstra's algorithm and initial sample generation, (b) Relocation of sampling nodes around the best path and (c) Potential optimization of route with sampling nodes around the best path.

2.4 Fourth Strategy: Sample Generation around Obstacles

The fourth strategy in the Smart-PRM algorithm focuses on strategically generating samples around obstacles encountered in the environment. After encountering newly identified obstacles during iterations, the Smart-PRM algorithm initiates the fourth strategy to systematically use several sampling points to explore and understand the areas around these obstacles. This strategy contributes to creating an optimal path, as optimal paths are often found around obstacles [31].

Strategic sampling around obstacles enhances the algorithm's flexibility and robustness, especially in scenarios where conventional approaches may face difficulties, such as in environments with narrow passages. An illustration of this fourth strategy can be seen in Figure 4. When the algorithm detects samples near an obstacle (purple points in the white gap in Figure 4a), the sides of the obstacle will be explored by more samples (as indicated by three purple points in the white gap in Figure 4b). When a sufficient number of areas on the sides of obstacles are explored by sample points (Figure 4c), there is the potential to discover a better path, as depicted in Figure 4d.



(a) (b) (c) (d)

Figure 4. Illustration of sample generation around the obstacles in Smart-PRM: (a) Detection of samples near obstacles and initial exploration, (b) Increased exploration of obstacle sides by additional samples, (c) Sufficient exploration of areas around obstacles by sample points and (d) Potential discovery of better paths around obstacles.

2.5 Fifth Strategy: Route Repair Using the Wrapping Procedure

The path-correction strategy using the wrapping process emulates the wrapping-based Informed RRT* algorithm discussed in [32]. This wrapping process aims to find a shorter path by creating new nodes close to obstacles. An illustration of this fifth strategy is shown in Figure 5.

In the example case depicted in Figure 5, there is an initial red path consisting of four nodes. The wrapping process begins by creating a temporary node (X_{temp}) at node X_{i+1} or node X_2. Node X_{temp} is connected to node X_1 with a blue line, as shown in Figure 5a. Then, the position of node X_{temp} is

advanced along the path connecting node X_{i+1} to node X_{i+2}, as in Figure 5b. The light blue area indicates the path covered by the blue line connecting X_1 to X_{temp}. The position of node X_{temp} continues to advance until an obstacle obstructs the blue line connecting X_1 to X_{temp}, as shown in Figure 5c. The position where the blue line meets the obstacle is marked as a new node for X_2 (denoted as X_2'). In the next iteration, the position of X_{temp} is advanced again, but because a new node X_2' has been found, the blue line now connects X_{temp} to X_2', as depicted in Figure 5d. The position of X_{temp} continues to advance until it reaches node X_{i+2} or node X_3. Once node X_3 is reached, the position of X_{temp} is further advanced along the path connecting node X_{i+2} to X_{i+3} (or node X_3 to X_4). This process is shown in Figure 5e. If the blue line connecting node X_2' to X_{temp} encounters an obstacle, the position where the blue line meets the obstacle is marked as a new node for X_3 (denoted as X_3'). This iteration continues until node X_{temp} reaches the destination node X_{goal}, as shown in Figure 5f. Figure 5g depicts a comparison of the initial path and the path produced by the wrapping operation. The red line is the original path and the blue line is the corrected/improved path as a result of the wrapping process. Green nodes represent new nodes created during the wrapping process.

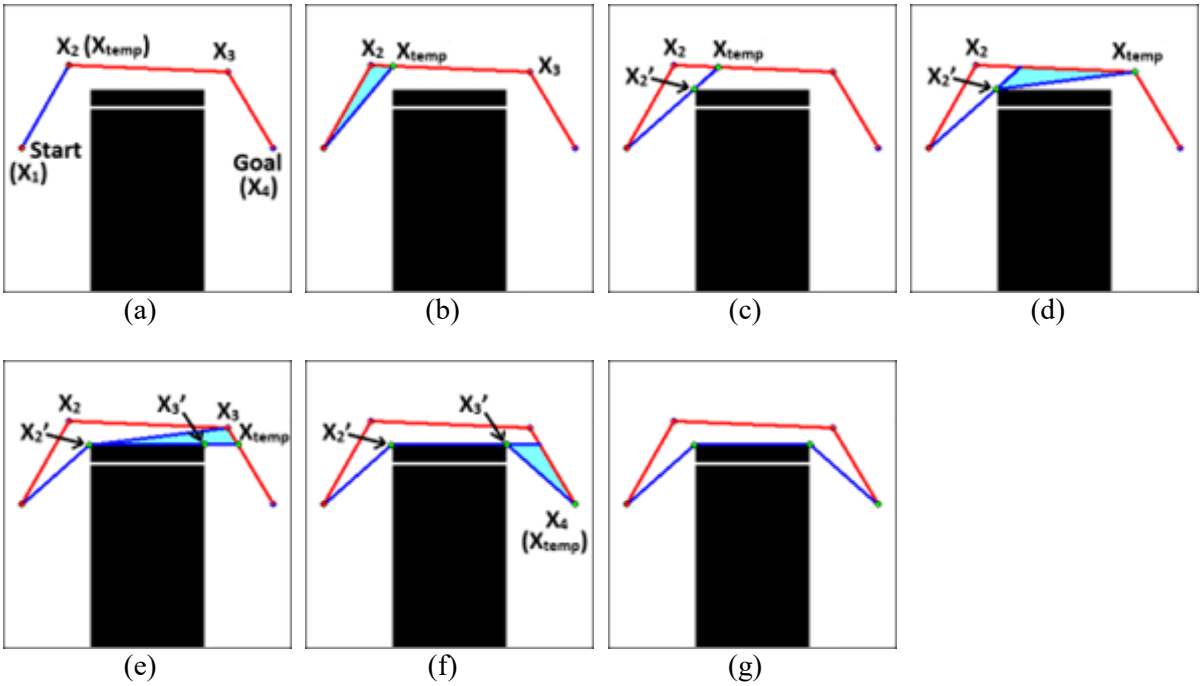


Figure 5. Illustration of the wrapping process to optimize the generated path. The red line represents the initial path, while the blue line represents the repairing/improved path: (a) Creation of temporary node (X_{temp}) and connection to X_1, (b) Advancement of X_{temp} along the path between nodes X_{i+1} and X_{i+2}, (c) Identification of obstacle obstruction and creation of new node X_2', (d) Continued advancement of X_{temp} towards node X_{i+2} or X_3, with connection to X_2', (e) Further advancement of X_{temp} along the path towards node X_{i+2} or X_3, with potential creation of new node X_3', (f) Completion of wrapping process when X_{temp} reaches destination node X_{goal} and (g) Comparison of initial and improved paths resulting from the wrapping operation.

2.6 Comprehensive Overview of the Smart-PRM Algorithm

The complete algorithm proposed is illustrated in Figures 6 and 7. The PRM algorithm consists of sample generation (lines 1-26 in Algorithm 1), roadmap construction (lines 30-37 in Algorithm 1) and path planning (the proposed algorithm uses Dijkstra's algorithm) connecting the start node to the goal node through the generated sample nodes (lines 38-39 in Algorithm 1).

The second strategy of the Smart-PRM algorithm is implemented in lines 3-16 of Algorithm 1. Setting the value of c_{max} to the minimum will create a small-sized ellipsoid subset area. If a path solution in this small area cannot be found, the ellipsoid area will be iteratively enlarged until a path solution connecting the start node to the goal node is found. The expansion process of the ellipsoid area during the path not being found is shown in line 44 of Algorithm 1.

The first strategy of the Smart-PRM algorithm is implemented in lines 17-25 of Algorithm 1 and Algorithm 2. In Algorithm 2, the generation of samples x_{rand} will only be done in the ellipsoid area surrounding x_{start} and x_{goal} with eccentricity depending on the length of c_{max}. Each time the algorithm finds a shorter path, the value of c_{max} will be updated (line 42 of Algorithm 1), therefore, the concentration of path search will increase.

Line 11 of Algorithm 2 implements the Smart-PRM algorithm's third strategy. Lines 27-29 of Algorithm 1 execute the Smart-PRM algorithm's fourth strategy. The fifth strategy of the Smart-PRM algorithm is implemented in Algorithm 1 (line 41).

Algorithm 1. $X_{sol} = (map, x_{start}, x_{goal})$

1. $c_{max} \leftarrow \|x_{goal} - x_{start}\|_2$
2. $X_{sol} \leftarrow \emptyset$
3. **while** $|V_{init}| < n$ **do**
4. **repeat**
5. $x_{rand} \leftarrow$ RandomSampling(map)
6. $q \leftarrow x_{rand}$
7. **until** q **is** collision-free
8. $V_{init} \leftarrow V_{init} \cup \{q\}$
9. **end**
10. **while** $termination\_condition\_not\_meet$ **do**
11. **if** $X_{sol} = \emptyset$ **then**
12. **while** $q \in V_{init}$ **do**
13. **if** q $inside\_ellipsoid\_area$ $(x_{start}, x_{goal}, c_{max})$
14. $V \leftarrow V \cup \{q\}$
15. **end**
16. **else**
17. $V \leftarrow \emptyset$
18. $E \leftarrow \emptyset$
19. **while** $|V| < n$ **do**
20. **repeat**
21. $x_{rand} \leftarrow$ Sample $(x_{start}, x_{goal}, c_{max})$
22. $q \leftarrow x_{rand}$
23. **until** q **is** free of collisions
24. $V \leftarrow V \cup \{q\}$
25. **end**
26. **end**
27. **if** $new\_obstcale\_found$
28. $V \leftarrow V \cup \{q_{around\_obstacle}\}$
29. **end** **for all** $q \in V$ **do**
30. $N_q \leftarrow$ the neighbors of q chosen from V based on $dist$
31. **for all** $q' \in N_q$ **do**
32. **if** (q, q') is free of collisions **then**
33. $E \leftarrow E \cup \{(q, q')\}$
34. **end**
35. **end**
36. **end**
37. $T = (V, E)$
38. $X_{sol} \leftarrow Djikstra(q_{init}, q_{goal}, T)$
39. **if** $X_{sol} \neq \emptyset$ **then**
40. $X_{sol} \leftarrow Wrapping(X_{sol})$
41. $c_{max} \leftarrow min(x_{sol} \in X_{sol})\{Cost(x_{sol})\}$
42. **else**
43. $c_{max} \leftarrow c_{max} \times expansion\_coefficient$
44. **end**
45. **end**

Figure 6. Smart-PRM algorithm.

Algorithm 2. Sample $(x_{start}, x_{goal}, c_{max})$

if $|V| < n/2$ **then**
    $c_{min} \leftarrow \|x_{goal} - x_{start}\|_2$
    $x_{centre} \leftarrow (x_{goal} + x_{start})/2$
    $C \leftarrow$RotationToWorldFrame(x_{start}, x_{goal})
    $r_1 \leftarrow c_{max}/2$
    $\{r_i\}_{i=2,\dots,n} \leftarrow \left(\sqrt{c_{max}^2 - c_{min}^2}\right)/2$
    $L \leftarrow$diag$\{r_1, r_2, \dots, r_n\}$
    $x_{ball} \leftarrow$SampleUnitBall
    $x_{rand} \leftarrow (CLx_{ball} + x_{centre}) \cap X$
else
    $x_{rand} \leftarrow Sampling\_Near\_Best\_Path(X_{sol}, d)$
return x_{rand}

Figure 7. Sample-generation strategy in the smart-PRM algorithm.

3. RESULTS AND DISCUSSION

Several tests were performed to validate the performance of the suggested path-planning algorithm. The first test aimed to verify the effectiveness of the first strategy of Smart-PRM, which generates samples using an informed search procedure. The second test was conducted to confirm the effectiveness of the second strategy of Smart-PRM, which employs incremental search techniques on increasingly dense samples. The third test aimed to verify the effectiveness of the third strategy of Smart-PRM, where samples are generated around the best solution. The fourth test was carried out to confirm the effectiveness of the fourth strategy of Smart-PRM, which generates samples around obstacles. The fifth test was conducted to verify the effectiveness of the fifth strategy of Smart-PRM, which repairs the found route using the wrapping procedure.

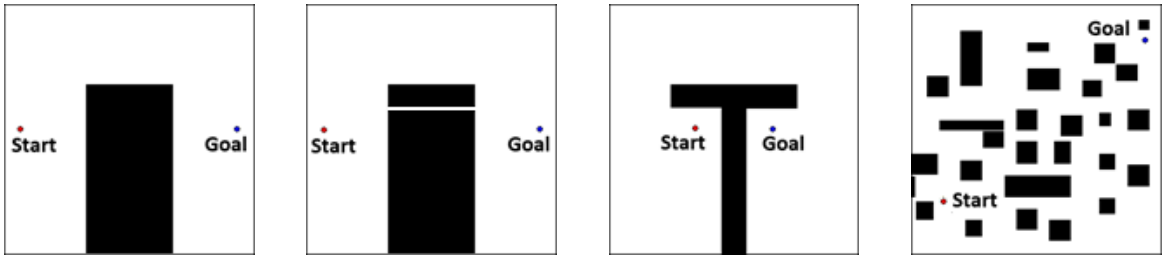
Meanwhile, the sixth test was developed to compare the Smart-PRM algorithm to the PRM algorithm [33], informed RRT*-Connect [18] and informed PRM [23]. The computational time for each approach to attain the optimal result was measured as a performance metric. All tests were done 40 times independently with the identical settings. The comparison was based on each algorithm's average performance across the 40 tests. All tests were carried out on a PC with a Core i5 3.20 GHz CPU and 4 GB RAM running Windows 10 64-bit. The Smart-PRM algorithm and the comparative algorithms were built in LabVIEW 7.1 using the Robotic Path-planning LabVIEW Libraries [34].

3.1 Experimental Scenarios

The proposed Smart-PRM method is compared to existing algorithms to validate its convergence speed and optimality performance. The performance of path-planning algorithms is evaluated using four common scenario cases. There are four scenarios: one with a single obstacle, one with narrow passages, one with a T-shaped obstacle and one with many randomly-scattered obstacles.

The testing scenario with a single obstacle is illustrated in Figure 8a. This scenario assesses whether an algorithm can produce an optimally convergent path. Mashayekhi et al. [18] utilized a testing scenario like this to evaluate their proposed path-planning algorithm. The testing scenario in an environment with narrow passages is depicted in Figure 8b. This scenario is employed to evaluate the effectiveness of path-planning algorithms when the goal node is hidden behind narrow passages. Gammel et al. [16] and Mashayekhi et al. [18] used testing scenarios like this.

The testing scenario in an environment with a T-shaped obstacle is shown in Figure 8c. This scenario assesses the algorithm's effectiveness in handling environments where the generated path needs to navigate turns. Islam et al. [35] used testing scenarios like this. The testing scenario in an environment with multiple randomly-scattered obstacles is illustrated in Figure 8d. This scenario is employed to evaluate the convergence speed of the path-planning algorithm. Gammel et al. [16] used testing scenarios like this.



(a) (b) (c) (d)

Figure 8. Testing scenarios: (a) environment with a single obstacle, (b) environment with narrow passages, (c) environment with T-shaped obstacle, (d) environment with multiple randomly-scattered obstacles.

3.2 Verification of the First-strategy Effectiveness: Informed Search Procedure for Sample Generation

The first test aims to verify the effectiveness of the first strategy; namely, sample generation based on information. The test compares the basic PRM algorithm with the improved PRM algorithm using the first Smart-PRM strategy, which involves generating samples based on information. Testing is performed on the four scenarios mentioned in sub-section 3.1. The measured performance is the computation time of each algorithm to achieve the optimal path. The test results can be seen in Table 1. Furthermore, an analysis of the average-percentage comparison of convergence time to reach the optimal path for both algorithms can be found in Table 2.

Based on the data in Table 2, it can be observed that the average time of the improved PRM algorithm using the first Smart-PRM strategy is 5.49 times faster than the basic PRM algorithm. This result is consistent with the performance measurements of the informed RRT* algorithm (which employs the same algorithm-enhancement strategy) reported by Gammel et al. in [16]. Gammel et al. said that by limiting the sample-acquisition area to the subset ellipsoid area with eccentricity matching the length of the path solution in that iteration, the informed RRT* algorithm becomes 3.4 times faster than the RRT* algorithm in achieving the optimal path. This result verifies the effectiveness of the first strategy, which involves generating samples based on information, in improving the performance of the PRM algorithm.

Table 1. Comparison of improved PRM algorithm using the first strategy against the basic PRM algorithm (in seconds).

Scenario	Convergence time to achieve the optimal path	Improved PRM algorithm using the first strategy	Basic PRM
Scenario I: Single Obstacle	Best	0.13	12.90
	Average	2.20	13.10
	Worst	4.92	13.56
Scenario II: Narrow Passages	Best	0.42	2.79
	Average	1.60	8.61
	Worst	3.71	13.42
Scenario III: T-shaped Obstacle	Best	0.76	10.47
	Average	2.10	14.01
	Worst	6.07	30.97
Scenario IV: Multiple Obstacles	Best	0.49	4.61
	Average	1.55	6.14
	Worst	3.95	13.51

Table 2. Comparison of average convergence time of the improved PRM algorithm using the first strategy against the basic PRM algorithm.

Scenario	Comparison of convergence time (how many times faster)
Scenario I: Single Obstacle	5.95
Scenario II: Narrow Passages	5.38
Scenario III: T-shaped Obstacle	6.67
Scenario IV: Multiple Obstacles	3.96
Average	**5.49**

3.3 Verification of the Second-strategy Effectiveness: Incremental Search Techniques on Increasingly Dense Samples

The second test aims to verify the effectiveness of the second strategy. In this second test, the first strategy is not included; so, the enhancement of the PRM algorithm in this test is solely derived from the second strategy. The test compares the basic PRM algorithm with the improved PRM algorithm using the second S-PRM strategy. Testing is performed on the four scenarios mentioned in sub-section 3.1. The measured performance is the computation time of each algorithm to achieve the optimal path. The test results can be seen in Table 3. Furthermore, an analysis of the average-percentage comparison of convergence time to reach the optimal path for both algorithms can be found in Table 4.

Based on the data in Table 4, it can be observed that the average time of the improved PRM algorithm using the second Smart-PRM strategy is 7.48 times faster than the basic PRM algorithm. This result is consistent with what was reported by Gammel et al. [27] regarding the performance measurements of the BIT* algorithm (which employs a similar strategy to enhance the RRT* algorithm). Gammel et al. reported that by sampling in a small-sized sub-set ellipsoid area first, the BIT* algorithm can achieve an optimal solution 6.8 times faster than the RRT* algorithm. This result verifies the effectiveness of the second strategy; namely, using incremental search techniques on increasingly dense samples.

Table 3. Comparison of improved PRM algorithm using the second strategy against the basic PRM algorithm (in seconds).

Scenario	Convergence time to achieve the optimal path	Improved PRM algorithm using the second strategy	Basic PRM
Scenario I: Single Obstacle	Best	0.08	12.90
	Average	1.36	13.10
	Worst	3.05	13.56
Scenario II: Narrow Passages	Best	0.27	2.79
	Average	1.05	8.61
	Worst	2.41	13.42
Scenario III: T-shape Obstacle	Best	0.89	10.47
	Average	2.45	14.01
	Worst	7.12	30.97
Scenario IV: Multiple Obstacles	Best	0.31	4.61
	Average	0.97	6.14
	Worst	2.45	13.51

Table 4. Comparison of average convergence time of the improved PRM algorithm using the second strategy against the basic PRM algorithm.

Scenario	Comparison of convergence time (how many times faster)
Scenario I: Single Obstacle	9.66
Scenario II: Narrow Passages	8.20
Scenario III: T-shaped Obstacle	5.73
Scenario IV: Multiple Obstacles	6.33
Average	**7.48**

3.4 Verification of the Third-strategy Effectiveness: Sample Generation around the Best Solution

The third test aims to verify the effectiveness of the third strategy. In this third test, neither the first nor the second strategy is included; so, the enhancement of the PRM algorithm in this test is solely derived from the third strategy. The test compares the basic PRM algorithm with the improved PRM algorithm, which is enhanced only by adding the third Smart-PRM strategy. Testing is performed on the four scenarios mentioned in sub-section 3.1. The measured performance is the computation time of each algorithm to achieve the optimal path. The test results can be seen in Table 5. Furthermore, an analysis of the average-percentage comparison of convergence time to reach the optimal path for both algorithms can be found in Table 6.

Based on the data in Table 6, it can be observed that the average time of the PRM algorithm, when

adding the third strategy, is 8.94 times faster than the basic PRM algorithm. This result verifies the effectiveness of the third strategy, which generates a sample around the best solution for improving the performance of the PRM algorithm.

Table 5. Comparison of improved PRM algorithm using the third strategy against the basic PRM algorithm (in seconds).

Scenario	Convergence time to achieve the optimal path	Improved PRM algorithm using the third strategy	Basic PRM
Scenario I: Single Obstacle	Best	0.05	12.90
	Average	1.20	13.10
	Worst	3.51	13.56
Scenario II: Narrow Passages	Best	0.18	2.79
	Average	0.91	8.61
	Worst	2.71	13.42
Scenario III: T-shaped Obstacle	Best	0.52	10.47
	Average	1.72	14.01
	Worst	6.02	30.97
Scenario IV: Multiple Obstacles	Best	0.20	4.61
	Average	0.86	6.14
	Worst	2.82	13.51

Table 6. Comparison of average convergence time of the improved PRM algorithm using the third strategy against the basic PRM algorithm.

Scenario	Comparison of convergence time (how many times faster)
Scenario I: Single Obstacle	10.96
Scenario II: Narrow Passages	9.48
Scenario III: T-shaped Obstacle	8.15
Scenario IV: Multiple Obstacles	7.18
Average	**8.94**

3.5 Verification of the Fourth-strategy Effectiveness: Sample Generation around Obstacles

The fourth test aims to verify the effectiveness of the fourth Smart-PRM strategy. The first, second and third strategies are not included in this fourth test. Therefore, this test's enhancement of the PRM algorithm is solely derived from the fourth strategy. The test compares the basic PRM algorithm with the improved PRM algorithm using the fourth Smart-PRM strategy. Testing is performed on the four scenarios mentioned in sub-section 3.1. The measured performance is the computation time of each algorithm to achieve the optimal path. The test results can be seen in Table 7. Furthermore, an analysis of the average percentage comparison of convergence time to reach the optimal path for both algorithms can be found in Table 8.

Based on the data in Table 8, it can be observed that the average time of the improved PRM algorithm, when using the fourth strategy, is 6.22 times faster than the basic PRM algorithm. This result verifies the effectiveness of the fourth strategy, which involves generating samples around obstacles, in improving the performance of the PRM algorithm.

3.6 Verification of the Fifth-strategy Effectiveness: Route Repair Using the Wrapping Procedure

The fifth test is aimed at verifying the effectiveness of the fifth Smart-PRM strategy. The first, second, third and fourth strategies are not included in this fifth test. Therefore, this test's enhancement of the PRM algorithm is solely derived from the fifth Smart-PRM strategy. The test compares the basic PRM algorithm with the improved PRM algorithm using the fifth strategy. Testing is performed on the four scenarios mentioned in sub-section 3.1. The measured performance is the computation time of each algorithm to achieve the optimal path. The test results can be seen in Table 9. Furthermore, an analysis of the average-percentage comparison of convergence time to reach the optimal path for both algorithms can be found in Table 10.

Table 7. Comparison of improved PRM algorithm using the fourth strategy against the basic PRM algorithm (in seconds).

Scenario	Convergence time to achieve the optimal path	Improved PRM algorithm using the fourth strategy	Basic PRM
Scenario I: Single Obstacle	Best	0.21	25.79
	Average	3.56	26.19
	Worst	7.96	27.12
Scenario II: Narrow Passages	Best	0.69	5.58
	Average	2.65	17.21
	Worst	6.12	26.84
Scenario III: T-shape Obstacle	Best	1.65	20.93
	Average	4.55	28.01
	Worst	13.19	61.93
Scenario IV: Multiple Obstacles	Best	0.79	9.22
	Average	2.52	12.28
	Worst	6.39	27.02

Table 8. Comparison of average convergence time of the improved PRM algorithm using the fourth strategy against the basic PRM algorithm.

Scenario	Comparison of convergence time (how many times faster)
Scenario I: Single Obstacle	7.37
Scenario II: Narrow Passages	6.49
Scenario III: T-shape Obstacle	6.16
Scenario IV: Multiple Obstacles	4.87
Average	**6.22**

Table 9. Comparison of improved PRM algorithm using the fifth strategy against the basic PRM algorithm (in seconds).

Scenario	Convergence time to achieve the optimal path	Improved PRM algorithm using the fifth strategy	Basic PRM
Scenario I: Single Obstacle	Best	0.03	12.90
	Average	1.01	13.10
	Worst	3.98	13.56
Scenario II: Narrow Passages	Best	0.09	2.79
	Average	0.79	8.61
	Worst	3.01	13.42
Scenario III: T-shaped Obstacle	Best	0.15	10.47
	Average	0.90	14.01
	Worst	4.92	30.97
Scenario IV: Multiple Obstacles	Best	0.10	4.61
	Average	0.78	6.14
	Worst	3.20	13.51

Table 10. Comparison of average convergence time of the improved PRM algorithm using the fifth strategy against the basic PRM algorithm.

Scenario	Comparison of convergence time (how many times faster)
Scenario I: Single Obstacle	12.97
Scenario II: Narrow Passages	10.89
Scenario III: T-shaped Obstacle	15.56
Scenario IV: Multiple Obstacles	7.87
Average	**11.82**

Based on the data in Table 10, it can be observed that the average time of the improved PRM algorithm, when using the fifth strategy, is 11.82 times faster than the basic PRM algorithm. This result verifies the effectiveness of the fifth strategy, which involves path refinement using the wrapping process, in improving the performance of the PRM algorithm.

3.7 Analyzing the Contribution of Each Sampling Strategy

Based on Tables 2, 4, 6, 8 and 10, a table illustrating the contribution of each sampling strategy, as demonstrated in Table 11, can be constructed. Table 11 presents a comparison of convergence time using each strategy against the basic PRM algorithm across various scenarios.

Table 11. Comparison of convergence time using each strategy against the basic PRM algorithm across various scenarios.

Scenario	Comparison of convergence time (how many times faster) using each strategy against basic PRM				
	1st Strategy	2nd Strategy	3rd Strategy	4th Strategy	5th Strategy
Scenario I	5.95	9.66	10.96	7.37	**12.97**
Scenario II	5.38	8.20	9.48	6.49	**10.89**
Scenario III	6.67	5.73	8.15	6.16	**15.56**
Scenario IV	3.96	6.33	7.18	4.87	**7.87**
Average	5.49	7.48	8.94	6.22	**11.82**

As depicted in Table 11, which compares the convergence time using each sampling strategy with the basic PRM algorithm across various scenarios, we can evaluate the relative contributions of each strategy to the overall algorithm performance. Upon examining the data, it is evident that based on the test results, the fifth strategy, Route Repair Using the Wrapping Procedure, demonstrates the most significant contribution to achieving superior performance across different scenarios.

3.8 Performance Comparison between the Smart-PRM Algorithm and Other Algorithms

The sixth test compares the Smart-PRM algorithm (which implements all five proposed techniques) to the informed RRT*-Connect and informed PRM algorithms. The test is run on the four scenarios described in sub-section 3.1. The calculation time of each algorithm to find the best path is assessed as performance. Table 12 displays the test results. Table 13 also contains a study of the average-percentage comparison of convergence time to reach the optimal path for both techniques.

Table 12. Comparison of the Smart-PRM algorithm against the informed RRT*-Connect and informed PRM algorithms (in seconds).

Scenario	Convergence time to achieve the optimal path	Smart-PRM	Informed RRT*-connect	Informed PRM
Scenario I: Single Obstacle	Best	0.02	0.56	0.13
	Average	0.60	3.24	2.19
	Worst	1.35	7.16	4.92
Scenario II: Narrow Passages	Best	0.06	1.87	0.42
	Average	0.47	11.63	1.62
	Worst	1.07	31.63	3.71
Scenario III: T-shaped Obstacle	Best	0.10	3.90	0.76
	Average	0.66	6.73	2.09
	Worst	2.70	19.35	6.07
Scenario IV: Multiple Obstacles	Best	0.06	3.52	0.49
	Average	0.43	13.67	1.57
	Worst	1.09	28.79	3.95

Table 13. Comparison of average convergence time of the Smart-PRM algorithm against the informed RRT*-Connect and informed PRM algorithms.

Scenario	Comparison of convergence time (how many times faster)	
	Informed RRT*-Connect	Informed PRM
Scenario I: Single Obstacle	5.36	3.62
Scenario II: Narrow Passages	24.92	3.46
Scenario III: T-shaped Obstacle	10.20	3.16
Scenario IV: Multiple Obstacles	31.78	3.64
Average	**18.06**	**3.47**

According to the statistics in Table 13, the Smart-PRM algorithm has an average time that is 18.06 times faster than the informed RRT* algorithm and 3.47 times faster than the informed PRM algorithm. Therefore, the Smart-PRM algorithm requires less computational time to design an optimal path than the informed RRT* and informed PRM algorithms. The results of the tests show that the Smart-PRM algorithm can create an optimal path in all test scenarios.

3.9 Evaluating the Stability of the Smart-PRM Algorithm

According to Xue [36], a path-planning algorithm is considered stable if it consistently produces the same path when planning the same task. Therefore, we will evaluate the stability of the Smart-PRM algorithm using the data provided in Table 14. Table 14 summarizes the statistical results of performance measurements obtained by Smart-PRM and other algorithms in various benchmark scenarios. Performance measurements include the best-path length, worst-path length, average-path length and standard deviation. A decrease in standard deviation indicates that the cost values of paths generated in each iteration are more consistent. As shown in Table 14, the standard deviation of the Smart-PRM algorithm is the smallest or relatively small compared to the standard deviation of other algorithms in each benchmark scenario. This smaller standard deviation suggests that the Smart-PRM algorithm tends to be more stable compared to other available algorithms

Table 14. Comparison of algorithm stability across various benchmark scenarios. Best results are highlighted for each section.

Scenario	Algorithm	Best	Worst	Mean	Std
Scenario I: Single Obstacle	Smart-PRM	**285.73**	**285.73**	**285.73**	**0**
	Informed RRT*-Connect	**285.73**	286.00	**285.73**	**0**
	Informed PRM	**285.73**	286.00	**285.73**	**0**
Scenario II: Narrow Passages	Smart-PRM	**258.84**	**259.26**	**258.84**	**0.001**
	Informed RRT*-Connect	**258.84**	262.40	259.89	0.004
	Informed PRM	**258.84**	259.89	259.47	**0.001**
Scenario III: T-shaped Obstacle	Smart-PRM	**275.54**	**275.54**	**275.54**	**0**
	Informed RRT*-Connect	277.42	280.70	279.06	0.004
	Informed PRM	**275.54**	278.35	276.24	0.003
Scenario IV: Multiple Obstacles	Smart-PRM	**307.35**	**307.79**	**307.57**	**0.007**
	Informed RRT*-Connect	307.27	314.86	309.41	0.08
	Informed PRM	308.56	311.39	309.78	0.023

3.10 Example Application

As an example of an application requiring fast asymptotically optimal path planning, we find that our algorithm, with its fast convergence, would be highly beneficial in the implementation of autonomous vehicles. The need for algorithms with fast convergence is paramount in traffic-safety contexts, where optimal path planning and rapid response to unforeseen situations are crucial. For instance, in Figure 9, we illustrate a scenario where an autonomous vehicle encounters a curve on the road while pedestrians are crossing unexpectedly. In such situation, autonomous vehicles must be able to respond quickly to plan alternative safe routes and avoid potential accidents. This study can be used as a reference for the current issues in vehicle automation, as discussed in previous studies [37]-[40].

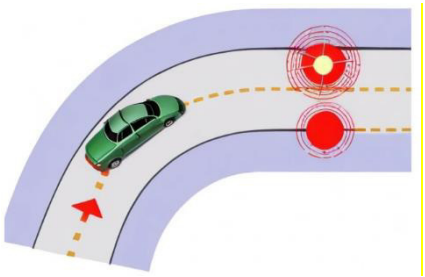


Figure 9. An illustration where autonomous vehicles (green car) must be able to quickly plan alternative routes when sudden changes in environmental conditions are encountered, such as sudden pedestrian crossings (illustrated by the red circle).

4. CONCLUSIONS

This research proposes a new fast, asymptotically optimal path-planning algorithm called the Smart-PRM algorithm. The method is improving the PRM algorithm. The results of the tests reveal that the Smart-PRM algorithm can provide optimal pathways in all test circumstances. The Smart-PRM algorithm takes less processing time to construct an optimal path than the PRM, informed PRM and informed RRT*-connect algorithms. The Smart-PRM algorithm can have good convergence speed, because it uses five smart sampling strategies. First, it generates samples using an informed search procedure. Second, it employs incremental search techniques on increasingly dense samples. Third, samples are generated around the best solution. Fourth, samples are generated around obstacles. Fifth, it repairs the found route using the wrapping procedure. The effectiveness of each strategy has been verified through test results. Thus, the smart-PRM algorithm has the potential to be implemented in various applications that need an optimal path-planning algorithm.

REFERENCES

[1] O. Salzman and D. Halperin, "Asymptotically Near-Optimal RRT for Fast, High-quality Motion Planning," IEEE Transactions on Robotics, vol. 32, no. 3, pp. 473-483, DOI: 10.1109/TRO.2016.2539377, June 2016.

[2] M. Larrañaga, M. Assaad, A. Destounis and G. S. Paschos, "Asymptotically Optimal Pilot Allocation over Markovian Fading Channels," IEEE Transactions on Information Theory, vol. 64, no. 7, pp. 5395-5418, DOI: 10.1109/TIT.2017.2772814, July 2018.

[3] I. Noreen, A. Khan and Z. Habib, "Optimal Path Planning Using RRT* Based Approaches: A Survey and Future Directions," Int. J. of Advanced Computer Science and Applications, vol. 7, no. 11, pp. 97-107, DOI: 10.14569/ijacsa.2016.071114, Jan. 2016

[4] M. A. Hossain and I. Ferdous, "Autonomous Robot Path Planning in Dynamic Environment Using a New Optimization Technique Inspired by Bacterial Foraging Technique," Robotics and Autonomous Systems, vol. 64, pp.137-141, 2015.

[5] C. Friedrich, A. Csiszar, A. Lechler and A. Verl, "Efficient Task and Path Planning for Maintenance Automation Using a Robot System," in IEEE Transactions on Automation Science and Engineering, vol. 15, no. 3, pp.1205-1215, DOI: 10.1109/TASE.2017.2759814, July 2018.

[6] H. Yang, J. Qi, Y. Miao, H. Sun and J. Li, "A New Robot Navigation Algorithm Based on a Double-layer Ant Algorithm and Trajectory Optimization," IEEE Transactions on Industrial Electronics, vol. 66, no. 11, pp. 8557-8566, 2018.

[7] Y. Rasekhipour, A. Khajepour, S. -K. Chen and B. Litkouhi, "A Potential Field-based Model Predictive Path-planning Controller for Autonomous Road Vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 5, pp.1255-1267, DOI: 10.1109/TITS.2016.2604240, May 2017.

[8] M. Diana and J. Marescaux, "Robotic Surgery," British Journal of Surgery, vol. 102, no. 2, pp. e15–e28, DOI: 10.1002/bjs.9711, January 2015

[9] M. Elbanhawi and M. Simic, "Sampling-based Robot Motion Planning: A Review," IEEE Access, vol. 2, pp. 56-77, DOI: 10.1109/ACCESS.2014.2302442, 2014.

[10] Y. Yang, J. Pan and W. Wan, "Survey of Optimal Motion Planning," IET Cyber-systems and Robotics, vol. 1, no. 1, pp.13-19, 2019.

[11] R. Mashayekhi, M. Y. I. Idris, M. H. Anisi and I. Ahmedy, "Hybrid RRT: A Semi-dual-tree RRT-based Motion Planner," IEEE Access, vol. 8, pp.18658-18668, DOI: 10.1109/ACCESS.2020.2968471, 2020.

[12] S. Karaman and E. Frazzoli, "Sampling-based Algorithms for Optimal Motion Planning," Int. J. of Robotics Research, vol. 30, no. 7, pp.846-894, Jun. 2011.

[13] H. Qureshi et al., "Triangular Geometry Based Optimal Motion Planning Using RRT*-motion Planner," Proc. of 2014 IEEE 13th Int. Workshop Adv. Motion Control (AMC), pp.380-385, DOI: 10.1109/AMC.2014.6823312, Yokohama, Japan, 2014.

[14] J. Nasir, F. Islam, U. Malik, Y. Ayaz, O. Hasan, M. Khan and M. S. Muhammad, "RRT*-SMART: A Rapid Convergence Implementation of RRT," Int. Journal of Advanced Robotic Systems, vol. 10, no. 7, p. 299, 2013.

[15] I. B.Jeong, S. J. Lee and J. H. Kim, "Quick-RRT*: Triangular Inequality-based Implementation of RRT* with Improved Initial Solution and Convergence Rate," Expert Systems with Applications, vol. 123, pp. 82-90, 2019.

[16] J. D. Gammell, T. D. Barfoot and S. S. Srinivasa, "Informed Sampling for Asymptotically Optimal Path Planning," IEEE Transactions on Robotics, vol. 34, no. 4, pp. 966-984, Aug. 2018.

[17] J. Wang, W. Chi, M. Shao and M. Meng, "Finding a High-quality Initial Solution for the RRTs Algorithms in 2D Environments", Robotica, vol. 37, no. 10, pp. 1677-1694, 2019.

[18] R. Mashayekhi, M. Y. I. Idris, M. H. Anisi, I. Ahmedy and I. Ali, "Informed RRT*-Connect: An Asymptotically Optimal Single-Query Path Planning Method," IEEE Access, vol. 8, pp. 19842-19852, DOI: 10.1109/ACCESS.2020.2969316, 2020.

[19] J. Dai, D. Li, J. Zhao and Y. Li, "Autonomous Navigation of Robots Based on the Improved Informed-RRT Algorithm and DWA," Journal of Robotics, vol. 2022, Article ID: 3477265, pp.1-9, 2022.

[20] H. Ryu and Y. Park, "Improved Informed RRT* Using Gridmap Skeletonization for Mobile Robot Path Planning," Int. J. of Precision Engineering and Manufacturing, vol. 20, no. 11, pp.2033-2039, 2019.

[21] D. Wu, L. Wei, G. Wang, L. Tian and G. Dai, "APF-IRRT*: An Improved Informed Rapidly-exploring Random Trees-star Algorithm by Introducing Artificial Potential Field Method for Mobile Robot Path Planning," Applied Sciences, vol. 12, no. 21, p. 10905, 2022.

[22] M. Aria, "Path Planning Algorithm Using Informed Rapidly Exploring Random Tree*-Connect with Local Search," J. of Eng. Science and Technology, Special Issue on INCITEST, pp. 50-57, 2020.

[23] M. Aria, "Optimal Path Planning Using Informed Probabilistic Road Map Algorithm," Journal of Engineering Research, ASSEEE Special Issue, pp.1-15, DOI: 10.36909/jer.ASSEEE.16105, 2021.

[24] G. Chen, N. Luo, D. Liu, Z. Zhao and C. Liang, "Path Planning for Manipulators Based on an Improved Probabilistic Roadmap Method," Robotics and Computer-Integrated Manufacturing, vol. 72, pp.102196, DOI: 10.1016/j.rcim.2021.102196, 2021.

[25] A. Ravankar, A. Ravankar, T. Emaru and Y. Kobayashi, "HPPRM: Hybrid Potential Based Probabilistic Roadmap Algorithm for Improved Dynamic Path Planning of Mobile Robots," IEEE Access, vol. 8, pp.221743-221766, DOI: 10.1109/ACCESS.2020.3043333, 2020.

[26] C. Liu, S. Xie, X. Sui, Y. Huang, X. Ma, N. Guo and F. Yang, "PRM-D* Method for Mobile Robot Path Planning," Sensors, vol. 23, no. 7, p. 3512, 2023.

[27] J. D. Gammell, T. D. Barfoot and S. S. Srinivasa, "Batch Informed Trees (BIT*): Informed Asymptotically Optimal Anytime Search," Int. J. of Robotics Research,, vol. 39, no. 5, pp. 543–567, DOI: 10.1177/0278364919890396, Jan. 2020.

[28] M. Iqbal, K. Zhang, S. Iqbal and I. Tariq, "A Fast and Reliable Dijkstra Algorithm for Online Shortest Path," SSRG Int. J. of Computer Science and Engineering, vol. 5, no. 12, pp. 24–27, DOI: 10.14445/23488387/ijcse-v5i12p106, Dec. 2018.

[29] M. A. R. Pohan, B. R. Trilaksono, S. P. Santosa and A. S. Rohman, "Path Planning Algorithm Using the Hybridization of the Rapidly-exploring Random Tree and Ant Colony systems," IEEE Access, vol. 9, pp. 153599–153615, DOI: 10.1109/access.2021.3127635, Jan. 2021.

[30] M. A. R. Pohan and J. Utama, "Efficient Sampling-based for Mobile Robot Path Planning in a Dynamic Environment Based on the Rapidly-exploring Random Tree and a Rule-template Sets," Int. J. of Engineering, vol. 36, no. 4, pp. 797-806, 2023.

[31] C. Scheffer and J. Vahrenhold, "Approximate Shortest Distances among Smooth Obstacles in 3D," J. of Computational Geometry, vol. 10, no. 1, pp. 389-422, 2019.

[32] M. A. R. Pohan, "Asymptoticcaly-Optimal Path Planning Using the Improved Probabilistic Road Map Algorith," Telekontran: Jurnal Ilmiah Telekomunikasi, Kendali and Elektronika Terapan, vol. 10, no. 2 , pp. 116-127, 2022.

[33] M. I. Abdulakareem and F. A. Raheem, "Development of Path Planning Algorithm Using Probabilistic Roadmap Based on Ant Colony Optimization," Engineering and Technology J., vol. 38, no. 3, pp. 343-351, 2020.

[34] M. A. R. Pohan, "LabVIEW Libraries Untuk Algoritma Perencanaan Jalur Robotik," Telekontran: Jurnal Ilmiah Telekomunikasi, Kendali and Elektronika Terapan, vol. 10, no. 1, pp. 47-62, 2022.

[35] J. Nasir, F. Islam, U. Malik, Y. Ayaz, O. Hasan, M. Khan and M. S. Muhammad, "RRT*-SMART: A Rapid Convergence Implementation of RRT," Int. J. of Advanced Robotic Systems, vol. 10, no. 7, p. 299, 2013.

[36] Y. Xue, X. Zhang, S. Jia, Y. Sun and C. Diao, "Hybrid Bidirectional Rapidly-exploring Random Trees Algorithm with Heuristic Target Graviton," Proc. of 2017 IEEE Chinese Automation Congress (CAC), pp. 4357-4361, DOI: 10.1109/CAC.2017.8243546, Jinan, China, 2017.

[37] M. Al-Khalil, R. Abu-Rhayyem, A. Hammoudeh and T. A. Edwan, "Unmanned Ground Vehicle with Virtual Reality Vision,"Jordanian Journal of Computers and Information Technology (JJCIT), vol. 4, no. 1, pp. 58-79, DOI: 10.5455/jjcit.71-1510942341, April 2018.

[38] M. Aria and J. Utama, "An Autonomous Parking System Using the Hybridization of the Rapidly-Exploring Random Trees Star and Ant Colony System," Journal of Advanced Research in Applied Sciences and Engineering Technology, vol. 31, no. 1, pp. 291-297, 2023.

[39] M. A. R. Pohan and J. Utama, "Efficient Autonomous Road Vehicles Local Path Planning Strategy in Dynamic Urban Environment Using RRT-ACS, Bi-directional Rule Templates and Configuration Time-Space," Journal of Engineering Science and Technology, vol. 18, no. 5, pp. 2388-2397, 2023.

[40] M. Aria, "New Sampling Based Planning Algorithm for Local Path Planning for Autonomous Vehicles," J. of Engineering Science and Technology, Special Issue on INCITEST, pp. 66-76, 2019.

ملخص البحث:

يؤدّي التّخطيط لإيجاد المسار الأمثل إلى ضمان الحصول على الحلّ الأمثل إذا ما أعطي النّظام الوقت الكافي للعمل.

يهدف هذا البحث إلى اقتراح خوارزميةٍ جديدةٍ وسريعةٍ للتّخطيط للمسار الأمثل. وتستخدم هذه الطريقة خمس استراتيجيات ذكية لأخذ العينات من أجل تحسين خريطة الطّريق لإيجاد المسار الأمثل. فأولاً، يتمّ إنتاجُ العينات باستخدام إجراءات بحثٍ قائمة على المعلومات. وثانياً، يتمّ استخدام تقنيات بحثٍ متزايدة على العينات متزايدة الكثافة. وثالثاً، يجري إنتاج العينات حوْل الحلّ الأمثل. أمّا الاستراتيجية الرابعة فتتمثل في أخذ العينات حوْل العوائق، بينما تقوم الاستراتيجية الخامسة على إصْلاح المسارِ الذي تمّ إيجاده.

تسمّى الخوارزمية المقترحة في هذا البحث خوارزمية (PRM) الذّكية، وقد جرت مقارنتها بعددٍ من الخوارزميات الأخرى، وتبين أنّ بإمكان الحوارزمية المقترحة إيجاد المسار الامثل لأيّ حالةٍ من حالات الاختبار. وكان معيار الأداء هو أقصر مسافةٍ بين عقدة البداية وعقدة الهدف. وباستطاعة الخوارزمية المقترحة إيجاد المسار الأمثل على نحوٍ أسرع من الخوارزميات الأخرى، وذلك يجعلها مرشّحةً للاستخدام في مدى واسعٍ من التّطبيقات التي تحتاج الى خوارزمية لإيجاد المسار الأمثل، ومنها المركبات ذاتية القيادة على سبيل المثال.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 02, June 2024.

BEYOND WORDS: HARNESSING SPEECH SOUND FOR SPEAKER AGE AND GENDER DETECTION USING 1D CNN ARCHITECTURE WITH SELF-ATTENTION MECHANISM

Umniah Hameed Jaid[1] and Alia Karim Abdulhasan[2]

(Received: 22-Dec.-2023, Revised: 9-Mar.-2024, Accepted: 20-Mar.-2024)

ABSTRACT

Beyond the immediate content of speech, the voice can provide rich information about a speaker's demographics, including age and gender. Estimating a speaker's age and gender offers a wide range of applications, spanning from voice forensic analysis to personalized advertising, healthcare monitoring and human-computer interaction. However, pinpointing precise age remains intricate due to age ambiguity. Specifically, utterances from individuals at adjacent ages are frequently indistinguishable. Addressing this, we propose a novel, end-to-end approach that deploys Mozilla's Common Voice dataset to transform raw audio into high-quality feature representations using Wav2Vec2.0 embeddings. These are then channeled into our self-attention-based convolutional neural network (CNN) model. To address age ambiguity, we evaluate the effects of different loss functions such as focal loss and Kullback-Leibler (KL) divergence loss. Additionally, we evaluate the estimation accuracy at different speech durations. Experimental results from the Common Voice dataset underscore the efficacy of our approach, showcasing an accuracy of 87% for male speakers, 91% for female speakers and 89% overall accuracy, as well as an accuracy of 99.1% for gender prediction.

KEYWORDS

Speaker age, Speaker gender, Speaker profiling, Wav2vec embedding, Attention mechanism.

1. INTRODUCTION

Beyond mere verbal content, the sound of a person's speech offers profound insights into the speaker's identity, revealing hints about age, gender, ethnicity and emotional state [1]. The capability to infer demographic information from speech plays a pivotal role in numerous applications, from forensics [2] to personalized advertising [3]-[4], healthcare systems and human-robot interactions [4].

However, the accurate estimation of demographics from speech is a challenging task. The multifaceted nature of human speech, influenced by factors like emotions, health status, weight and context not only enriches the vocal expressions, but also makes them complex. A particular challenge lies in segregating the textual content from a speaker's physical attributes [5].

Traditionally, the process of speaker profiling has been structured in three stages: data accumulation and preprocessing, feature extraction and selection and finally, the estimation of physical attributes. Historically, voice-pattern analysis has largely relied on time-frequency representations, such as mel-frequency cepstral coefficients (MFCCs) [6], linear predictive coding (LPC) [7] and formant frequencies [8]. However, some studies have leaned towards statistical methods or Gaussian mixture models for speech modeling [9]-[11].

Recent developments in deep-learning (DL) techniques have emerged as powerful tools for identifying complex patterns in data. The multilayered architecture of DL models has demonstrated superior performance in speech processing and speaker-profiling tasks [12]. For instance, long short-term memory (LSTM) networks combined with features like MFCC have been employed for age estimation [13]. Additionally, research by Kalluri et al. [14] and Kaushik et al. [15] delved into the potential of deep neural networks (DNNs) for the estimation of various speaker attributes.

Traditional approaches have relied heavily on handcrafted feature-extraction techniques, such as MFCC and LPC, with classical machine-learning models, resulting in significant limitations in terms of accuracy, generalizability and efficiency. Other studies employed handcrafted features with DL models.

1. U. Jaid is with Department of Computer Science, College of Science, Uni. of Baghdad, Iraq. Email: umniah.h@sc.uobaghad.edu.iq
2. A. Abdulhassan is with Department of CS, University of Technology, Iraq. Email: Alia.K.AbdulHassan@uotechnology.edu.iq

These approaches, while being effective, introduce limitations in capturing the nuanced patterns within speech indicative of age and gender. Handcrafted features play a crucial role in the performance and accuracy of the recognition system; however, their implementation is challenging due to the complexity of feature engineering, as well as the significant time investment needed. Additionally, handcrafted feature extraction can underperform when the manually selected features aren't aligned with the task requirements.

To utilize the rich source of information available in the signal, including spatial cues, several studies adopted the utilization of raw input signal to DL models. Researches used raw signal directly as input to the DL model [16]-[17] or employed hybrid architectures to utilize both the spatial domain of the speech signal with handcrafted features [18]. Several researchers utilized pre-trained models, such as wave2vec and Titanet, to extract features from raw-speech signals directly [19]-[20].

The challenges of age-group prediction are further compounded by the intrinsic diversity of human speech, influenced by factors, such as emotion, health and accent, which can obscure critical demographic indicators. One major limitation to age-group prediction from speech is the ambiguity of the age, where speakers from adjacent age groups are often indistinguishable, due to the gradual change in speech characteristics with age. This problem is further emphasized with data imbalance with more samples in certain age groups than others. One approach to address this problem is the use of distribution learning, emphasizing the model's capability to output probability distributions that reflect the likelihood of each possible outcome, incorporating uncertainty into the predictions [21].

KL-divergence loss naturally accommodates this by comparing the predicted probability distribution against a target distribution that can represent soft labels, improving the model's ability to learn from nuanced differences in speech related to age. Instead of making hard predictions for a specific age group, using KL-divergence encourages the model to output a probability distribution over all possible age groups. This probabilistic approach is beneficial for capturing the uncertainty in age-group prediction, where speech features might not clearly distinguish between adjacent age groups.

Building on the strengths of using raw-speech signals with DL models and the strengths of KL-divergence loss, our proposed model addresses the aforementioned limitations and challenges, by introducing an end-to-end model that integrates Wav2Vec 2.0 embeddings with a self-attention-based CNN, utilizing Mozilla's Common Voice dataset. This methodology not only simplifies the feature extraction process, but also introduces a robust framework capable of discerning subtle age-related variations and gender characteristics in speech. By incorporating the principles of KL-divergence loss within a more comprehensive and advanced modeling approach, we address critical gaps in speaker profiling, including the challenges of age ambiguity and the need for robust, data-driven feature extraction.

In addition to employing KL-divergence loss and raw-speech signal with pre-trained feature extractors, the proposed model employs a self-attention mechanism. Attention mechanisms have recently revolutionized several fields, such as emotion recognition [22], natural-language processing [23] and speech recognition [24], enabling models to focus selectively on parts of the speech signal that are most relevant to the task at hand, by weighting different parts of the input differently, allowing the model to consider the context of the entire speech sequence when making predictions. A specific type of attention, self-attention allows models to capture dependencies and relationships between different parts of the speech signal, regardless of their distance within the sequence. This is particularly beneficial for understanding long-range dependencies in speech, where context from earlier parts of a sequence may influence the interpretation of later parts. This is particularly advantageous for age and gender prediction, where temporal dynamics across the entire speech sequence are analyzed, identifying patterns that are characteristic of different age groups and genders, allowing the model to dynamically focus on segments that are more informative for these predictions.

Additionally, this work provides an insight into the role that loss-function choice plays in the performance of the model, as we compare the performance of the model with several loss functions, such as regular corss-entropy loss, KL-divergence loss that is designed to handle age ambiguity and focal loss that is designed to handle age-group imbalance. A hybrid loss function is introduced in this work, focal-KL to introduce a balance between age-group imbalance and age ambiguity. Further, analyzing the relation between age-group sample size and the accuracy obtained for that age group,

showcases the effectiveness of the loss function in addressing the problem of data imbalance and age-group ambiguity.

The paper also demonstrates the robustness of the proposed model by conducting a thorough investigation into the impact of speech-segment duration on prediction accuracy with varying durations of speech ranging from 1 to 5 seconds of speech. This analysis informs our understanding of the balance between computational efficiency and the quality of our model's predictions.

Our proposed system outperforms existing DNN methods reliant on time-consuming, handcrafted feature extraction. Our work contributes to multi-task age group and gender detection from raw speech and introduces a novel combination of the self-attention mechanism with distributional learning.

Subsequent sections will explore our methodology in detail, present experiment results, compare our findings with existing literature, highlight potential applications and suggest future research directions in speaker profiling.

2. DATASET

To achieve our research objectives, it was crucial to select a dataset that is diverse and comprehensive. In light of this, we chose the Common Voice dataset by Mozilla [25]. This dataset is a crowdsourced, multi-language resource of spoken sentences. The dataset is rich in its demographic diversity, with data collected from speakers of various ages, genders and accents, making it an ideal resource for our research goal.

Each data entry consists of short spoken sentences, textual transcription and the demographic information of the speaker, including age group and gender. The age groups are categorized as 'Teens', 'Twenties', 'Thirties', 'Forties', 'Fifties', 'Sixties', 'Seventies' and 'Eighties and older'. Gender information as self-reported by the contributors is categorized as 'Male', 'Female' and 'Other'. The dataset is continually updated with new contributions; thus, the version used in this work is common_voice_11.

To maintain consistency and avoid ambiguity in the training data, only records marked as ‘Male’ and ‘Female’ were incorporated. Following the completion of data cleaning and removal of empty records, the dataset included 35,846 English-speaking samples from an array of global accents. This diverse collection includes accents from the USA, England, Australia, India, Canada, Malaysia, Scotland, Philippines, Singapore, Hong Kong and several other countries.

The dataset was divided as follows: 33,794 samples for training, 1,511 for validation and 577 for testing. From a gender-distribution perspective, it comprises 25,355 male samples and 8,439 female samples.

A detailed breakdown of the data distribution across various age and gender groups is provided in Table 1.

In the pre-processing phase, the audio data in the dataset originally stored in MP3 format, was converted into waveform samples for compatibility with the Wav2Vec model. For reasons of efficiency and memory management, longer utterances were cropped to 3 seconds, resulting in a maximum length of 48000 at a 16 kHz sampling rate.

Table 1. Description of the common voice dataset used in this work.

Age group	Training		Validation		Testing		Total
	Male	Female	Male	Female	Male	Female	
Teens	1,960	503	48	28	34	13	2,586
Twenties	8,601	1,830	389	87	149	39	11,095
Thirties	6,274	2,155	256	88	107	26	8,906
Forties	4,093	1,033	180	60	67	16	5,449
Fifties	2,307	2,055	116	87	42	32	4,639
Sixties	1,328	795	55	40	18	18	2,254
Seventies	691	58	36	1	14	-	800
Eighties	101	10	4	-	2	-	117
Total	25,355	8,439	1,084	391	433	144	35,846

3. PROPOSED METHOD

In this study, we propose an end-to-end methodology for speaker age and gender detection, leveraging the advanced capabilities of Wav2Vec2.0 for feature extraction from raw-audio signals. This approach eliminates the need for manual feature engineering, allowing the model to automatically learn the most informative aspects of the audio for our tasks.

The proposed methodology is comprised of three key aspects:

- Feature Extraction: The pre-trained Wav2Vec2.0 model is utilized to transform raw audio into high-quality feature representations. This unsupervised-learning technique captures complex speech characteristics essential for distinguishing speaker demographics.
- Self-attention-based CNN: The extracted features are processed through a self-attention-based convolutional neural network. This combination allows our model to dynamically focus on the most relevant parts of the audio signal for age and gender prediction.
- Loss Function Evaluation: To tackle the challenges of age ambiguity and class imbalance, various loss functions are explored, including focal loss and KL-divergence loss.
  A hybrid loss function combining focal loss and KL loss is introduced to offer a mixture for handling class imbalance and age ambiguity. This comparative analysis is crucial for optimizing our model's performance across diverse speech samples.

3.1 Network Architecture

In this study, we propose a novel architecture for audio-based gender and age classification. Our model employs the Wav2vec2.0 transformer-based architecture as an upstream model for feature extraction. Wav2Vec is an unsupervised-learning approach that transforms raw audio into rich, dense vector representations. These embeddings, also known as latent representations, capture significant information from the audio, such as speech content and speaker characteristics. Wav2vec2.0, pre-trained on a large corpus of unlabeled audio data, has demonstrated robustness in extracting meaningful representations from audio signals [26].

The extracted features are then passed through a series of three 1-dimensional convolutional layers, each followed by batch normalization. Each convolutional layer consists of filters of size 3, with the number of filters changing from 512 to 256 to 128 across the layers. The stride of 1 and padding of 1 are maintained in all convolutional layers.

Following feature extraction and convolutional processing, we employ adaptive average pooling with output size 64 to capture global temporal information. The output of the adaptive-pooling layer is then flattened before being passed to a self-attention mechanism. The self-attention mechanism assigns weights to features in the sequence based on their importance, thereby focusing the model's attention on the most informative parts of the audio signal. The attention mechanism consists of a linear transformation followed by a softmax activation function to generate attention scores.

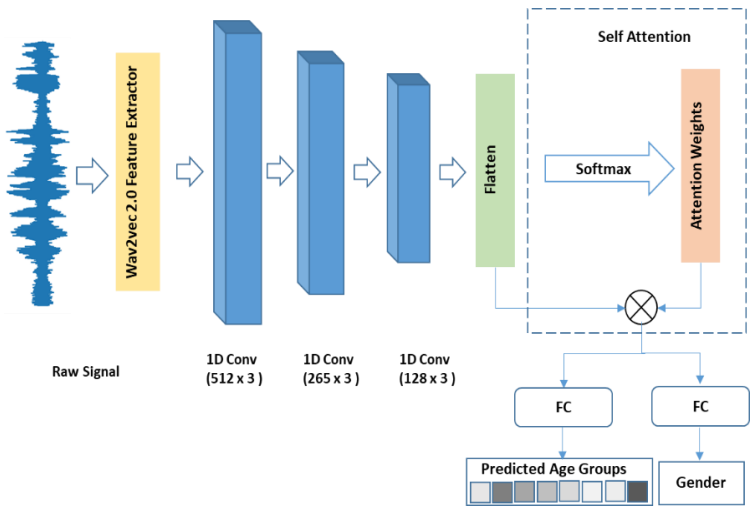


Figure 1. Overview of the proposed architecture.

Post the self-attention mechanism, a dropout layer is applied with a dropout rate of 0.5 to prevent overfitting. Subsequently, the processed features are passed to two separate fully-connected layers for the task of gender and age classification. The gender classifier consists of a linear layer with a single-output unit followed by a sigmoid activation function, classifying the audio clip into either of the two gender categories.

The age classifier, on the other hand, consists of a linear layer with an output size equal to the number of age categories. The proposed model effectively combines the strengths of transformer-based audio representation learning, convolutional processing, adaptive average pooling, self-attention mechanism and task-specific classification layers to perform the dual task of age and gender classification from raw-audio signals.

3.2 Loss Function

In our approach, the model is designed to simultaneously predict both age and gender. Thus, the composite loss function, as shown in Equation (1), merges the individual losses corresponding to age and gender predictions:

$$Loss = age_{loss} + gender_{loss} \tag{1}$$

For gender prediction, the loss is computed using the Mean Squared Error (MSE):

$$MSE_{gender} = \frac{1}{N}\sum_{i=1}^{N}(y_i - y\,pred_i)^2 \tag{2}$$

where N is the total number of predictions, y_i is the actual value of the i^{th} prediction and $y\,pred_i$ is the predicted value of the i^{th} prediction.

Regarding age prediction, we explore various loss functions including cross-entropy, focal loss and KL divergence. These will be detailed in the subsequent sub-sections.

3.2.1 Cross Entropy

Cross Entropy Loss is one of the most widely used loss functions for classification tasks. It measures the dissimilarity between the true label distribution and the predicted probabilities from the model.

Given a classification task with C classes, where each instance is assigned a label in the range [1, C], for a single data point, the predicted probabilities for each class can be determined using the softmax normalization function applied to the model's outputs. The predicted probability p_c of class c is computed as:

$$p_c = \frac{e^{x_c}}{\sum_{j=1}^{C} e^{x_j}} \tag{3}$$

where x_c is the output of the model corresponding to class c.

Given p_c as the probability of the predicted class and y_c as the true label, the cross entropy loss for that data point is defined as:

$$CE(p_c) = -y_c \log(p_c) \tag{4}$$

3.2.2 Focal Loss

While Cross Entropy is effective for many classification tasks, it may not perform as well in scenarios with significant class imbalance. In such cases, the model might become biased towards the majority class, often misclassifying the minority class.

To address this, Focal Loss was introduced as an enhancement over the standard Cross Entropy Loss [27]. It is specifically designed to give more importance to misclassified examples and is especially helpful for imbalanced datasets.

After obtaining the probability of each age class with softmax normalization as in (3), the Focal Loss for a true class c is defined as:

$$FL(p_c) = -\alpha(1 - p_c)^{\gamma} \log(p_c) \tag{5}$$

where p_c is the probability of the true class, α is a scaling factor for the loss and γ is a focusing parameter used to weigh down easy examples.

3.2.3 KL Loss

The proposed model leverages KL divergence (often referred to as the Kullback-Leibler divergence or relative entropy) as the loss function. KL divergence is a measure of how one probability distribution differs from a second, reference probability distribution. It's especially fitting for our problem since our model predicts a distribution over labels, rather than a singular label for each input.

For age-group detection, the KL divergence gauges the dissimilarity between the predicted label distribution and the true label distribution for each instance in the training set.

Given Q as the predicted probabilities for each instance, after softmax normalization, the true label for an instance with label c is represented as a one-hot encoded vector, P, defined as:

$$P = \begin{cases} 1 & if\ i = c \\ 0 & otherwise \end{cases}$$

The KL divergence is then computed as:

$$KL\,(P||Q) = \sum_{i=1}^{C} p_i \log(p_i/q_i)$$

where p_i and q_i are the true and predicted probabilities, respectively, for the i[th] age group.

3.2.4 Focal-KL

The Focal-KL Loss is a hybrid loss that is a combination of the focal loss and the KL divergence loss, which attempts to leverage the benefits of both losses, where Focal Loss addresses the class-imbalance problem by giving more weight to the misclassified examples, while KL Divergence measures the divergence between two probability distributions, making it especially suitable when the model's predictions are distributions over labels.

To create a hybrid loss, we take a linear combination of the Focal Loss and KL Divergence:

$$\text{Focal} - \text{KL} = \boldsymbol{\lambda} \times \text{Focal\_Loss} + (1 - \boldsymbol{\lambda}) \times \text{KL}$$

where λ is a weighting coefficient in the range [0, 1] determining the contribution of each loss. A higher λ gives more weight to the Focal Loss, while a lower λ emphasizes the KL Divergence Loss.

4. EXPERIMENTS AND RESULTS

In this section, we evaluate our model's performance against various benchmarks, different loss functions and input durations to understand its strengths and potential areas of improvement.

To demonstrate the effectiveness of the proposed model, several experiments are performed. The first set of experiments compares the performance of a baseline model with 3 convolutional layers and no attention mechanism and the proposed model in age-group and gender detection. Next, we compare the performance of different loss functions on the proposed model. Finally, duration analysis is performed by performing tests on different durations of the model ranging from one to five seconds. The experiments are performed with a learning rate (1×10^{-6}) and a batch size of 32.

4.1 Self-attention Mechanism

The primary objective here is to discern the impact of incorporating a self-attention mechanism into our model as compared to a baseline model that lacks this feature. To investigate the efficacy of integrating a self-attention mechanism, we compare our proposed model against a baseline architecture. This baseline encompasses three convolutional layers, employs wav2vec for feature extraction and incorporates adaptive pooling. Notably, it lacks the self-attention mechanism characteristic of our proposed design. Both models were trained under identical settings using cross-entropy loss. As presented in Table 2, the inclusion of the self-attention mechanism manifests in marked improvements in age-prediction accuracies for both male and female categories. Conversely, the gender-recognition capability remains consistent across the two models, underscoring the specific advantages of self-attention in age-prediction tasks.

Table 2. Age and gender accuracy of the proposed and baseline models.

	Age Accuracy			Gender
	Male	Female	All	
Baseline	0.72	0.76	0.734	0.98
Proposed	0.76	0.83	0.78	0.98

4.2 Effect of Different Loss Functions

This sub-section aims to evaluate and compare how the model performs when trained with various loss functions, emphasizing the model's adaptability and optimization potential. To demonstrate the effects of the loss function on the model's performance, we evaluate the model with different loss functions; namely, CE, CE with focal loss, Kl divergence loss and a hybrid Kl with focal loss. As seen in Table 3, across the board, it's evident that the model is highly adept at gender classification, achieving an accuracy range of 0.98 to 0.99 regardless of the loss function used. This underscores the robustness of the architecture in distinguishing gender-based audio features.

For age-group detection, using the plain CE, we observed that the model had a higher accuracy for female speakers at 0.841 compared to male speakers at 0.796, yielding an overall average accuracy of 0.807. However, incorporating the focal loss, which is especially effective in addressing class imbalance, shows a marked improvement in performance for both genders. The gap between male and female accuracy narrows, with females achieving a commendable 89.5% accuracy. Switching to the KL divergence loss sees further improvements, especially for female speakers who achieve a 91.5% accuracy. The overall accuracy, taking into account both genders, reaches 86.7%, marking a substantial enhancement over the traditional CE loss.

Combining KL with focal loss produces results that are marginally better than using CE alone, but slightly lag when compared to using either the focal loss or KL divergence loss separately. This could indicate that while both focal and KL loss individually address certain nuances of the dataset, their combination may not necessarily be synergistic for this specific task.

Table 3. Age and gender accuracy of the proposed model using different loss functions.

Loss Function	Age Accuracy			Gender
	Male	Female	All	
CE	76	83	78	98
Focal	84.5	89.5	85.8	98.9
KL	85	91.5	86.7	98.9
Focal_KL	85.4	88.5	86.2	99

4.3 Duration Analysis

In this experiment, we explore how speech input duration influences age-prediction accuracy across different loss functions, aiming to identify the optimal speech duration for accurate predictions. Comparing the age-prediction accuracy of different loss functions at various durations of speech input, it can be seen in Figures 2, 3 and 4 that as the duration of speech input increases, the accuracy tends to increase for all loss functions. This suggests that having more speech data generally results in better age prediction. However, The KL loss seems to consistently provide the highest or one of the highest accuracies across different speech durations. It's especially dominant in the 1-second and 2-second durations. Similarly, Focal-KL shows an interesting trend, where it jumps to 89% at 2 seconds, leading all other methods, but it then aligns more closely with the rest at longer durations.

At 1 second, the KL loss seems to be the most effective with an accuracy of 40%, while other losses are somewhat close, between accuracies of 35% and 38%. However, starting at 2 seconds, there is a significant improvement with KL loss and Focal-KL loss providing the best performance with accuracies of 79.9% and 89%, respectively. At 3 seconds, all loss functions are in the mid-80s range with KL loss leading at 86.7%. The highest accuracy is achieved at the duration of 4 seconds, with KL loss slightly ahead at 88.6%. The performance plateaus at 5 seconds with KL still leading at 88.2% with other losses performing very closely.

In general, between 1 and 2 seconds, there is a large accuracy improvement, jumping from 38% to 89%.

However, between 3 and 5 seconds, there's very minimal improvement across all methods, suggesting a diminishing return of increased speech duration beyond 3 seconds for age prediction with the given model and dataset. Table 4 summarizes the accuracies achieved at 4 seconds.

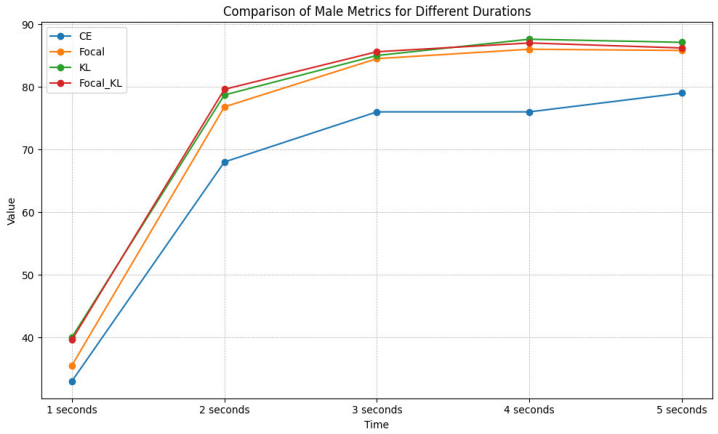


Figure 2. Comparison of male accuracies for different durations.

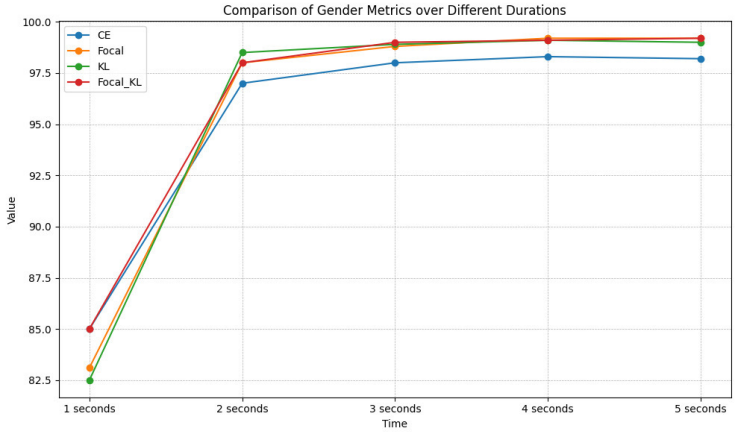


Figure 3. Comparison of female accuracies for different durations.

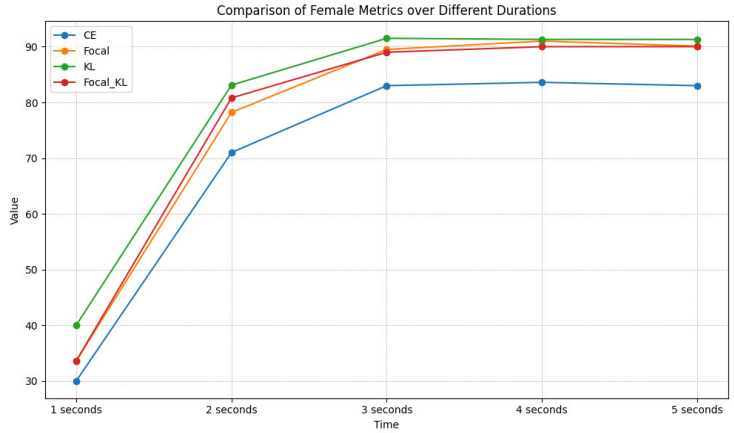


Figure 4. Comparison of gender detection for different durations.

Table 4. Age and gender accuracy of the proposed model at 4-second duration.

Loss Function	Age Accuracy			Gender
	Male	Female	All	
CE	76.7	83.6	78.5	98.3
Focal	86	91	87.3	**99.2**
KL	**87.6**	**91.3**	**89**	99.1
Focal_KL	87	90	87.7	99.1

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 02, June 2024.

4.4 Discussion

In this sub-section, we delve deeper into the results obtained from the experiments, aiming to extract insights and understand patterns in the model's performance. The series of experiments performed in this work not only establishes the effectiveness of our proposed model, but also uncovers intriguing insights regarding age and gender prediction from audio data.

Integrating the self-attention mechanism led to a discernible improvement in age-prediction accuracies for both genders. This enhancement particularly highlights the capacity of the self-attention mechanism to discern age-related attributes in audio data. Contrastingly, the gender-recognition performance remained consistent across the models, implying that the impact of the self-attention mechanism on gender prediction, given the current architectural choices, is relatively limited.

Notably, there's a distinct gender disparity in age accuracy when using the CE loss, pointing to potential inherent biases or differentiating features in the dataset. The introduction of alternative loss functions not only boosts the overall performance, but significantly narrows this gender discrepancy. This is indicative of the effectiveness of these losses in managing potential class imbalances. Combining KL loss and focal loss offered a slight improvement over the individual focal loss; however, it didn't outperform KL loss, suggesting that the performance improvement might be attributed to the KL part of the hybrid loss.

Duration analysis offered an understanding of the relationship between the amount of speech data and prediction accuracy. An evident rapid surge in accuracy with increased duration emphasizes the additional informative value extracted from longer speech samples. Yet, the performance plateau beyond 3 seconds hints at a saturation point, suggesting an optimal duration window that offers the maximum informational value without redundancy.

To provide a comparative perspective on the effectiveness of various methodologies in the field, Table 5 summarizes the classification accuracies achieved by different studies.

Table 5. Comparison of classification accuracies across different studies and numbers of classes.

Study	No. of Classes	Accuracy (All)	Accuracy (Gender)
H. Abdulmohsin et al. [28]	2	87.97%	-
Sánchez-Hevia et al. [29]	6	83.23%	98.24%
D. Kwasny et al. [30]	8	-	99.6%
A. Tursunov et al. [31]	6	73%	96%
Sánchez-Hevia et al. [32]	8	80%	98.14%
Proposed Method	8	89%	99.1%

Our experimental results showcase not only a high degree of accuracy in age and gender detection, but also a significant improvement over existing state-of-the-art methods. Compared to the latest reported accuracies in speaker age-group detection, as reported in Table 5, our model demonstrates a marked increase in precision, especially in distinguishing between closely adjacent age groups—a longstanding challenge in the field. The proposed model achieved an overall accuracy of 89% in age detection and 99.1% in gender detection. Differently from similar studies presented in Table 6 [28]-[29], [31], the age detection accuracy is achieved over eight age groups while similar studies divided the dataset into 2 or 6 classes. These results are notably superior to those of existing models, indicating the effectiveness of our approach in capturing and analyzing the nuanced features of speech that correlate with age and gender.

Figures 5 and 6 show the confusion matrix of the best-performing model with a 4-second duration of the speech and KL loss for age-group and gender prediction respectively. The confusion matrix analysis for age-group classification reveals a detailed performance of the model across various age brackets. For the "teens" group, the model correctly classified 81% of the samples, suggesting a reasonable accuracy, but leaving room for improvement. The model's performance peaks for individuals in their twenties, forties, fifties and sixties with accuracy rates of 89%, 89%, 93% and 94%, respectively. The "thirties" group witnesses a slightly lower accuracy at 87%. Remarkably, the model's efficacy ascends as it approaches the "seventies" age group, achieving a 97% accuracy. However, this trend takes a downturn for the oldest age bracket in the dataset. The "eighties and more" group observes a significant

decline in accuracy of 60%; however, 20% of the misclassified instances were misclassified as the adjacent age group seventies.

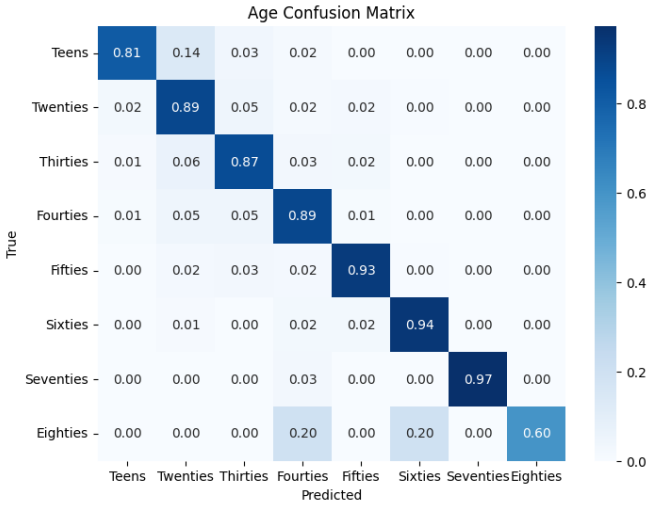


Figure 5. Confusion matrix of age-group prediction of the proposed model.

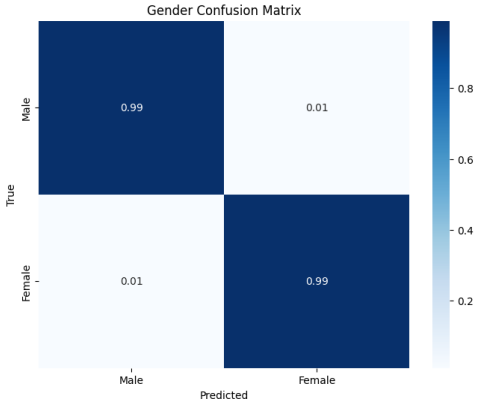


Figure 6. Confusion matrix of gender prediction of the proposed model.

Comparing the number of training and testing instances with the acquired accuracies (Figures 7 and 8) shows that there doesn't appear to be a direct linear relationship between dataset size and accuracy. Larger datasets (like "twenties") don't necessarily have the highest accuracy and smaller datasets (like "seventies") don't necessarily have the lowest accuracy. However, the sharp drop in accuracy for the "eighties and more" group suggests that a minimum threshold of data might be essential for achieving reasonable performance.

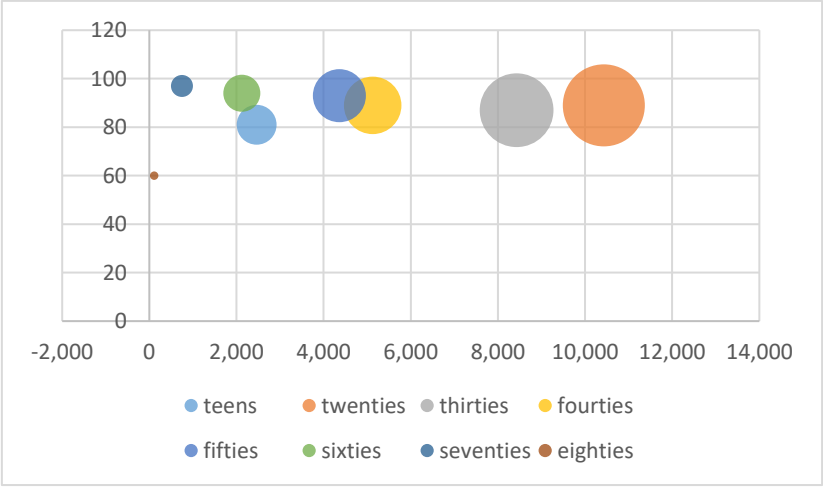


Figure 7. Correlation between obtained accuracies and number of instances in the training dataset.

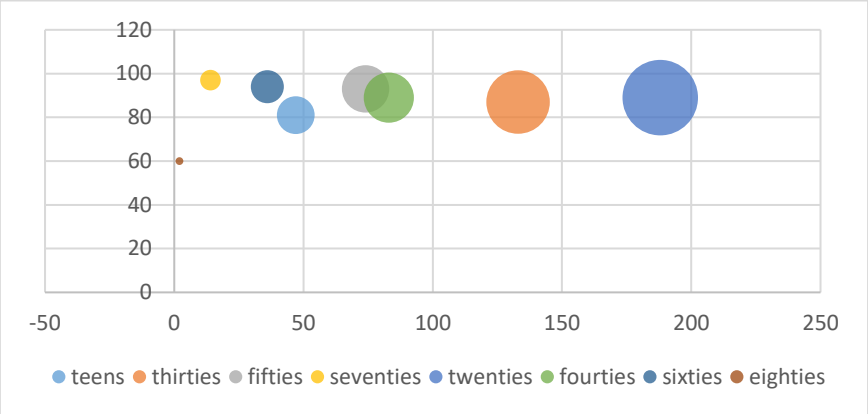


Figure 8. Correlation between obtained accuracies and number of instances in the testing dataset.

Unlike traditional approaches that rely on handcrafted features, our model facilitates an end-to-end learning process by utilizing the Wav2Vec2.0 for feature extraction, benefiting from rich, pre-trained representations of audio data. This unsupervized learning approach allows the model to leverage large amounts of unlabeled audio data, providing a robust foundation for understanding complex speech characteristics without the need for extensive manual feature engineering.

The integration of a self-attention mechanism within the CNN architecture enables the model to dynamically focus on the most informative parts of the audio signal. This aspect is particularly beneficial for age detection, where subtle variations in speech patterns can significantly impact accuracy. Our findings indicate that the self-attention mechanism contributes to a marked improvement in age-prediction accuracies for both male and female speakers.

The proposed model also demonstrates consistent performance across a range of speech durations, from short clips to longer utterances. This versatility suggests that the model can effectively extract and utilize relevant information from audio signals of varying lengths, enhancing its applicability in real-world scenarios where speech samples may not be uniformly sized.

While our proposed method demonstrates promising results in speaker age and gender detection, it is not without limitations. One of the main difficulties lies in the reliance on high-quality, diverse training data. The performance of our model, especially its ability to generalize across different accents, dialects and speech patterns, is heavily dependent on the breadth and depth of the dataset used for training. The Common Voice dataset, while being extensive, may not fully represent the global diversity of speech, potentially limiting our model's applicability in real-world scenarios across various languages and socio-linguistic backgrounds.

Additionally, the computational complexity of our model, driven by the sophisticated feature extraction with Wav2Vec2.0 and the self-attention mechanism, presents a challenge for deployment in low-resource environments or in real-time applications. The balance between model complexity and practical usability is a critical consideration, especially for applications requiring rapid processing or deployment on devices with limited computational capabilities.

Moreover, while our approach addresses age ambiguity to some extent, distinguishing between speakers of closely adjacent age groups remains a challenge. The subtle vocal variations that differentiate age groups may not always be captured or deemed significant by the model, particularly in cases where the training data lacks sufficient examples of such subtle differences.

5. CONCLUSIONS

Our study introduces a novel, end-to-end 1D CNN model for detecting speaker age and gender from speech signals, achieving an overall accuracy of 89% for age groups and a 99.1% accuracy in gender detection, thereby demonstrating significant improvements over traditional methods. This network architecture, built upon three convolutional layers, integrates a self-attention mechanism and leverages direct-speech representations from the advanced pre-trained wav2vec2.0 model, eliminating the need for manual feature extraction. Our evaluation, conducted on the Common Voice dataset comprised of 35,845 speech samples, not only yields promising results in age-group classification and gender

detection, but also showcases the model's versatility by accommodating variable audio lengths. This paves the way for its application in real-world scenarios, particularly enhancing user experiences in mobile devices and human-computer interaction domains where adaptability to varying speech inputs is crucial. The distinct influence of the loss function on model efficacy, with a marked preference for KL and the innovative focal-KL loss functions, underscores the nuanced approach required for optimal performance. Despite the robust performance of our model, the challenge of differentiating between adjacent age groups underscores the complexity of vocal age markers and highlights an avenue for future exploration. Delving deeper into neural-network architectures or innovative feature representations could unveil more granular age-related vocal characteristics. Moreover, expanding our dataset to encompass a broader spectrum of languages, dialects and recording conditions will be imperative for enhancing the model's generalizability and mitigating potential biases.

REFERENCES

[1] G. Assunção, P. Menezes and F. Perdigão, "Speaker Awareness for Speech Emotion Recognition," Int. J. of Online and Biomedical Engineering, vol. 16, no. 4, pp. 15-22, 2020.

[2] A. H. Poorjam and M. H. Bahari, "Multitask Speaker Profiling for Estimating Age, Height, Weight and Smoking Habits from Spontaneous Telephone Speech Signals," Proc. of the 2014 4th IEEE Int. Conf. on Computer and Knowledge Engineering (ICCKE), Mashhad, Iran, pp. 7-12, 2014.

[3] C. Müller, "Automatic Recognition of Speakers2 Age and Gender on the Basis of Empirical Studies," Proc. of the 9th Int. Conf. on Spoken Language Processing (Interspeech 2006), pp. 2118–2121, paper 1031-Wed3CaP.11, DOI: 10.21437/Interspeech.2006-195, 2006.

[4] C. Müller and F. Burkhardt, "Combining Short-term Cepstral and Long-term Pitch Features for Automatic Recognition of Speaker Age," Proc. of the 8th Annual Conf. of the Int. Speech Communication Association, (Interspeech 2007), pp. 2277–2280, Antwerp, Belgium, 2007.

[5] S. B. Kalluri, A. Vijayakumar, D. Vijayasenan and R. Singh, "Estimating Multiple Physical Parameters from Speech Data," Proc. of the 2016 IEEE 26th Int. Workshop on Machine Learning for Signal Processing (MLSP), pp. 1-5, Vietri sul Mare, Italy, 2016.

[6] S. Galgali, S. S. Priyanka, B. Shashank and A. P. Patil, "Speaker Profiling by Extracting Paralinguistic Parameters Using Mel Frequency Cepstral Coefficients," Proc. of 2015 IEEE Int. Conf. on Applied and Theoretical Computing and Communic. Technology (iCATccT), pp. 486-489, Davangere, India, 2015.

[7] A. A. Badr and A. K. Abdul-Hassan, "Estimating Age in Short Utterances Based on Multi-class Classification Approach," Computers, Materials & Continua, vol. 68, no. 2, pp. 1713-1729, 2021.

[8] I. Mporas and T. Ganchev, "Estimation of Unknown Speaker's Height from Speech," International Journal of Speech Technology, vol. 12, no. 4, pp. 149-160, DOI: 10.1007/s10772-010-9064-2, 2010.

[9] K. A. Williams and J. H. Hansen, "Speaker Height Estimation Combining GMM and Linear Regression Subsystems," Proc. of 2013 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2013), pp. 7552-7556, 2013.

[10] H. Arsikere, G. K. F. Leung, S. M. Lulich and A. Alwan, "Automatic Estimation of the First Three Subglottal Resonances from Adults Speech Signals with Application to Speaker Height Estimation," Speech Communication, vol. 55, no. 1, pp. 51-70, DOI: 10.1016/j.specom.2012.06.004, 2013.

[11] A. A. Mallouh, Z. Qawaqneh and B. D. Barkana, "New Transformed Features Generated by Deep Bottleneck Extractor and a GMM-UBM Classifier for Speaker Age and Gender Classification," Neural Computing & Applications, vol. 30, no. 8, pp. 2581-2593, DOI: 10.1007/s00521-017-2848-4, 2018.

[12] O. Buyuk and M. L. Arslan, "Combination of Long-term and Short-term Features for Age Identification from Voice," Advances in Electrical and Computer Engineering, vol. 18, no. 2, pp. 101-108, 2018.

[13] R. Zazo et al., "Age Estimation in Short Speech Utterances Based on LSTM Recurrent Neural Networks," IEEE Access, vol. 6, pp. 22524-22530, DOI: 10.1109/access.2018.2816163, 2018.

[14] S. B. Kalluri, D. Vijayasenan and S. Ganapathy, "A Deep Neural Network Based End to End Model for Joint Height and Age Estimation from Short Duration Speech," Proc. of ICASSP 2019 - 2019 IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 6580-6584, Brighton, UK, 2019.

[15] M. Kaushik, V. T. Pham and E. S. Chng, "End-to-End Speaker Height and Age Estimation Using Attention Mechanism with LSTM-RNN," arXiv preprint arXiv: 2101.05056, 2021.

[16] S. Kwon, "1D-CNN: Speech Emotion Recognition System Using a Stacked Network with Dilated CNN Features," Computers, Materials & Continua, vol. 67, no. 3, pp. 4039-4059, 2021.

[17] U. H. Jaid and A. K. AbdulHassan, "End-to-End Speaker Profiling Using 1D CNN Architectures and Filter Bank Initialization," Int. J. of Online & Biomedical Engineering, vol. 19, no. 10, 2023.

[18] Mustaqeem and S. Kwon, "Optimal Feature Selection Based Speech Emotion Recognition Using Two-stream Deep Convolutional Neural Network," Int. J. of Intellig. Syst., vol. 36, no. 9, pp. 5116-5135, 2021.

[19] M. Z. Tarashandeh, A. Torkanloo and M. H. Moattar, "AgeNet-AT: An End-to-End Model for Robust

Joint Speaker Age Estimation and Gender Recognition Based on Attention Mechanism and Titanet," Proc. of the 2023 13th IEEE Int. Conf. on Computer and Knowledge Engineering (ICCKE), pp. 414-419, Mashhad, Iran, 2023.

[20] T. Gupta, D.-T. Truong, T. T. Anh and C. E. Siong, "Estimation of Speaker Age and Height from Speech Signal Using Bi-encoder Transformer Mixture Model," arXiv preprint, arXiv: 2203.11774, 2022.

[21] S. Si, J. Wang, J. Peng and J. Xiao, "Towards Speaker Age Estimation with Label Distribution Learning," arXiv preprint, arXiv: 2202.11424, 2022.

[22] S. Kwon, "Att-Net: Enhanced Emotion Recognition System Using Lightweight Self-attention Module," Applied Soft Computing, vol. 102, p. 107101, 2021.

[23] A. Galassi, M. Lippi and P. Torroni, "Attention in Natural Language Processing," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 10, pp. 4291-4308, 2020.

[24] N.-Q. Pham, T.-S. Nguyen, J. Niehues, M. Müller, S. Stüker and A. Waibel, "Very Deep Self-attention Networks for End-to-End Speech Recognition," arXiv preprint, arXiv:1904.13377, 2019.

[25] R. Ardila et al., "Common Voice: A Massively-multilingual Speech Corpus," arXiv: 1912.06670, 2019.

[26] A. Baevski et al., "Wav2vec 2.0: A Framework for Self-supervised Learning of Speech Representations," Advances in Neural Information Processing Systems, vol. 33, pp. 12449-12460, 2020.

[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection," Proc. of the IEEE Int. Conf. on Computer Vision, pp. 2980-2988, 2017.

[28] H. A. Abdulmohsin, J. J. Stephan, B. Al-Khateeb and S. S. Hasan, "Speech Age Estimation Using a Ranking Convolutional Neural Network," Proc. of Int. Conf. on Computing and Communication Networks (ICCCN 2021), pp. 123-130, Springer, 2022.

[29] H. A. Sánchez-Hevia, R. Gil-Pita, M. Utrilla-Manso and M. Rosa-Zurera, "Age Group Classification and Gender Recognition from Speech with Temporal Convolutional Neural Networks," Multimedia Tools and Applications, vol. 81, no. 3, pp. 3535-3552, 2022.

[30] D. Kwasny and D. Hemmerling, "Gender and Age Estimation Methods Based on Speech Using Deep Neural Networks," Sensors, vol. 21, no. 14, p. 4785, 2021.

[31] A. Tursunov, Mustaqeem, J. Y. Choeh and S. Kwon, "Age and Gender Recognition Using a Convolutional Neural Network with a Specially Designed Multi-attention Module through Speech Spectrograms," Sensors, vol. 21, no. 17, p. 5892, 2021.

[32] H. A. Sánchez-Hevia, R. Gil-Pita, M. Utrilla-Manso and M. Rosa-Zurera, "Convolutional-recurrent Neural Network for Age and Gender Prediction from Speech," Proc. of the 2019 IEEE Signal Processing Symposium (SPSympo), pp. 242-245, Krakow, Poland, 2019.

ملخص البحث:

يمكن أن يزودنا الصّوت بمعلوماتٍ غنية عن بعض الخصائص الشخصية للمتكلّم، بما في ذلك عمره وجنسه. وإنّ تقدير عمر المتكلّم وجنسه يوفر لنا مدىً واسعاً من التّطبيقات يمتدّ من التّحليل الشّرعي للصّوت إلى الإعلان المشخّص والرّصد المتعلّق بالرّعاية الصّحية وتفاعل الإنسان مع الحاسوب. ومع ذلك، يبقى التّحديد الدّقيق للعُمْر أمراً مَشوباً بالصُّعوبة والغموض. ويتعلّق الأمر بتشابه سِمات أصوات الأشخاص ذوي الأعمار المتقاربة إلى درجةٍ يصْعب معها تمييزها. ولمعالجة هذا الأمر، نقترح نموذجاً يستخدم ما يعرف بمجموعة بيانات الصّوت العام (Common Voice) لتحويل الصّوت الخام إلى تمثيلاتٍ للخصائص عالية الجودة. ومن ثمّ يجري إدخال هذه الخصائص إلى سلسلة من الشّبكات العصبية الالتفافية. وللتّغلُّب على غُموض العُمر، فإنّنا نعمل على تقييم آثار مجموعة متنوعة من دوالّ الفقْد على دقّة النّموذج في تقدير إشاراتٍ صوتية مختلفة الفترة الزمنية لبقائها.

وقد أثبتت التّجارب التي أجريناها على مجموعة بيانات (الصّوت العام) نجاعة النّموذج المقترح وتفوقّه على عددٍ من النّماذج المماثلة الواردة في أدبيات الموضوع؛ فقد حقّق النّموذج المقترح في تقدير العُمْر دقّةً وصلت إلى 87% للمتكلمين الذّكور، و 91% للمتكلمات الإناث، بدقةٍ إجمالية بلغت 89%، بينما بلغت الدّقة الإجمالية فيما يتعلّق بجنس المتكلّم 99.1%.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 02, June 2024.

APPLYING TOGAF-BASED ENTERPRISE ARCHITECTURE IN THE HEALTHCARE SECTOR: A CASE STUDY OF THE NATIONAL CENTER FOR DIABETES IN JORDAN

Hania Al Omari[1], Abedalrhman Alkhateeb[2] and Bassam Hammo[1]

(Received: 19-Jan.-2024, Revised: 27-Mar.-2024 and 20-Apr.-2024, Accepted: 24-Apr.-2024)

ABSTRACT

Technology implementation can significantly benefit organizations, but ensuring that it aligns with their business needs and goals is crucial. Adopting an enterprise-architecture approach can aid healthcare enterprises in overcoming challenges during the transformation process. In particular, this study examines how The Open Group Architecture Framework (TOGAF) can facilitate digital transformation while ensuring alignment with business needs. Using the Architecture Development Method (ADM) of TOGAF, the study analyzes the current architecture of the National Center for Diabetes, Endocrinology and Genetics (NCDEG) in Jordan, intending to develop a target architecture that helps NCDEG effectively achieve its goals by aligning technology implementations with business objectives. Utilizing TOGAF's ADM, the study navigates the complexities of technological advancements while ensuring seamless integration and effective utilization of resources. Furthermore, the findings highlight the critical role of enterprise architecture in facilitating organizational evolution, emphasizing the need for continuous evaluation and refinement to adapt to changing business landscapes and technological advancements for NCDEG and similar organizations. The proposed changes were validated through simulation using Rockwell Arena Simulation Software. Results showed significant improvements in patient handling, process efficiency, waiting times and resource utilization by implementing virtual clinics and digital solutions.

KEYWORDS

Digital transformation, Enterprise architecture, Healthcare informatics, TOGAF, Diabetes.

1. INTRODUCTION

Enterprise Architecture (EA) is a practical approach to strategically managing an organization's technology landscape. By aligning technology with business goals, EA ensures that suitable applications and technologies support business processes. As organizations grow and evolve through mergers and acquisitions, EA must govern and guide new projects, systems and processes added to the technology ecosystem. This practice helps control costs by eliminating duplication and ensuring standardization across processes and technologies [1]-[2].

Several factors have recently influenced the healthcare industry, making embracing digital transformation in services and operations imperative. These factors include restrictions, social distancing and the immense strain on the healthcare sector due to the COVID-19 pandemic. Additionally, emerging technologies, such as the Internet of Medical Things (IoMT), mobile health apps, artificial intelligence (AI) and big data have played a crucial role [3].

To keep pace with the digital era, healthcare providers must incorporate modern technologies into their existing systems or transition traditional practices to digital ones. To do this, a thorough assessment of their current status is required, as well as the development of a comprehensive digital transformation plan that aligns with their objectives and purpose.

Organizations can adopt EA through different approaches or frameworks. The most popular frameworks are TOGAF, The US Federal EA Framework (FEAF), the US Department of Defence Architecture Framework (DODAF), Zachman and the UK Ministry of Defence Architecture Framework (MODAF), which was withdrawn in 2021 and replaced by the NATO Architecture Framework (NAF) [1]. Healthcare institutions may benefit from implementing an EA to facilitate and oversee their transformation efforts.

1. H. Alomari and B. Hammo (corresponding author) are with the Software Engineering Department, Princess Sumaya University for Technology, Amman, Jordan. Emails: han20208079@std.psut.edu.jo and b.hammo@psut.edu.jo

2. A. Alkhateeb is with the Computer Science Department, Lakeahead University, Canada. Email: aalkhate@lakeheadu.ca

This study explores the adoption of TOGAF in the healthcare sector, specifically in the case study of the National Centre of Diabetes, Endocrinology and Genetics (NCDEG) in Jordan [4]. In addition, it serves as a milestone to motivate healthcare providers in Jordan to consider EA a valuable tool for guiding and governing their digital-transformation initiatives.

The following sections introduce the significance of technology implementation in healthcare organizations and the challenges that they face in aligning technology with business objectives. We also discuss the rationale behind selecting The Open Group Architecture Framework (TOGAF) as the methodology for this study in the Background and Related Work sections. Collecting the required data and applying the framework are detailed in the Research Methodology section. The simulation's setup and running are discussed in the Results and Analysis section and finally, the Conclusion section summarizes the study.

2. BACKGROUND

2.1 The Open Group Architecture Framework (TOGAF) Standard

The TOGAF Standard, initially released in 1995, is an EA framework widely used to assist organizations in developing EA for their entire organization or specific parts of it based on their needs [4]. This framework can be used in its entirety or tailored to suit the objectives of the EA. According to the TOGAF Standard, EA's primary goal is to help organizations enhance and integrate their processes, allowing them to better respond to change and support their business strategy. Additionally, it can be advantageous for organizations seeking to establish a seamless data flow within or among multiple organizations. The TOGAF Standard can be accessed online for free or organizations can obtain a licensed copy for downloading and storage purposes [5].

The diagram presented in Figure 1 depicts the central component of the TOGAF Standard 9.2 ADM [6]. The preliminary phase is responsible for preparing and adjusting the ADM and is continuously repeated [6]. It identifies the relevant units impacted by EA and the stakeholders and governance involved. This phase also establishes the architecture-governance framework and any additional support frameworks necessary for managing the architectural materials and the relationship between governance processes and ownership of architectural artifacts. Moreover, it defines the architecture principles, such as business, data, application and technology principles, that are crucial for effective architecture governance. The remaining ADM phases include the following:

A. Architecture Vision phase: develops the vision of the capabilities and business value that the proposed architecture would achieve. In addition, it approves the work plan required to build and deploy the proposed architecture.
B. Business Architecture phase: describes how the business operates or needs to operate to achieve the business goals.
C. Information-system Architecture phase: describes how the information-system architectures (data architecture and application architecture) enable the achievement of architecture vision and business architecture.
D. Technology Architecture phase: describes the target-technology architecture that enables the achievement of architecture vision, target business and information-system architectures.
E. Opportunities and Solutions phase: sets the foundation for delivering the architectures, including the migration and implementation plan, which provides the timeline for the projects required to produce the target architectures.
F. Migration Planning phase: finalizes the migration and implementation plan required to deliver the target architectures.
G. Implementation Governance phase: ensures that the implementation or any ongoing projects within the enterprise conforms with the target architectures.
H. Architecture Change Management phase: describes managing any changes to the new architectures.

The core of the ADM diagram, the Requirements Management phase, is a crucial part of ADM, as it identifies and stores the EA requirements fed into other phases. Throughout each phase, enterprise architects should pinpoint functional and non-functional requirements. These requirements dictate what the architecture should meet. The Architecture Requirements Repository contains all approved

architecture requirements, new-architecture requirements, out-of-scope architecture requirements and any changes to these requirements.

2.2 The National Center for Diabetes, Endocrinology and Genetics

The National Center for Diabetes, Endocrinology and Genetics (NCDEG) was established in 1996 as one of the centers affiliated with the Higher Council for Science and Technology [9]. Diabetes Mellitus (DM) is a chronic disease of inadequate control of blood levels of glucose; it has many subclassifications, including type 1 and type 2 [7]. Type 2 is considered the most common type and is regarded as an epidemic in many countries. The Middle East and North Africa Region (MENA) recorded the highest level of DM worldwide in 2019, with 12.2% of its population [8]. DM can lead to severe health conditions if left unmanaged; heart and blood vessel disease, kidney failure and retinal diseases are among the consequences of DM [9].

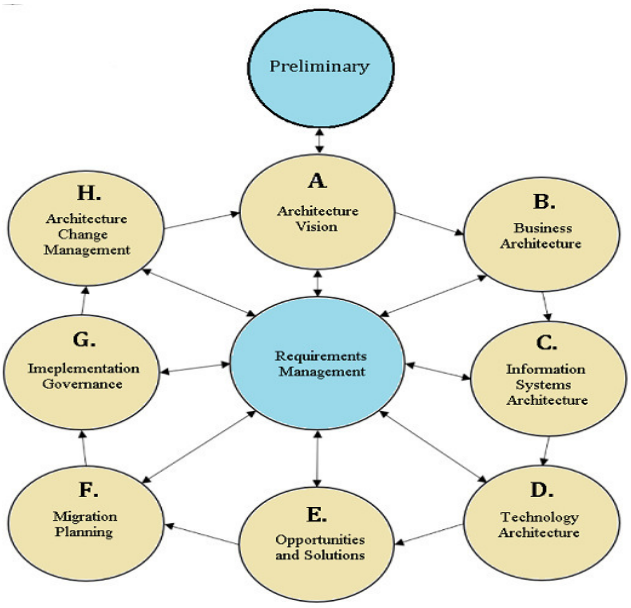


Figure 1. TOGAF Standard ADM [10].

In Jordan, 15% of the Jordanian population (aged 20-79) have been diagnosed with diabetes [11]. The prevalence of DM in the country has steadily increased; for example, it reached 32.4 % of males over 25, representing 125% of its prevalence among the same population in 1994 [12]. It is also projected that a fifth of Jordanian people will have type-2 DM by 2050 [13].

NCDEG is the only specialized healthcare provider in Jordan that provides comprehensive treatment services for DM, its complications, endocrine diseases and genetic disorders. These services include DM-specialized clinics, Endocrinology clinics, Genetics clinics, Diabetic Foot Care Clinics and other specialized clinics. In addition, NCDEG has a Radiology, Imaging and Nuclear Medicine department, four labs and a specialized pharmacy dedicated to providing medications for DM and its complications, endocrine disease and genetic disorders.

Given the substantial demands placed on the services offered by NCDEG, the findings of this study are anticipated to contribute significantly to developing an EA that effectively enhances NCDEG's service delivery and outcomes.

3. Related Work

Numerous research studies have delved into EA's various applications and roles in healthcare organizations. In our work, we have come across a few such studies. One such comprehensive systematic review was conducted by [14], which analyzed 46 studies in 19 countries between 2015 and 2019. These studies explored the practical applications of EA in various healthcare settings, including hospitals, public and private health systems, e-health, health-information systems (HIS), public-health systems, pharmaceutical corporations and health companies. According to the review results, TOGAF, Adaptive Integrated Digital Architecture Framework (AIDAF), Weil and Ross and Zachman are the four most

commonly implemented EA frameworks in the healthcare sector, accounting for 43% of the studies included in the review. TOGAF was the most widely adopted, featured in 22% of the total studies, followed by AIDAF in 11%, Weil and Ross in 6% and Zachman in 4%. The review also highlighted that several studies employed a combination of multiple frameworks.

A Dharmais Cancer Hospital case study in Indonesia identified the need for an effective information technology solution [15]. However, this proved challenging due to the need for an EA to guide the implementation and execution of technology solutions and evaluate their effectiveness. The study's authors found that TOGAF was the ideal framework for creating an EA, focusing on processes and their alignment with business strategy. Using TOGAF's ADM, the authors identified 36 gaps between the baseline and target architecture needing improvement.

A recent study discussed the Queensland State in Australia model to create a digital health vision enabling all stakeholders to access health and medical information in a consumer-centric system. However, to achieve this vision, adopting an EA framework was necessary. The study noted that in addition to realizing the vision, the framework would need to address several challenges, including inefficient data sharing, difficulty in integrating diverse systems and databases across Queensland, insecure data access, lack of adequate IT governance and a need to increase digital literacy among medical professionals and staff to adapt to new technologies. After conducting a literature review of common EA frameworks, such as TOGAF, Zachman and FEAF, the study concluded that TOGAF was the most suitable framework for the case of Queensland and explained how it would address the identified challenges.

A conceptual EA framework, the Hospital EA Framework (HEAF), was created and specifically designed for hospitals in Iran [16]. This framework is based on the well-established TOGAF framework, but was adapted to meet the unique needs of Iranian hospitals. Through a rigorous methodology for selecting criteria, the authors could justify their reliance on TOGAF and determine the architectural elements involved in each phase. Their study found that this new EA framework could be implemented in hospitals throughout Iran.

Another recent case study was conducted at the Setiabudi District Public Health Center in South Jakarta, Indonesia. The study employed the TOGAF-ADM method, analyzing business, application and technology architecture. Primary and secondary data was gathered through observational studies and interviews at the Public Health Center. Gap analysis is used to compare the target architecture with the current-state architecture. The primary outcome of the research is the presentation of an EA design for the Setiabudi District Public Health Center, aiming to enhance the effectiveness and efficiency of its services [17].

This study contributes to the field of healthcare enterprise architecture by examining the digital transformation process within NCDEG in Jordan. The primary contribution lies in applying The Open Group Architecture Framework (TOGAF) as a methodology to facilitate this transformation while ensuring alignment with the organization's business goals. This approach offers a systematic and comprehensive framework for analyzing the current architecture of NCDEG, identifying areas for improvement and developing a target architecture to address the organization's needs effectively. The simulation provides a strong indication of how this model can assist the center in achieving its goals with better performance measurements. The novelty of this paper lies in its focus on leveraging TOGAF for healthcare enterprise architecture, particularly in the context of a specialized medical center like NCDEG. By employing TOGAF's Architecture Development Method (ADM), our study provides a structured and rigorous methodology tailored to the unique requirements of healthcare organizations, thus filling a gap in the existing literature. This research contributes to advancing the understanding of digital transformation in healthcare settings and provides valuable insights for practitioners and researchers seeking to implement similar initiatives.

The comprehensive nature of TOGAF, along with its flexibility and the ability to align with business goals, are the main reasons for selecting it as an EA framework that could address the complexity of the technological landscape and operational concerns at NCDEG.

4. Research Methodology

This qualitative exploratory study utilized specific research steps to construct the NCDEG's EA. The

following steps were implemented:

Phase 1: Obtaining primary data from NCDEG through observations and semi-structured interviews with officials from NCDEG's IT and quality departments. The first phase aimed to comprehend NCDEG's IT structure, direct medical services and role in providing them. Additionally, necessary data was obtained to establish a baseline and target for the EA.

Phase 2: The NCDEG's EA was built by implementing TOGAF ADM (refer to Figure 1) on its direct medical services. The second phase was accomplished through a sequence of steps.

S1: In the preliminary phase and phase A of the TOGAF Standard ADM, we define NCDEG's business goals based on its mission and vision. We also determine the scope of the EA and establish the EA's principles and governance requirements.

S2: Developing the baseline EA is crucial to understanding NCDEG's current architecture. It involves creating the baseline core architectures of TOGAF Standard ADM: Phase B-Baseline (business architecture), Phase C-Baseline (information-system architecture) and Phase D-Baseline (technology architecture).

S3: To identify areas for improvement and build the target EA, we conduct a gap analysis of the baseline architectures following NCDEG's primary business drivers and goals.

S4: Developing the target EA involves addressing the shortcomings identified in the baseline EA analysis to enhance NCDEG's performance and achieve its business goals. We develop the target core architectures of TOGAF Standard ADM: Phase B-Target (business architecture), Phase C-Target (information-system architecture) and Phase D-Target (technology architecture).

Phase 3: Validating the results by utilizing the Rockwell Arena Simulation Software. This study aims to create an EA for NCDEG's direct medical services, which includes the elements, units, stakeholders, processes and technologies specific to this scope. We will follow the TOGAF ADM phases pertinent to EA development, but not those related to its implementation and governance.

4.1 Research Limitations

This study applies the TOGAF Standard 9.2 ADM phases to the NCDEG case, specifically focusing on the development phases ranging from preliminary to phase D for current/baseline and target architectures. It is worth noting that we have excluded the implementation and governance phases at this study stage.

Moreover, TOGAF recommends several artifacts that architects may develop in each phase of the TOGAF ADM. These artifacts are a collection of catalogs, matrices and diagrams that can address different stakeholders' concerns and requirements within the organization. In this research, the authors decided to use a limited number of diagrams that fit this research.

5. RESULTS AND ANALYSIS

This section presents the NCDEG's baseline EAs, the gap analysis to identify areas for improvement and the target EA that resulted from following the research methodology's phases.

5.1 The Preliminary Phase

The preliminary phase encompasses four main steps: (1) understanding the organizational context, (2) identifying the EA's key business drivers and objectives, (3) identifying and defining the fundamental units, organizations and stakeholders impacted by EA and (4) defining the EA principles. The following sub-phases will cover the preliminary phase of TOGAF ADM.

Organizational Context, Business Drivers and EA Objectives: By interviewing NCDEG officials, we could determine the following business objectives:

1. Providing a comprehensive, safe, high-quality treatment facility as a one-stop destination for patients with diabetes, endocrinology and genetic disorders.
2. Ensuring that all provided medical services and treatments adhere to national and international quality standards.

3. Addressing the prevalence of diabetes, endocrinology and genetic disorders in Jordan, seeking to limit their impact on the population.
4. Integrating medical services with scientific research, training and education and establishing a cohesive synergy between healthcare provision and advancing knowledge in the field.

In addition to the abovementioned objectives, essential business drivers pose challenges, yet offer potential opportunities. These business drivers are the core reasons for dedicating time, resources and effort toward developing and implementing an EA for NCDEG. Figure 2 depicts the EA's objectives derived from NCDEG's business drivers and objectives:

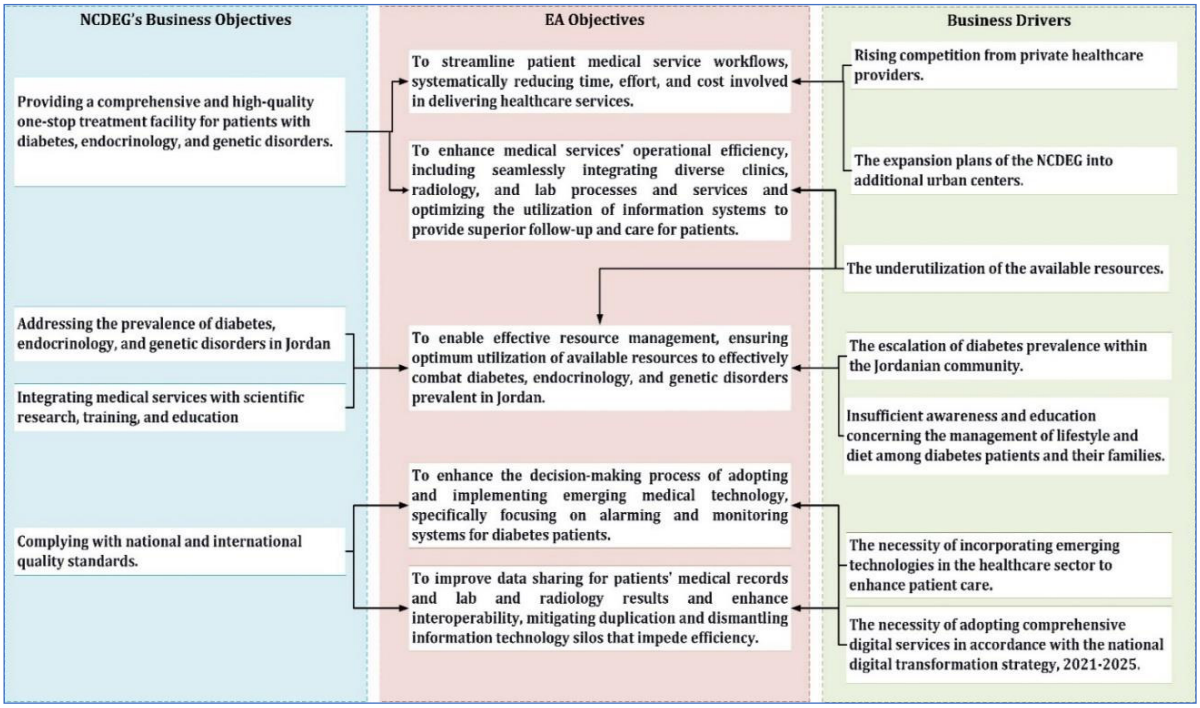


Figure 2. NCDEG's EA objectives.

Identifying and Defining the Fundamental Units, Organizations and Stakeholders Impacted by EA: To maximize the benefits of implementing EA in the NCDEG's medical services, the study has identified the core units that will be significantly impacted. These include the Diabetes Clinics and their labs (including regular and consultation clinics), as well as other specialized clinics, such as Ophthalmology, Cardiology, Pulmonology, Gynaecology, Nephrology, Urology, Dermatology, Neurology and Diabetic Foot Clinic. The Central Lab, imaging department, Pharmacy and Appointment department are also expected to derive maximum value from implementing EA.

Defining the EA Principles: During this phase, the governance mechanism and architecture principles are established to oversee the ADM cycle and the creation, upkeep and utilization of EA and IT resources. These principles guide decision-makers, as the TOGAF Standard outlines. The architecture principles themselves can be categorized into various groups, including business principles (e.g. the importance of principles in all NCDEG divisions), data principles (e.g., treating data as a strategic asset and implementing data governance) and application and technology principles (e.g. prioritizing ease of use and technology independence).

Architecture Vision (Phase A): In this TOGAF ADM phase, we identify the business scenario and stakeholders involved. Specifically, we focus on the NCDEG, the only center for diabetes in Jordan and the challenges that the medical staff and patients face. The busy schedules of physicians and consultants make it challenging to schedule appointments flexibly or quickly, while patients must undergo a lengthy, manual process to access medical services and prescriptions. Stakeholders include NCDEG's administrative and medical staff, management and patients. Figure 3 outlines NCDEG's primary and support activities, which aim to increase efficiency and resource utilization while improving patient service quality. The EA's vision is to achieve these goals.

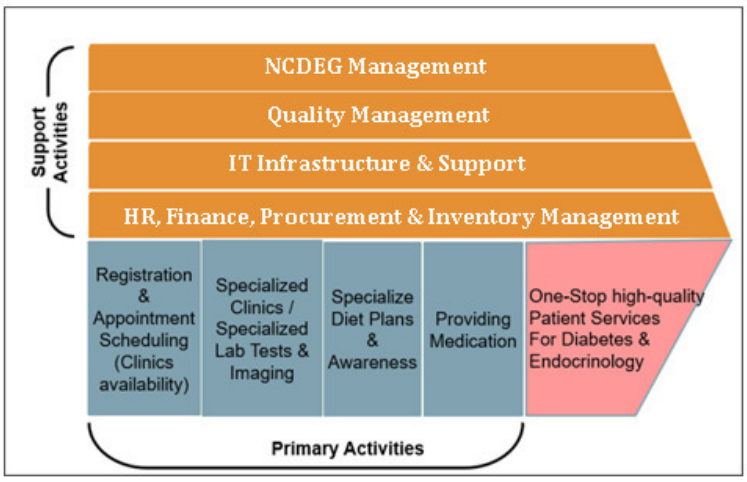


Figure 3. Value chain of NCDEG.

5.2 Developing the Baseline Enterprise Architecture

This step is crucial in comprehending the present state of NCDEG architecture. It involves the development of the current or baseline core architectures of NCDEG by the TOGAF Standard, comprising business, information system and technology architectures.

Baseline-Business Architecture: A detailed baseline process diagram (Figure 4) has been created to gain a comprehensive understanding of the processes in line with this study's scope. It represents the primary activities necessary to achieve NCDEG's goals.

Baseline-Information System Architecture: The information-system architecture entails understanding the data and application architectures. Based on the information gathered from NCDEG, the current NCDEG Health Information System (HIS) comprises the components depicted in Figure 5. In this phase, the baseline data and application architectures were not presented thoroughly due to data access restrictions; however, the gathered data to create the HIS functional decomposition diagram was provided and validated by the NCDEG.

Baseline-Technology Architecture: The NCEDG's IT management must provide adequate details to construct the current technology architecture. However, they have mentioned that users connect to servers *via* wired or wireless connections. Oracle DBMS and the backup appliance are utilized to manage data on database servers. Additionally, NCDEG must employ cloud services to safeguard patient privacy and information security.

5.3 Baseline EA Gap Analysis

After analyzing the baseline EA, we have identified several gaps and areas for improvement. One significant issue is the improper follow-up process for patients referred for laboratory tests or imaging procedures. Currently, patients only receive updates on their results from the referring physician when their next appointment is scheduled, which can take up to three months in diabetes clinics and 4-5 months in specialized clinics, like cardiology and neurology clinics, due to limited available appointments. This inadequate follow-up process can lead to severe health complications for affected patients, making it crucial to address this issue.

The Lengthy Prescription Renewal Process: Patients requiring only a renewal of their medical prescriptions must undergo an extensive procedure to obtain their medications, much like any other patient. This process worsens the problem of appointment unavailability in heavily crowded clinics and it also costs patients considerable time, effort and financial resources. It is essential to streamline this process and make it more efficient to improve patient outcomes and reduce the burden on the healthcare system.

Inefficient Monitoring of Glucose Levels: For diabetes patients, especially those who rely on insulin therapy, monitoring blood sugar levels is crucial to assess and adjust their therapy and diet plans continuously. During a visit to the NCDEG, it was noted that patients are given booklets to record their daily glucose readings, known as self-monitoring blood glucose (SMBG). However, some patients

overlook this step and resort to the accumulative sugar blood test HbA1c, which alone cannot offer a comprehensive analysis of the patient's condition to establish an effective diet plan or support the doctor's assessment of the current treatment plan, particularly for insulin therapy, as SMBG can.

Nevertheless, providing doctors and dieticians with a booklet containing three months of daily readings is impractical. It would be challenging for doctors to allocate sufficient time to review and analyze all these readings effectively.

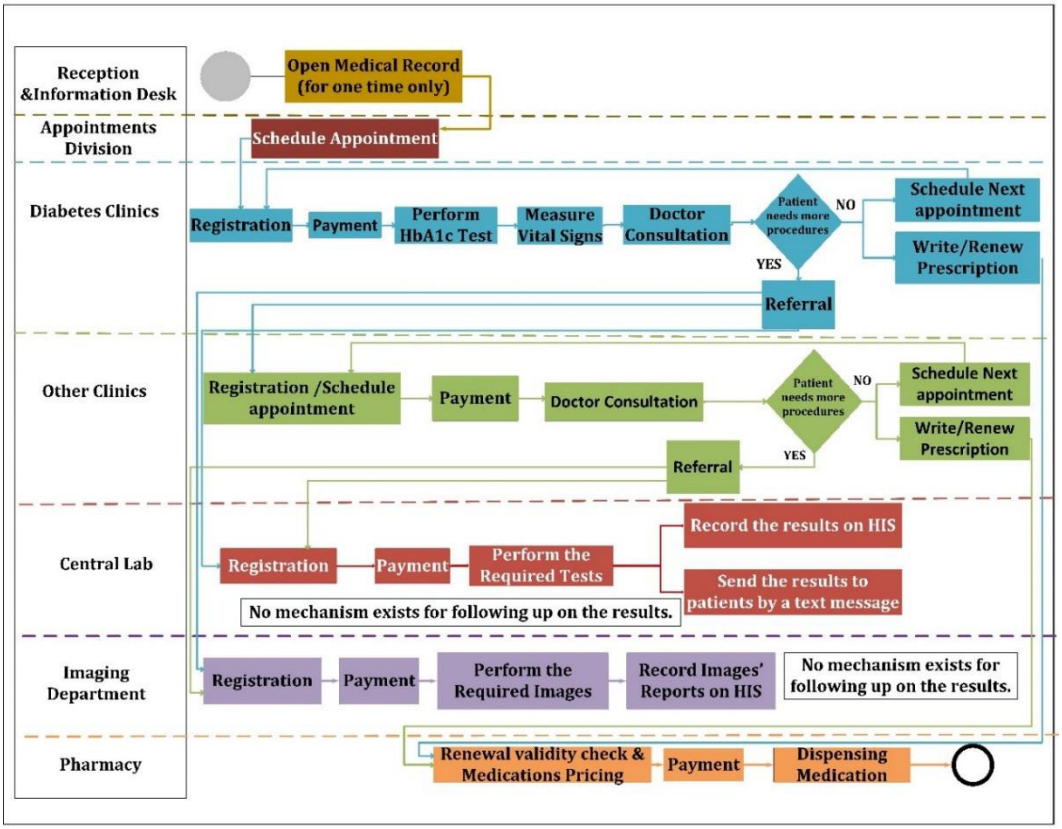


Figure 4. Baseline business architecture.

Appointment Availability: Securing a timely consultation with NCDEG physicians is problematic. In specific clinics, the next available appointment could be months away. This task poses a significant challenge to patients' treatment plans and could worsen their health. Furthermore, the unavailability of appointments may put an added financial strain on insured patients with NCDEG. In such situations, patients may be compelled to seek treatment from private healthcare providers, which could result in a loss of revenue for NCDEG.

Incomplete Process Automation and Modules Integration: The existing process lacks full automation, requiring patients to complete specific paperwork at various stations during their visit, such as the clinic registration desk and pharmacy. This could lead to increased time that patients spend on a particular service.

Resources Underutilization (Utilization of Education and Dietician Services): Research has highlighted the significance of managing one's diet and weight and engaging in physical activities to mitigate diabetes complications and enhance the patient's quality of life. The NCDEG offers a dietetics clinic and a diabetes-education clinic to provide patients with the necessary resources for self-management. However, a visit to the NCDEG reveals that both clinics suffer from low attendance rates, which could jeopardize patients' well-being and health. Furthermore, this underutilization of services undermines the NCDEG's mission to disseminate diabetes education and raise awareness of self-management requirements [9].

Table 1 summarizes the gap analysis, highlighting the current shortcomings in the NCDEG process, their impact on critical patient aspects, like health, cost and time, and their influence on NCDEG's

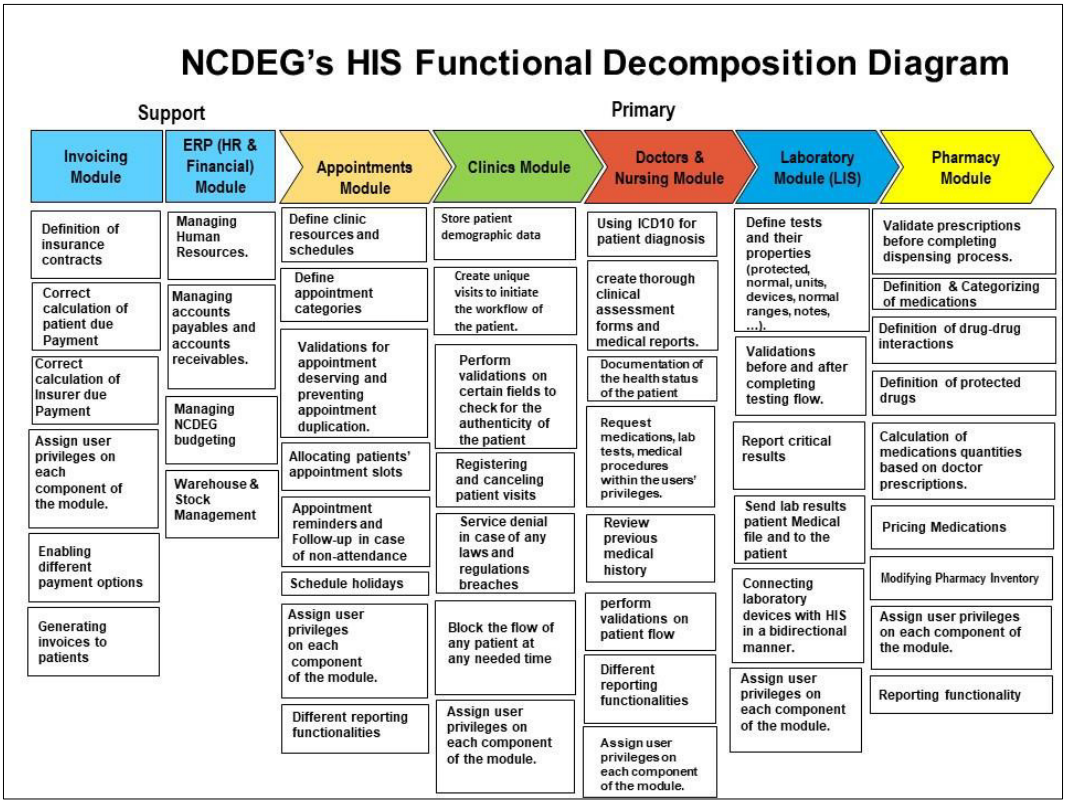


Figure 5. HIS functional decomposition diagram – Baseline information-system architecture.

primary activities. The red cells in Table 1 indicate the impacted aspects out of seven ones. After analyzing the table, it was found that improper follow-up and unmonitored glucose levels are the most critical issues affecting patient well-being and NCDEG's primary activities. These issues represent 71% (5/7) and 57% (4/7) of criticality, respectively, assuming that all aspects have the same weight/importance. Therefore, it is crucial to prioritize resolving these shortfalls.

Furthermore, the analysis highlights the significant effect of current process shortfalls on patient health and NCDEG's overall operations. Process automation, HIS module integration and streamlined prescription renewals are recommended to address prolonged service times and appointment unavailability. These measures would enhance patient care, improve efficiency and resource utilization and reduce waiting times.

In the following part, we will discuss NCDEG's EA, which aims to fill the gaps in the baseline architecture and address these analysis outcomes.

5.4 Developing the Target Enterprise Architecture

The targeted EA encompasses business, information-system and technology architecture changes.

The Target Business Architecture (Phase B): The target business process outlined in Figure 6 addresses issues related to medical prescription renewals, appointment availability and underutilization of resources. The process described in Figure 6 focuses on several key areas, including:

1. Advanced technologies, like telemedicine and virtual clinics, save patients' time when renewing prescriptions. Through online sessions with a doctor, patients can have their case assessed and prescription renewal approved through the HIS. Telemedicine and mobile apps can also offer dietician services and awareness programs to improve resource utilization. Integrating AI technologies with wearable devices or mobile apps can provide personalized tools for diabetes prevention and management, including the detection of diabetic complications, such as hypoglycemia, diabetic retinopathy and cardiovascular risks, and the enablement of the artificial pancreas.
2. The process also includes follow-up sessions with patients after receiving test or imaging results to ensure that they are updated on their results and any necessary actions or procedures as soon as possible, rather than waiting until their next visit, which could be up to three months away.

Table 1. Gaps impact on patients and NCDEG's primary activities (impacted cells are in red color).

	Impacts on Patient			Impacts on NCDEG's Primary Activities			
Shortfalls	Health Impacts	Financial Impacts	Service Time	Registration & Appointment Scheduling	Specialized Clinics / Lab Tests & Imaging	Specialized Diet Plans & Awareness	Providing Medication
Improper Following up 71% (*5/7* impacted aspects)							
Lengthy Prescriptions Renewal Process 42% (*3/7*)							
Inefficient GCM 57% (*4/7*)							
Appointments Unavailability 42% (*3/7*)							
Incomplete Process Automation & Modules Integration 42% (*3/7*)							
Resources Underutilization 42% (3/7)							

Figure 6 depicts the newer technology in daily-basis activity in NCDEG, where the new architecture comprises several interconnected elements to enhance the organization's operational efficiency and patient-care quality. These elements include adopting digital health solutions, including telemedicine and virtual clinics, integrating AI technologies for personalized patient care and optimizing existing processes and workflows to streamline operations.

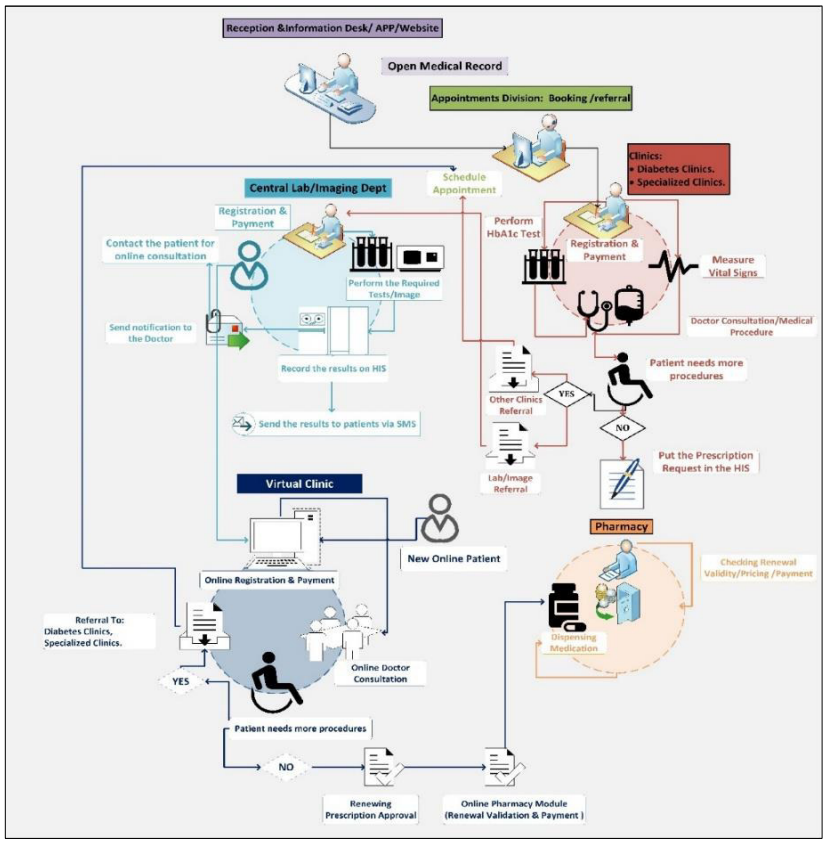


Figure 6. The target business architecture.

The Target Information-system Architecture - Data (Phase C): Regarding the information-system architecture, proposed changes to the business process outlined in Figure 6 necessitate adjustments to the current system. The system consists of data and application architectures that depict the various entities within it and their relationships. Specifically, the data architecture of the system must allow for the electronic capture of patient data and glucose-level readings and integration with the patient's record in HIS. Currently, NCDEG's patients manually record their daily blood-sugar levels, which requires six

inputs daily in a notebook provided by NCDEG. The new system will also enable patients to schedule appointments, update their data and profile and allow NCDEG staff to interact with the system's data per their roles, as illustrated in Figure 7.

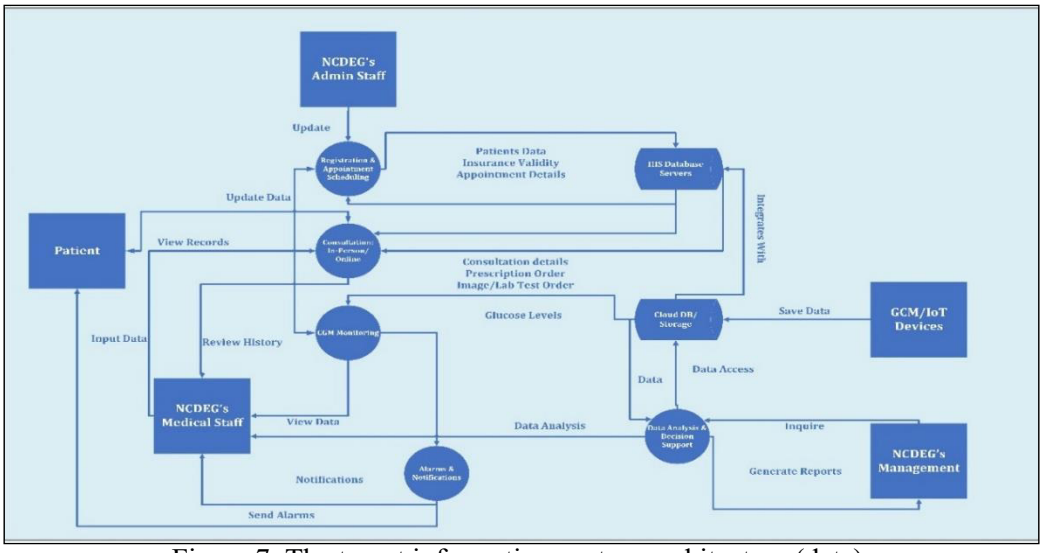


Figure 7. The target information-system architecture (data).

The Target Information-system Architecture - Application (Phase C): The architecture comprises several components that seamlessly provide a comprehensive healthcare experience. These components include a telemedicine module that enables online consultations, with added features, such as online payments and insurance-validity checks, if necessary. Additionally, the system allows for collecting and managing patients' data, including important factors, like sugar levels and blood pressure. The platform includes AI-powered patient-data analysis and visualization for better data interpretation and decision-making. Finally, the system provides follow-up reminders and alarms for patients whose sensed data or test results reveal risky conditions. For a more detailed understanding of the target architecture, please refer to Figure 8. Figure 8 depicts the suggested three platforms (interfaces): web, mobile-phone interfaces and data warehouse for data storage and archiving. The main modules that incorporate AI-powered services are included in the figure.

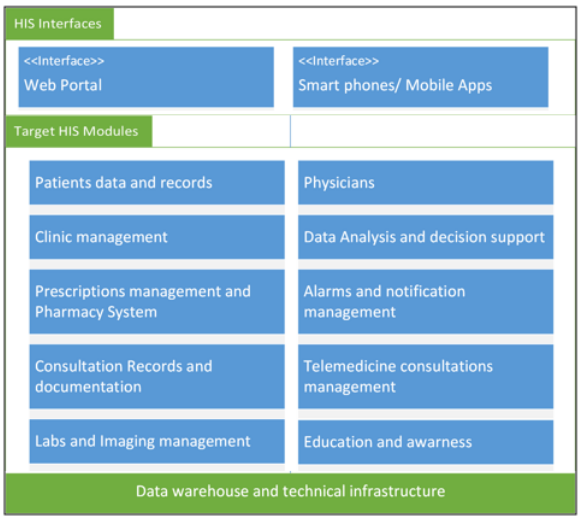


Figure 8. The target information-system architecture (application).

The Target Technology Architecture (Phase D): Figure 9 describes the target information system with the required technological advancements that can be implemented to improve services' quality and performance. The figure shows the sequences of actions that the user and NCDEG staff privileges can take. The HIS modules manage the interaction between the patient and staff views to verify the procedures before the integration process and data analytics. Data management and warehouse store the data for future reference and reporting.

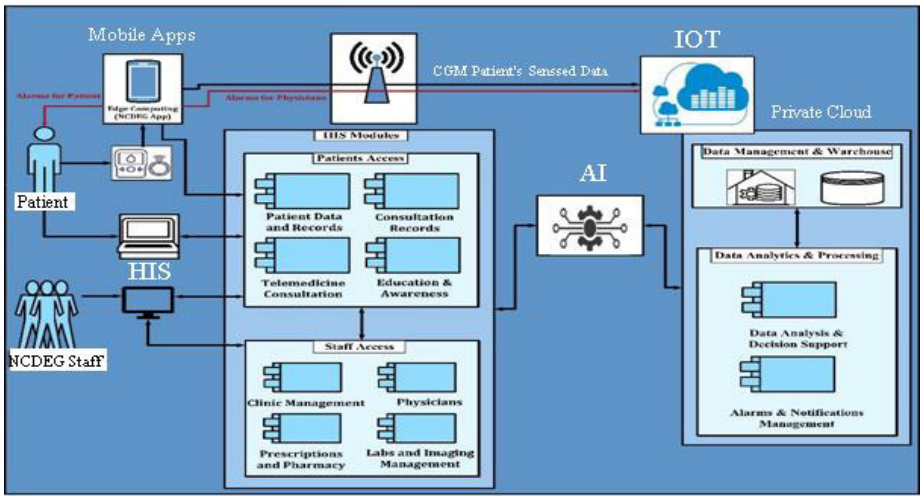


Figure 9. The target technology architecture.

Figure 10 outlines additional improvements and enabling technologies that the NCDEG can leverage to enhance its services. The enablers represent the cutting-edge technologies that can improve the center's services in general to achieve its goals.

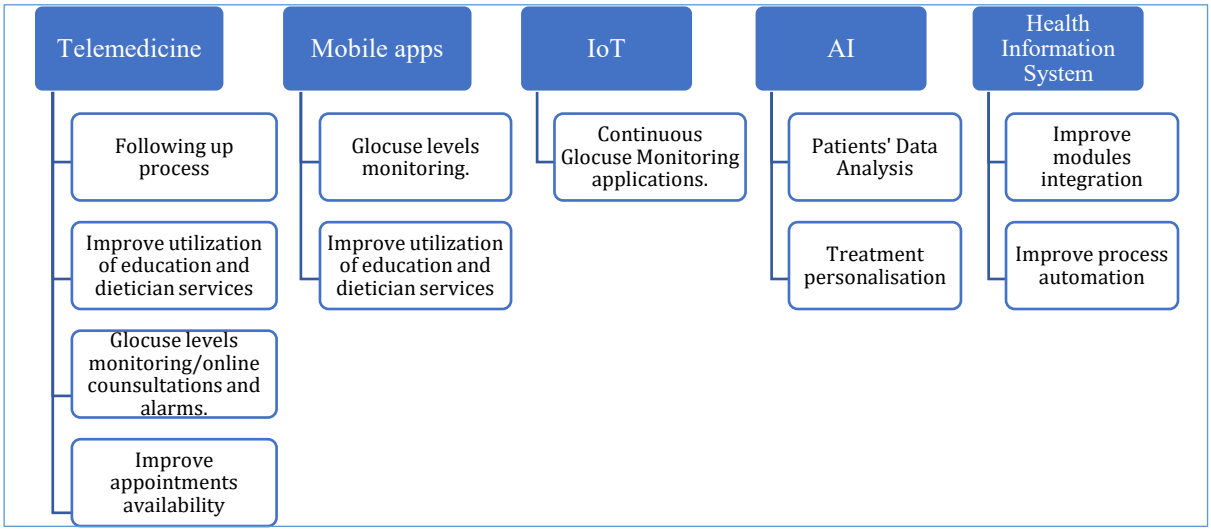


Figure 10. Proposed technologies and improvements.

5.5 Validation

The proposed changes have been validated by simulating the current and target processes to enable comparison and analysis of each process's key performance indicators (KPIs). Figure 11 illustrates the simulation of the proposed changes (target business process) using Arena Simulation Software.

The Data Gathering Phase: Data gathering consisted of visiting NCDEG for four days to collect observational data (5 cases per day) from diabetes clinics. The data collection focused on gathering data on patients' visits to diabetes clinics, the radiology department, the labs and the pharmacy.

The Simulation Phase: Rockwell Automation's Arena Simulation Software (version 16) was used to conduct five replications based on observational data. Each replication was designed to mimic six days with eight operational hours per day and the Base Time Units were represented in minutes. The simulation of the current process resulted in the following main KPIs shown in Table 2 and Table 3. In contrast, the simulation of the target process, which includes physical and virtual clinics running simultaneously, resulted in the main KPIs shown in Tables 4 and 5.

Analysis of the Simulation Results

1. The analysis of simulation results indicates that establishing a virtual clinic, even with minimal resources, can significantly improve NCDEG's ability to handle an enormous patient load. The

simulations demonstrate a marked increase in the total patient count, representing a 32% rise, from 2322 to 3067. The virtual clinic enables patients who require prescription renewals to consult with doctors online without undergoing the typical clinical process. Such a digital health solution can relieve stress for NCDEG's constrained resources, enhancing the overall availability of medical appointments.

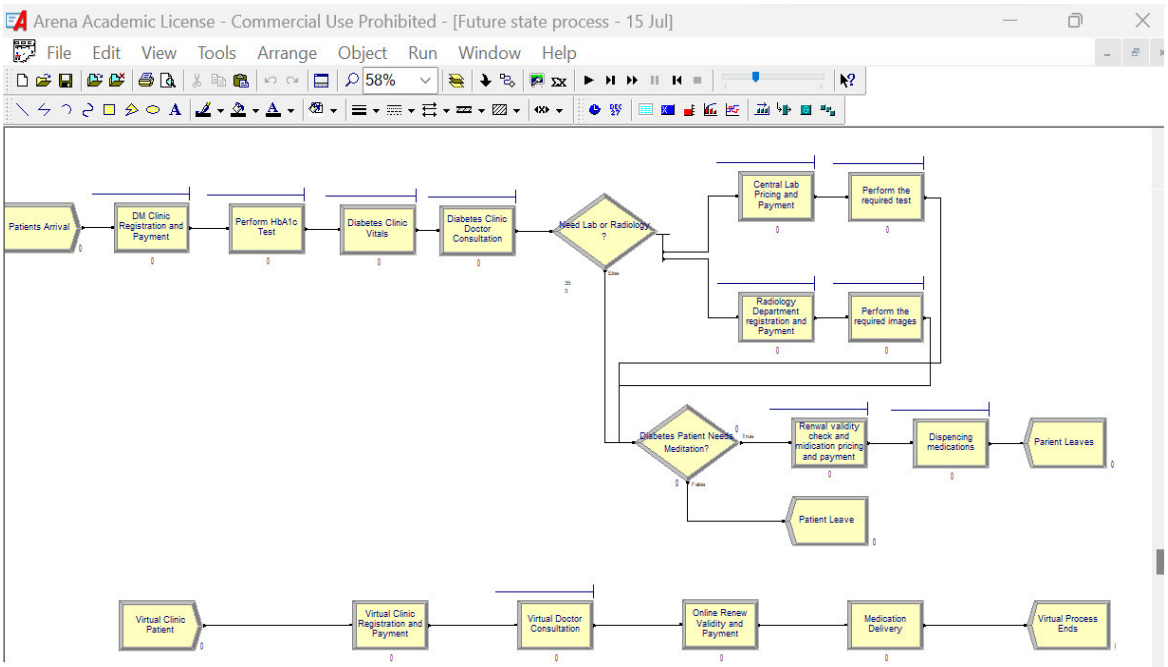


Figure 11. Illustration of the simulation of the proposed changes using the Arena Simulation Software.

Table 2. Entity (patient) KPIs.

Entity	Average Number
Patients' Total Number in	2380
Patients' Total Number out	2322

Table 3. Time KPIs.

Waiting Time (Minutes)	Average	Maximum
Total Process Length	46.1643	83.8235
Total Waiting Time	12.3707	64.5179
Total Time to Service (Process Length+ Waiting Time)	58.5350	142.33
DM Clinics Queuing Time (Registration Queue + Payment Queue + Nursing Vitals Queue+ HbA1c Test Queue)	0.011499	3.5663
DM Clinics Doctors Consultation Waiting Time	10.0711	62.5091
The Central Lab Queuing Time (Central Lab Payment Queue + Central Lab Pricing Queue+ Performing the required Test Queue)	0.853825	7.6104
Pharmacy Queuing Time (Pricing Queue + Payment Queue+ Dispensing Medications Queue)	2.2335	29.2735
Radiology Dept Queuing Time (Registration Queue + Payment Queue + Performing the Required Image Queue)	2.742817	42.0904

Table 4. Entity (patient) KPIs.

Entity	Average Number
Inpatient number (in-person clinics)	2389
Virtual inpatient number	723
Total inpatient number	3112
Outpatient number (in-person clinics)	2350
Virtual outpatient number	717
Total outpatient number	**3067**

Table 5. Time KPIs.

Waiting Time (Minutes)	Average	Maximum
Traditional In-Person DM Clinic KPIs		
Total Process Length	43.7283	75.5309
Total Waiting Time	11.5185	72.9616
Total Time to Service (Process Length+ Waiting Time)	55.2468	124.71
DM Clinics Queuing Time (DM Clinic Registration and Payment + Nursing Vitals Queue+ HbA1c Test Queue)	0.000971	1.872
DM Clinics Doctors Consultation Waiting Time	9.6415	61.1566
The Central Lab Queuing Time (Central Lab Pricing and Payment Queue + Performing the Required Test Queue)	0.018707	2.853
Pharmacy Queuing Time (Renewal Validity Check and Medication Pricing and Payment Queue+ Dispensing medications Queue)	1.9254	19.8665
Radiology Dept. Queuing Time (Registration and Payment Queue + Performing the Required Image Queue)	3.455178	37.2402
Virtual DM Clinic KPIs		
Total Process length	12.0500	12.0500
Total Waiting Time	0.00	0.00
Total Time to Service (Process Length+ Waiting Time)	12.0500	12.0500
Pharmacy Queuing Time	It is not applicable as the medicine should be delivered to the patient.	

2. The proposed changes can reduce the maximum service time, including process duration and waiting time, by 12%. The current process has too many unnecessary steps, making it long, complicated and wasteful. Merging the clinic registration and payment into one step can save time and simplify the process's complexity. The same idea can work for the central lab, radiology and pharmacy departments by joining the registration/pricing and payment into one more straightforward step.
3. The virtual clinic offers a streamlined process where registration and payment can be completed online. Additionally, medications can be dispensed to these patients *via* a paid delivery service. The virtual clinic also provides a practical and accessible platform for follow-up interactions between patients and doctors. For instance, if lab tests or radiology reports reveal concerning results, the doctor and the patient can swiftly arrange a virtual consultation session, avoiding the necessity for an in-person visit, subject to appointment availability.
4. Digitalization of services can significantly improve existing processes. For instance, a separate step for medicine pricing becomes redundant within the proposed virtual clinic. This action is possible, because the system automatically registers approved medications, enabling their readiness for online payment. This strategy can be effectively applied to in-person clinic visits as well. The electronic health assistant (EA) can improve interoperability and integration among NCDEG's Health Information System (HIS) modules. When a doctor prescribes medications through the system, the integrated clinic and pharmacy modules can autonomously calculate the cost and prepare the prescription for payment, thus eliminating the need for a distinct queue for medication pricing.
5. Moreover, digital transformation has the potential to assist patients visiting the NCDEG in person in bypassing the queue for payments. This action can be achieved by facilitating online payments *via* the NCDEG's app. or website or installing smart kiosks that streamline clinic registration and payment for various services, such as clinics, labs, radiology and pharmacies.
6. The virtual-clinic model can enhance dietary and diabetes awareness services. Although the NCDEG has yet to provide data on the number of patients using these services, several visits at varying intervals suggest a low usage rate. However, transitioning these services to a virtual format could increase their reach and accessibility, benefiting NCDEG patients and the wider Jordanian community. Given the high prevalence of diabetes in Jordan, such a transition could significantly enhance diabetes awareness, understanding and management, aligning well with the business goals of NCDEG.

6. DISCUSSION

The integration of the new architecture at NCDEG is justified by its potential to address critical

challenges and capitalize in emerging opportunities within the healthcare landscape. By utilizing digital health solutions and AI technologies, NCDEG can enhance its capacity to deliver comprehensive, safe, high-quality treatment facilities while adhering to Jordanian and international quality standards. Moreover, optimizing existing processes and workflows enables NCDEG to streamline operations, reduce waiting times and improve overall service delivery, ultimately contributing to better patient outcomes and organizational efficiency.

The implementation of EA at the NCDEG in Jordan marks a significant milestone in the organization's digital-transformation journey. Several vital challenges were encountered throughout this process, reflecting the inherent complexities of integrating technological advancements within a healthcare framework. The various elements of the new architecture interact synergistically to support NCDEG's overarching business goals and objectives. For example, integrating telemedicine and virtual clinics enables patients to access medical services remotely, reducing the need for in-person visits and enhancing appointment availability. Additionally, AI technologies facilitate data-driven decision-making and personalized patient interventions, improving clinical outcomes and patient satisfaction.

We should keep in mind all difficulties that may arise during the implementation, including user resistance, the complexity of automating the process and facing procedures that may stop the adoption of the suggested framework. Looking ahead, several opportunities for future enhancements and refinements within NCDEG's EA framework become apparent. Firstly, ongoing monitoring and evaluation of the implemented solutions will be essential to assess their effectiveness and identify areas for further optimization. Additionally, continued collaboration with healthcare practitioners and IT experts will ensure that the EA framework remains adaptive and responsive to evolving patient needs and technological advancements. Moreover, efforts to enhance data security and privacy measures will be paramount, particularly in light of the increasing reliance on digital health solutions and the sensitive nature of patient information.

The successful integration of the new architecture at NCDEG underlines the transformative potential of technology in improving healthcare delivery and patient outcomes. By providing clear explanations of the elements, interactions and justification for the new architecture, NCDEG has taken significant strides toward providing better medical services while reducing the time and cost for the institution and creating a better patient experience.

7. CONCLUSION

The National Cenetr for Diabetes, Endocrinology and Genetics (NCDEG) in Jordan underwent a study utilizing the TOGAF Standard framework. Primary data was obtained through interviews with NCDEG officials despite restrictions on data provision. The application of TOGAF Standard ADM revealed areas for improvement in NCDEG's business, information systems and technology architectures and proposed changes to better align their digital-transformation initiatives with their vision, mission and business objectives. To ensure better results, improved interoperability and integration among processes, data and technologies, the study recommends developing EA while ensuring top-management engagement and assessing available resources for successful development. The TOGAF Standard 9.2 ADM phases applied to the NCDEG case include the development phases, from preliminary to phase D for current/baseline and target architectures. Implementation and governance phases were excluded. A baseline and target processes simulation was conducted to validate proposed changes and highlight their impact. It is recommended that the healthcare sector in Jordan can utilize EA as a strategic tool to plan, manage and govern its digital transformation and IT landscape using suitable EA tools or frameworks.

The study relied on available data from NCDEG, which may have limitations in terms of completeness and accuracy. Another limitation is the narrow scope of the evaluating simulation compared to open-scope real-life systems. The future direction is to explore the applicability of the proposed approach to other healthcare organizations and settings. Security, privacy and other confidential measures must be examined in future works.

REFERENCES

[1] D. R. Banger, Enterprise Systems Architecture, Apress, 2022.
[2] E. Niemi and S. Pekkola, "The Benefits of Enterprise Architecture in Organizational Transformation," Business & Information Systems Engineering, vol. 62, no. 6, pp. 585–597, Dec. 2020.

[3] S. Razdan and S. Sharma, "Internet of Medical Things (IoMT): Overview, Emerging Technologies and Case Studies," IETE Technical Review, vol. 39, no. 4, pp. 775–788, 2022.
[4] A. Okhrimenko, Comparing Enterprise Architecture Frameworks – A Case Study at the Estonian Rescue Board, M.Sc. Thesis, Institute of Computer Science, University of Tartu, pp. 1–71, 2017.
[5] TOGAF, "TOGAF Introduction," [Online], Available: https://pubs.opengroup.org/architecture/togaf9-doc/arch/index.html, 2018.
[6] TOGAF, "TOGAF-Preliminary Phase," TOGAF Standard 9.2, [Online], Available: https://pubs.opengroup.org/architecture/togaf9-doc/arch/index.html, 2018.
[7] B. P. I. Sapra A and S. Vaqar, "Diabetes Mellitus," StatPearls [Internet], no. December 2019, [Online], Available: https://www.ncbi.nlm.nih.gov/books/NBK513253/?report=classic, 2019.
[8] I. M. El-Kebbi, "Epidemiology of Type 2 Diabetes in the Middle East and North Africa: Challenges and Call for Action," World Journal of Diabetes, vol. 12, no. 9, pp. 1363–1586, 2021.
[9] WHO, "Diabetes," WHO-Health Topics, [Online], Available: https://www.who.int/health-topics/diabetes#tab=tab_1, 2021.
[10] TOGAF, "TOGAF ADM," TOGAF Documentation Chapter 5, [Online], Available: http://pubs.opengroup.org/architecture/togaf9-doc/arch/chap05.html, 2018.
[11] IDF, "IDF Jordan Diabetes Report 2000-2045," International Diabetes Federation Diabetes Atlas, 2021. [Online], Available: https://diabetesatlas.org/data/en/country/102/jo.html.
[12] K. Ajlouni et al., "Time Trends in Diabetes Mellitus in Jordan between 1994 and 2017," Diabetic Medicine, vol. 36, no. 9, pp. 1176–1182, DOI: 10.1111/dme.13894, 2019.
[13] S. F. Awad et al., "Characterizing the Type 2 Diabetes Mellitus Epidemic in Jordan up to 2050," Scientific Reports, vol. 10, no. 1, pp. 1–10, DOI: 10.1038/s41598-020-77970-7, 2020.
[14] S. H. Silvano et al., "Frameworks, Methodologies and Specification Tools for the Enterprise Architecture Application in Healthcare Systems: A Systematic Literature Review," Proc. of 2020 IEEE Int. Conf. on E-health Networking, Application & Services (HEALTHCOM), DOI: 10.1109/HEALTHCOM49281.2021.9398916, Shenzhen, China, 2021.
[15] A. S. Girsang and A. Abimanyu, "Development of an Enterprise Architecture for Healthcare Using TOGAF ADM," Emerging Science Journal, vol. 5, no. 3, pp. 305–321, 2021.
[16] A. Haghighathoseini, H. Bobarshad, F. Saghafi, M. S. Rezaei and N. Bagherzadeh, "Hospital Enterprise Architecture Framework (Study of Iranian University Hospital Organization)," Int. J. of Medical Informatics, vol. 114, no. March, pp. 88–100, DOI: 10.1016/j.ijmedinf.2018.03.009, 2018.
[17] D. Prayitno, B. Indiarto and M. B. Legowo, "Enterprise Architecture Planning for Public Health Centers," Aijbm.Com, vol. 6, no. 8, pp. 132–143, [Online], Available: https://www.aijbm.com/wp-content/uploads/2023/08/P68132143.pdf, 2023.

ملخص البحث:

يُمكن لتطبيق التّكنولوجيا أن يفيد المنظّمات بشكلٍ كبير، ويجب أن تنسجم مع احتياجات المنظّمات وأهدافها. ويُمكن لتطبيق "بنية مشروع" مساعدة مؤسسات الرّعاية الصّحية في التّغلُّب على التّحدّيات في أثناء عملية الانتقال الرّقمي.

هذه الورقة تبيّن كيف يُمكن لتطبيق "بنية مشروع" أن يسهل عملية الانتقال الرّقمي مع ضمان أن يتماشى ذلك مع احتياجات العمل. وباستخدام طريقة تطوير بنية المشروع، تعمل هذه الدّراسة على تحليل البنية الرّاهنة للمركز الوطني للسّكري والغدد الصّمّاء والوراثة في الأردن بهدف تطوير بنية جديدة للمركز من شأنها أن تساعد المركز على تحقيق أهدافه، وذلك من خلال التّناغم بين تطبيق التّكنولوجيا وغايات المركز المراد تحقيقها. وتبحث الدّراسة في التّعقيدات المتعلّقة بالتّطورات التّكنولوجية في الوقت الّذي يُحْرص فيه على دمج التّكنولوجيا ببنية المشروع بسلاسة وفاعلية في استغلال الموارد المتاحة. يُضاف الى ذلك التّركيز على الحاجة الى تقييم بنية المشروع وضبطها باستمرار لمواجهة التّطورات الّتي تحدث في بيئة العمل. وقد تمّت محاكاة التّغيرات الضّرورية في بنية المشروع باستخدام البرمجيات المناسبة، وبيّنت النّتائج تحسُّناً ملحوظاً في التّعامل مع المرضى، وفاعلية العمليات، وأوقات الانتظار، واستغلال الموارد عبر تطبيق العيادات الافتراضية والحلول الرّقمية.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 02, June 2024.

TEXT TO VIDEO USING GANS AND DIFFUSION MODELS

Nikita Singhal, Praval Pratap Singh, Nikhil Singh, Mahipal Singh and Harsimrat Singh

(Received: 21-Feb.-2024, Revised: 5-Apr.-2024, Accepted: 24-Apr.-2024)

ABSTRACT

The challenging endeavour of text-to-video creation requires transforming text descriptions into realistic and cohesive videos. This field of study has made substantial progress in recent years, with the development of diffusion models and generative adversarial networks (GANs). This study examines the most modern text-to-video generation models, as well as the various steps involved in text-to-video generation, including temporal coherence, video generation and text encoding. We additionally emphasise the challenges involved with text-to-video generation, as well as recent advances to overcome these issues. The most frequently used datasets and metrics in this field are also analyzed and reviewed.

KEYWORDS

Text-to-video, Coherence, GAN, Diffusion.

1. INTRODUCTION

Video generation has grown dramatically in recent years, gaining popularity due to its various advantages and applications in a variety of sectors, including marketing, branding, content development, artificial video-dataset generation and so on. The objective of this article is to review and compare various text-to-video (T2V) generation approaches. Our goal is to investigate various models across various stages. Very few articles investigated video generation in depth. Furthermore, as new approaches are discovered, it becomes necessary to compare them in order to identify the limitations and constraints of existing techniques, which may then be used by other researchers for future study and enhancements. Table 1 compares the proposed survey study to existing T2V survey studies, including their features and limitations.

T2V is the next step after text-to-image (T2I). Like T2I, T2V began with the use of GAN[3] models and proceeded to the use of different techniques, the most common of which is diffusion, which allows us to use previously existing text to image models. A lot of study has been done in the text-to-image field, since it is used in T2V in diffusion video models.

In the beginning, GANs, which were excellent at generating images at the time, were used to generate images from text. However, stable diffusion has grabbed the lead in producing images of excellent quality in recent years. Different methodologies and tactics for addressing additional concerns, such as temporal and spatial consistency, were considered. Furthermore, the metrics used to evaluate text-to-video models have changed and novel metrics, such as FVD [4], have been established to provide further insight into text-to-video models. Some well-known models, such as Text2VideoZero [5] and Hotshot-XL [6], are also evaluated in terms of how well they perform using an FVD matrix.

The rest of the paper is organized as follows. Section 2 summarizes the various stages and approaches employed in T2V. In Section 3, we discussed the most often used datasets in T2V. In Section 4, we reviewed numerous metrics for evaluating T2V performance. In Section 5, we discussed open challenges and directions for future research and in Section 6, we concluded the work.

2. LITERATURE SURVEY

2.1 Video Generation

Video generation, a dynamic field at the intersection of artificial intelligence and multimedia, encompasses a spectrum of techniques dedicated to converting conditional and unconditional information into captivating visual content. The process involves a thoughtful blend of natural language

N. Singhal, P. P. Singh, N. Singh, M. Singh and H. Singh are with Department of Computer Engineering, Army Institute of Technology, Pune, India. Emails: [ngupta,pravalsingh_20250, nikhilsingh_20229,mahipalsingh_20134,harsimatsingh_20060]@aitpune.edu.in

processing (NLP), machine learning and creative design principles.

Table 1. Comparison of existing studies with proposed studies.

Study & Year	Advantages	Limitations
[1], 2023	• Provides an overview of existing literature of T2I and T2V AI generation • Theoretical comparison of different T2I and T2V models	• Performance evaluation is not conducted • Does not describe the processes involved in T2I or T2V generation
[2], 2023	• Comprehensive coverage-covering domains : video generation, editing and understanding • In-depth examination of the diffusion-model applications in the context of video • Conducted Performance evaluation	• Does not explain internal processes involved in T2V generation
Proposed Review	• Comprehensive coverage of video generation using textual input • Discussed the internal processes involved in T2V generation (including T2I, cross frame attention, motion dynamics and frame interpolation) • Conducted performance evaluation of various models using FVD score • Evaluated FVD score for models that were not included in [2]	

As video generation continues to evolve, researchers explore novel ways to dynamically generate scenes, integrate user feedback and enhance content creation through adaptive systems. This fusion of technology and creativity not only automates the process of video production, but also opens new frontiers for personalized and engaging multimedia experiences. Whether used in education, entertainment or communication, video generation stands as a testament to the ever-expanding capabilities of AI in transforming textual narratives into visually compelling stories.

Zhen Xing et al. [2] categorized video generation into four categories: Text-to-video generation, video-generation using different conditions, unconditional video generation and video-completion. In the proposed survey, we will delve into a comprehensive exploration of text-to-video generation thoroughly examining the various steps involved in the text-to-video generation process.

2.2 Text-to-video Generation

Video generation using GANs and diffusion models represents a cutting-edge approach in the realm of artificial intelligence and computer vision. GANs, pioneered by Ian Goodfellow et al. [3], consist of two neural networks, the generator and the discriminator, engaged in a game-like scenario, where the generator strives to create realistic data (in this case, video frames), while the discriminator aims to differentiate between real and generated data. This adversarial training process leads to the generator producing increasingly realistic video sequences over time. The entire system optimizes a function denoted as in Equation (1):

$$V(D, G) = E_{x \sim p_{\text{data}}(x)}[\log D(x)] + E_{z \sim pz(z)}[\log(1 - D(G(z)))] \quad (1)$$

where D(x) denotes the output of the discriminator for real data (x), G(z)) denotes the output value of the generator for latent vector (z), Ex∼pdata(x)[log D(x)] is the desired output value of the discriminator for actual data (x), Ez∼pz (z)[log(1 − D(G(z)))] denotes the desired output value of the discriminator for generated data (G(z)), x is derived using data distribution $p_{data}(x)$ and z is derived using the latent space distribution pz (z). Video GAN generators employ four primary strategies to effectively generate realistic video sequences. Firstly, they often utilize a hybrid approach that combines Recurrent Neural Network (RNN) architectures using 2D CNN to handle both temporal and spatial information. Secondly, some models opt for 3D convolutional networks instead of 2D ones to directly capture spatiotemporal

features from video data. Additionally, inspired by progressive growing GAN architecture, video GANs implement a coarse-to-fine strategy, refining generated data progressively for enhanced output. Lastly, they may adopt a two-stream architecture, with parallel streams specialized in processing different aspects of video data, aiding in capturing spatial and temporal features effectively. For the discriminator, strategies such as using a two-stream architecture or a 3D convolutional network are employed to distinguish between real and generated video data based on their effectiveness. These strategies collectively contribute to the advancement of video generation techniques. In recent years, researchers have integrated diffusion models into the video generation process to optimize the quality and realism of generated content. Diffusion models, inspired by the concept of Brownian motion, simulate the gradual spreading or diffusion of information or features across a data space. Diffusion models can capture long-range dependencies and temporal coherence, allowing for the creation of smoother and more natural-looking video sequences.

Text-to-video generation using diffusion models involves a series of interconnected steps orchestrated to transform textual prompts into coherent video sequences, as depicted in Figure 1. Initially, a text prompt is provided as input, which undergoes encoding by a text encoder to capture its semantic meaning, resulting in a fixed-length text embedding. Concurrently, a scheduler controls the diffusion process, modulating noise application to a latent image over successive time steps. This latent image represents the evolving state of video generation and serves as a canvas for subsequent transformations. Incorporating temporal information, timestamp embedding encodes frame sequences, facilitating coherent motion dynamics throughout the video-generation process. Alongside, text embedding, derived from the text prompt, as well as timestamp embedding, are concatenated and utilized by the decoder to synthesize each frame. The motion dynamics within the latent code are shaped by these embeddings, aligning the generated video with the provided text and ensuring smooth temporal progression. Integral to the process is the diffusion model or noise predictor, like U-NET, which models the conditional distribution of subsequent frames based on the current frame and noise level. Cross-frame attention mechanisms capture dependencies between frames, enabling the model to maintain coherence and consistency across the video sequence. Finally, frame-interpolation techniques may be employed to generate intermediate frames for smooth transitions, while background smoothing enhances visual quality and reduces artifacts, ensuring the fidelity of the generated video. Through this orchestrated flow, text-to-video generation using diffusion models seamlessly translates textual descriptions into visually compelling video content.

Diffusion models [7]-[8] acquire the ability to create data by progressively refining samples taken from a noise distribution. Gaussian diffusion models operate under the assumption of a forward noising process, where noise (ϵ) is gradually added to genuine data ($x0 \sim p_{data}$). The mathematical definition denoting the forward noising process is represented in Equation (2):

$$x_t = \sqrt{\gamma(t)}x_0 + \sqrt{1-\gamma(t)}\epsilon, \epsilon \sim N(0, I), t \in [0,1] \quad (2)$$

where $\gamma(t)$ represents a function that steadily decreases from 1 to 0 (referred to as the "noise schedule"). Diffusion models are trained for converse procedure, which counteracts the initial corruptions introduced during the forward process. The mathematical definition denoting the converse noising procedure is represented in Equation (3):

$$E_{x \sim p_{data}, t \sim U(0,1),\ \epsilon \sim N(0,I)}[\|y - f_\theta(x_t; c, t\|^2] \quad (3)$$

where $f\theta$ represents the denoiser model, which is defined by a neural network's parameters, conditioning information is donoted by c, such as textual prompts or class labels, while the target y can be arbitrary noise ϵ,and the denoised input x_0 or $v = \sqrt{1-\gamma(t)}\epsilon - \sqrt{\gamma(t)}x_0$. Combining GANs and diffusion models for video generation involves leveraging the strengths of both approaches. GANs excel at capturing high-frequency details and local structure in video frames, while diffusion models are effective at modeling long-term dependencies and global temporal coherence. By integrating these techniques, researchers have achieved significant advancements in generating high-resolution, photorealistic videos with coherent motion and semantic consistency. The synergy between GANs and diffusion models opens up new possibilities for applications, such as video synthesis, content creation and video editing. Furthermore, ongoing research in this field continues to push the boundaries of what is achievable, paving the way for more sophisticated and life-like video-generation systems in the future.

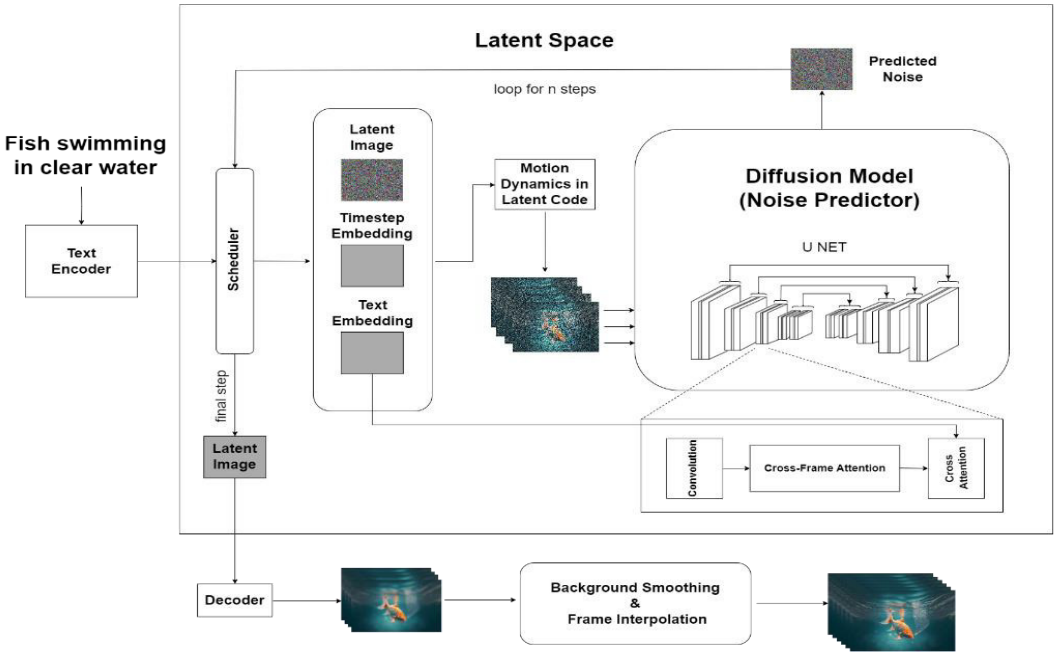


Figure 1. Architecture of a general text-to-video model using stable diffusion.

Video generation with text condition is further divided into two distinct categories: training-based and training-free T2V diffusion methods.

1. Training-based T2V Diffusion Methods: Diverse approaches are used for video synthesis in training- based text-to-video diffusion models, which improve quality by introducing novel training tactics and techniques. The most widely used training-based text-to-video diffusion models are listed here, along with some of their key features.

    - Make-A-Video, as described by Singer et al. [9], revolutionizes the process of learning visual- textual associations by leveraging paired image-text data and extracting video dynamics from unsupervised video datasets. This approach minimizes the need for extensive data-collection efforts, facilitating the generation of diverse and life-like videos through the integration of multiple super-resolution models and interpolation networks.
    - Imagen Video [10], an extension of the established T2I model Imagen [11], introduces a cascaded video-diffusion model consisting of seven interconnected sub-models. The effectiveness of training methodologies, such as classifier-free guidance, conditioning augmentation and v-parameterization, is validated, with additional benefits achieved through progressive distillation techniques aimed at enhancing sampling efficiency.
    - Show-1 [12], introduced by Zhang et al. (2023), innovates by combining pixel-based and latent- based diffusion models for T2V generation. This model operates across four distinct stages, each focusing on different aspects, such as key-frame generation, frame interpolation, super-resolution and latent super-resolution modules, thereby enhancing the overall video-synthesis process.
    - MagicVideo [13], developed by Zhou et al. (2022), employs the Latent Diffusion Model (LDM) to generate videos in latent space, effectively reducing computational overhead and accelerating processing speed. A frame-wise lightweight adaptor is introduced to align distributions, thereby improving temporal relationship modeling and the overall video quality.
    - Latent-Shift [14], as presented by An et al. (2023), prioritizes lightweight temporal modeling inspired by Temporal Shift Module (TSM). This approach involves channel shifting between adjacent frames within convolutional blocks, ensuring the retention of T2I capabilities while generating videos.
    - ModelScope [15], as described by Wang et al. (2023), integrates spatial-temporal convolution and attention mechanisms into the Latent Diffusion Model (LDM)

framework for T2V tasks. Leveraging a mixed training approach utilizing LAION and WebVid datasets, it serves as an open-source benchmark for T2V-synthesis methods.

- VideoFusion [16], proposed by Luo et al. (2023), addresses content redundancy and temporal correlations by decomposing the diffusion process using shared base noise and residual noise along the temporal axis for each frame. Two co-training networks are employed for noise decomposition, ensuring coherence in frame motion and improving the overall video-synthesis quality.

2. Training-free T2V diffusion methods: Training-free text-to-video (T2V) diffusion methods involve direct synthesis or utilize pre-existing models without dedicated training, bypassing explicit training processes.

    - Text2Video-Zero [5] uses a pre-trained text-to-image model for the purpose of video generation, incorporating a cross-attention mechanism and modifying latent code sampling to enhance motion dynamics.
    - DirecT2V [17] and Free-Bloom [18] employ large language models (LLMs) for frame-to-frame descriptions based on user prompts. DirecT2V uses dual-softmax filtering and value mapping for continuity between frames, while Free-Bloom introduces enhancements, like joint noise sampling and step-aware attention shifting.

2.3 Processes Involved in Text to Video Generation

In the intricate process of T2V generation, several key stages unfold. Initially, the system undertakes T2I generation, crafting a single frame that encapsulates the visual representation of the provided textual input. Subsequently, the model engages in cross-frame attention and motion dynamics, employing attention mechanisms to intricately link frames and model the dynamic motion inherent in the video sequence. This step ensures a coherent and realistic flow between frames. Finally, frame Interpolation comes into play, facilitating the creation of intermediate frames. This interpolation process enhances temporal continuity, contributing to the seamless generation of a cohesive and visually compelling video sequence. Together, these stages form a comprehensive pipeline for the transformation of text descriptions into dynamic and visually engaging video content.

2.3.1 Text-to-Image (Generation of a Single Frame)

T2I models serve as a foundational stage and point of entry for T2V models. Currently, there are various cutting-edge models based on stable diffusion and GANs. Table 2 provides a summary of popular studies in this field of study.

Generative Adversial Networks (GANs) [3] are unsupervised machine learning methods that function like supervised ones. Discriminators and generators are the two components of a GAN. While discriminators attempt to determine whether an image is real or not, generators attempt to create new images from the original dataset. In a zero-sum game, both players participate and the game ends when generators trick the discriminator more often than not. The two main advantages of GANs are their speed of inference and their ability to manipulate latent spaces to influence the synthesized outcome. A well-researched latent space in StyleGAN enables principled control over generated images. Diffusion models have made significant strides toward speeding up, but they are still far behind GANs, which only need a single forward pass.

Large-scale text-to-image (T2I) synthesis poses unique challenges, all of which are addressed by StyleGAN- T [19]. The specific requirements include extensive capacity, robust training across diverse datasets, precise text alignment and the ability to balance variation against text alignment according to user preferences. In terms of sample quality and speed, StyleGAN-T performs noticeably better than earlier GANs and surpasses distilled-diffusion models, which were the prior state-of-the-art models in quick T2I synthesis.

Diffusion models were first presented in [7], drawing inspiration from thermodynamics' non-equilibrium state. Fundamentally, in diffusion models, we gradually add Gaussian noise and then figure out how to reverse it.

Encoders play a pivotal role in both text and image encoding, providing a bridge between disparate data

modalities. The encoder's function is to transform complex information from textual and image inputs into a compressed, abstract representation conducive to further analysis or synthesis. In the context of text encoding, natural-language processing techniques are employed to distill semantic meaning, contextual nuances and sentiment from textual data. Simultaneously, in image encoding, visual features, patterns and spatial arrangements are captured and encoded to represent the essence of the visual content. It has been demonstrated that contrasting models, such as CLIP [20], are able to learn stable representations of images that capture both style and meaning to create images by using these representations. An effective technique for learning picture representation from natural-language supervision is Contrastive Language-Image Pre-training (CLIP), which trains both a text encoder and an image encoder simultaneously to anticipate the right pairings of a batch of (image, text) training samples. The target dataset's class names or descriptions are embedded by the learnt text encoder, which then uses this information to create a zero-shot linear classifier at test time.

Two methods were merged by A. Ramesh et al. [21] to solve the text-conditional image-creation problem. They initially trained the CLIP image encoder inverted using a diffusion decoder. The inverter exhibited non-determinism and had the ability to generate several pictures that corresponded to a particular embedding. Beyond T2I translation, the existence of an encoder and its near inverse (the decoder) enabled other possibilities. Encoding and decoding an input image yield semantically identical output images, just like in GAN inversion. By inverting the interpolations of the input images' image embeddings, this technique also made it possible to interpolate between them. But one important benefit of employing the CLIP latent space is that, unlike GAN latent space, which requires trial and error and intensive manual analysis to find these directions, one can semantically edit images by moving in the direction of any encoded text vector. Moreover, the processes of encoding and decoding images offer instruments for discerning the aspects of the image that CLIP acknowledges or ignores. The authors coupled the CLIP-image embedding decoder with an earlier model that produced CLIP image embeddings from a given text caption in order to create a comprehensive generative model of images.

A new degree of flexibility is provided by T2I, which allows users to direct the creative process using natural language. Customizing these models to match user-supplied visual conceptions is still a difficult issue, though. Combining several personalized concepts into a single image, keeping a modest model size and preserving high visual fidelity while permitting creative flexibility are just a few of the difficult problems that face T2I penalization. The Where Pathway and the What Pathway were utilized by Y. Tewel et al. [22] to enhance the user's control over what and where objects should be present in the final image. Using a text encoder, a text input is first turned to a sequence of word embeddings, which is subsequently changed into a sequence of encodings. Next, these encodings are projected *via* the W_k and W_v cross- attention matrices. K routes or W_k, were used to direct objects to their proper locations in the final image. On the other hand, W_v, sometimes called V routes, determined what should be included in the final image.

2.3.2 Motion Dynamics

Motion dynamics in the realm of video generation encompass the representation and understanding of temporal changes, spatial relationships and the flow of motion within a sequence of frames. This concept involves capturing the evolution and transitions of objects or scenes over time, ensuring that the generated videos exhibit realistic and coherent motion patterns. Key considerations include modeling how objects move in relation to each other, recognizing various actions or activities and representing the flow of motion with attention to factors, such as acceleration, deceleration and changes in direction. Effective motion-dynamics modeling also accounts for long-term dependencies, ensuring that the generated videos maintain consistency and contextually relevant temporal sequences. Now let's discuss some of the pivotal methodologies employed to attain motion dynamics.

- Carl Vondrick et al. [23] harnessed the wealth of unlabeled video data to develop a robust model centered around scene dynamics, emphasizing its applicability in both video recognition tasks, such as action classification and video generation tasks, like future prediction. A key contribution lies in the introduction of a generative adversarial network with a spatio-temporal convolutional architecture, strategically designed to disentangle foreground and background components within scenes.
- MoCoGAN [24] explicitly addresses the distinction between content and motion dynamics. It

employs a unique framework where a sequence of video frames is generated by mapping random vectors, with each vector comprising fixed content and a dynamic motion component modeled as a stochastic process. The innovation lies in its adversarial learning scheme, integrating video discriminators with images, to achieve unsupervised motion and content decomposition. The framework excels in generating videos with consistent content yet diverse motion and *vice versa*, showcasing its prowess in capturing and manipulating intricate motion dynamics within generated content.

Table 2. Text-to-image models and their features.

Study & Year	Algorithm	Dataset	Advantages	Limitations	Accuracy
[21], 2022	unCLIP (AR), unCLIP (diffusion)	• MS-COCO	• Complex, diverse & realistic images	• Not good at binding attributes	• unCLIP (AR prior) 10.63 FID Score • unCLIP (Diffusion prior) 10.39 FID score
[22], 2023	Gated Rank-1	• MS	• Less overfit • Better Pareto front	• Over-generalization • High amount of prompt engineering required when combining two or more concepts	• 2.18±0.02 (SEM)
[19], 2023	StyleGAN-T	• CC12m • CC • YFCC 100m (filtered) • Redcaps • LAION-aesthetic- 6+	• Better than DM at low resolution	• Less resolution • Struggles in producing images	• 13.90 FID score

- While existing methods struggled with entangling content tasks and motion in a sole-generator network, Ximeng Sun et al. proposed Two-Stream Variational Adversarial Network (TwoStream- VAN) [25] that adopts a two-stream model to disentangle these tasks. By progressively generating and fusing multiscaler motion alongside corresponding spatial content, the model excels in creating clear and consistent motion, resulting in photorealistic videos.
- Kangyeol Kim et al. [26] introduced an innovative method that entails learning distinct distributions for motion and appearance. Unlike previous methods that discretize motion dynamics, the proposed model utilizes neural ODE to capture the continuous nature of physical-body motion. The two-stage approach involves generation of a sequence of key points using a noise vector and synthesizing videos based on this sequence and an appearance noise vector. The model outperforms recent baselines quantitatively and showcases versatile functionalities, like motion-transfer among different datasets and dynamic frame-rate conversion, indicating promising applications for diverse video-generation scenarios.

2.3.3 Cross-frame Attention

It is imperative to verify that the video accurately depicts the event, that every frame is part of the same film and that there are no jumps in the video. To do this, it is essential to make sure that the frames generated after this are comparable. Cross-frame attention [27][9][28] approaches by various models were used to accomplish this. Table 3 presents the summary of work done in this area. A key idea in the fields of deep learning and artificial intelligence is attention [29]. Regarding neural networks and natural language processing, attention can be conceptualized as a technique that facilitates the model's ability

to concentrate on key information while processing data. Think of it like a spotlight that is focused on the most important portions of the input data. The mathematical definition denoting attention is represented in Equation (4):

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (4)$$

The attention function takes Q, K and V as inputs and its scaling factor is $\frac{1}{\sqrt{d_k}}$. While Q, K and V in self-attention are all part of the same sequence, in cross-attention, K and V are diffusion's conditioning parameters that control generation.

Fundamentally, attention enables the model to give distinct parts in a sequence various levels of importance. For example, when parsing a statement, some words or phrases could be more important than others to grasp the sentence's meaning. The model can adaptively weigh these words or phrases by highlighting the important ones and downplaying the less important ones according to attention mechanisms. When exploring self-attention, where the model evaluates the relationships between components inside the same sequence or cross-attention, which enables it to take into account interactions between elements from separate sequences, this idea becomes quite potent. Many natural language processing tasks now perform much better because to these methods, which also make them more accurate and context-aware. Essentially, attention functions as a kind of spotlight for neural networks, assisting them in identifying and concentrating on the most important data, ultimately producing increasingly complex and contextually-aware models.

A new issue in video creation arises from the addition of time as a dimension. The creation of context-aware videos presents whole new difficulties. In addition to having to pay attention to what is being done, there may be other things requiring focus that might be challenging to manage.

Using a U-Net architecture, J. Ho et al. [28] were able to create longer scenes by utilizing autoregressive extension and classifier-free guidance, which improved text-image linkages. Interleaved Spatial Super Resolution Model (SSR) and Temporal Super Resolution Model are employed by J. Ho et al. [10]. That enabled them to produce 128x128 videos at 24 frames per second, equivalent to 128x768 frames.

In order to introduce the time dimension into a two-dimensional (2D) conditional network, Uriel Singer et al. [9] employed spatiotemporal layers, specifically pseudo-3d convolutional layers and pseudo-3d attention layers. Consequently, cross-frame focus and consistent video were guaranteed. Inspired by separable convolutions, they inserted a 1D convolution after every 2D convolutional layer [30]. This improved temporal information fusion and made greater use of text-to-image (T2I).

Use of Large Language Models or LLMs is a very novel concept that H. Fei et al. [27] proposed. By employing existing LLMs for better simulation that is much closer to reality and produces better results, LLMs are used for scene simulation, making them more performant. Using Dysen (Dynamic Scene Manager)-VDM, a three-layer system combines scene imagination, event-to-DSG conversion and action planning. These layers provide more complicated scenarios by allowing several activities to occur simultaneously by using ChatGPT to interpret the action in a scene. When compared to Make-a-Video, it works noticeably better in longer, more intricate sequences.

2.3.4 Frame Interpolation

Frame Interpolation in text-to-video generation is a crucial step that enhances the temporal flow and smoothness of the generated video sequence. This process involves creating intermediate frames between existing frames to fill the gaps and improve the overall frame rate. By intelligently estimating the content and motion dynamics between consecutive frames, frame interpolation ensures a more fluid and natural transition in the video, contributing to a visually coherent and realistic output. This technique is particularly valuable in scenarios where the original frame rate is low or when generating high-quality videos to achieve a more lifelike and dynamic visual experience from the textual input.

1) Adaptive Separable Convolution:

- Niklaus et al. [30] employed a unique method to utilize a fully convolutional neural network to predict spatially adaptive convolutional kernels for each pixel, eliminating the need for independent-motion estimation and resampling stages. This approach efficiently captures local movement, rendering it resistant to occlusion, brightness fluctuations and blur.

Table 3. Algorithms for cross attention and their features.

Study & Year	Algorithm	Dataset	Advantages	Limitations	Accuracy
[27], 2022	Dysen (Dynamic Scene Manager)-VDM	• UCF-101 • MSR-VTT	• Performs better in scenarios with complex actions. • Uses LLMs for better dynamic	• Depends on ChatGPT	• IS 95.23 • FVD 255.42
[9], 2022	Spatiotemporal layers (pseudo- 3D convolution)	• WebVid-10M • HD-VILA-100M	• Better leverage a T2I architecture. • Allows for better temporal information fusion.	• Can not learn associations between text and phenomenon that can only be inferred in videos • Can generate short videos, and single scene/event.	• FVD 367 • IS 33
[28], 2023	U-Net classifier-free guidance autoregressive video extension	• Kinetics-600 • BAIR Robot Pushing • UCF101	• Enables joint training of text and video.	• Much worse compared to Make-A-Video in temporal information fusion	• FID 295±3 • IS 57±0.62
[10], 2022	SSR & TSR autoregressive video extension	• LAION-400M	• 128×128 videos at 24 frames per second equivalent to 128×768 frames.	• Lack of customizatio n options	• Clip Score 24.27 • Clip R-Precision 86.18

The introduction of separable convolutions significantly reduces processing requirements and the authors achieved substantial memory savings by estimating spatially adaptable 1D convolution kernels. This breakthrough improved previous methods, such as AdaConv, using a specialized encoder-decoder network to estimate kernels for all pixels simultaneously. The study addresses challenges in CNN-based frame interpolation, including occlusion management and resolution adaptation, marking a significant advancement in the field's effectiveness and accessibility.

- To mitigate computational complexity, Chen et al. [31] introduced deformable separable convolution (DSepConv). This technique aims to adaptively estimate kernels with suitable features to handle substantial motion. Subsequent enhancements in their model, known as EDSC [32], enabled the generation of numerous interpolated frames between consecutive frames. Nevertheless, achieving optimal performance for interpolating at arbitrary times remained challenging.

2) Path Selective Interpolation:

- Path Selective Interpolation is a powerful approach based on the principle that each pixel in interpolated frames follows a distinct path in preceding frames. Pioneered by Mahajan et al. [33], this method employs a path-based framework coupled with inverse optical flow

to calculate background motion. By moving and duplicating pixel gradients along anticipated paths, it minimizes issues like holes, chromatic aberrations and visual blur associated with traditional optical-flow techniques. Notably, path-based interpolation preserves the original frequency content and deterministically identifies veiled zones by prioritizing flow consistency, setting it apart from blending-based techniques.

- B. Yan et al. [34] incorporated standard optical-flow algorithms to the framework to control path direction and maintain global path coherency. They also introduced a pixel interlacing model to optimize optical-flow estimation, which significantly boosted efficiency.
- Y. Fan et al. [35] integrated semantic information acquired from input frames to identify crucial pixels *via* optimal-energy minimization, enhancing the precision of motion pattern detection. The inclusion of feature points from input frames resulted in more realistic and visually pleasing results. The approach, designed for processing larger input images and generating an arbitrary number of intermediate frames, aimed to provide exceptional visual quality.

3) Efficient Optical-flow Estimation:

- L. Khachatryan et al. [36] proposed an efficient optical-flow estimation method based on the local all-pass approach, operating in real time at high spatiotemporal resolutions. Using quadratic approximations, a higher-order approach compared to conventional first-order methods, the technique offers precise flow estimations. This unique methodology significantly enhances interpolated frame clarity by reducing motion boundary blur and preserving local geometric information. Notably, the approach addresses challenges in capturing fast, large- scale motion in optical flow-guided frame interpolation, employing a Laplacian cotangent mesh constraint for accurate motion representation, even in the presence of complex non- rigid motion. The implementation of a mesh system with one vertex per pixel demonstrates remarkable results in the Middlebury interpolation-error criterion, showcasing its potential applicability in optical flow-guided frame interpolation.

4) Real-time Frame Interpolation via GAN:

- J. van Amersfoort et al.'s work, "FIGAN" [37], represents a significant advancement in GAN- based frame interpolation. Demonstrating an impressive average runtime speedup of ×47 compared to rival approaches, FIGAN excelled in real-time YouTube 8M movies, establishing itself as the most sophisticated technique in the field. The authors introduced a multi-scale network with a mixed perceptual loss function, integrating spatial transformer networks with conventional optical-flow modeling. FIGAN's notable improvement over prior methods, such as SepConv-Lf, lies in its ability to generate higher-quality interpolated frames with fewer training parameters. This efficiency is particularly crucial in resource-limited scenarios, such as real-time video processing.
- S. Wen et al. [38] introduced a network comprising two concatenated GANs. The first GAN captures motion from training video clips and integrates finer frame data to enhance output quality, while the other generates frame details. To address issues related to noise that affected earlier approaches, they employed the Normalized Product Correlation Loss (NPCL). This innovative framework achieved visually appealing effects and demonstrated remarkable performance, particularly attributed to the effective use of NPCL, showcasing notable progress in the domain of GAN-based frame interpolation.
- J. Xiao et al.'s work, "FI MSAGAN"[39], used multi-scale dense attention generative adversarial networks to interpolate interim frames. FI MSAGAN accomplished more efficient fusion of local and global information details by using multiple generators and discriminator networks with varied sized input images. Its accuracy and runtime were found to be comparable to those of other state-of-the-art approaches.

5) Phase-based Frame Interpolation:

- P. Didyk et al. led the first study on phase-based frame interpolation. Their approach, as presented in [40], was based on the hypothesis that individual pixel phase-shift values

might contain limited motion information. However, the method struggled to effectively handle large movements, resulting in less than optimal results.

- S. Meyer et al. [41] developed a coarse-to-fine framework with a multi-scale pyramid level structure to communicate phase information. To address the issue of tolerating large motions, they capped phase-shift values. Phase-shift values were computed, phase differences were used to interpolate frames and amplitude values were used to blend the interpolated frames. Their algorithm was made up of these three essential steps. This method's inability to handle high- frequency motion resulted in blurry output in areas with small but high-frequency motion.
- The earliest phase-based techniques, including those introduced by the authors of [41], were characterized by the manual adjustment of parameters, such as amplitude and phase shift, to produce images. However, this manual adjustment process imposed limitations on the method's adaptability and efficiency. In order to estimate amplitude and phase-shift values directly, S. Meyer et al. [42] proposed the Phase-Net neural network architecture. Eliminating the need for manually adjusted parameters, this innovation significantly expanded the technique's ability to handle a broader spectrum of motion and frequencies. The authors used a decoder-only Phase-Net design, in which all levels were identical except for the last layer. Simulating a level-wise decomposition of phase information, the interpolated frame's resolution progressively grew as one proceeded up the network levels. By estimating parameters directly, Phase-Net was able to achieve more robust and diversified frame-interpolation capabilities than it could have achieved with hand-tuned phase-based approaches.

6) Bidirectional Optical Flow Estimation:

- H. E. Ahn et al. [43] proposed a method to effectively estimate bidirectional optical flow at lower resolutions and then reconstruct high-resolution optical flow. This multi-scale motion-reconstruction network works especially well with 4K footage and other high-resolution video frames. The method begins with bidirectional optical flow estimation at a lower resolution (e.g. one-fourth of the original resolution for 4K recordings). A multi-scale reconstruction strategy is then used to reconstruct the estimated optical-flow to match the original resolution. The authors trained their network using a variety of loss functions, such as adversarial loss, consistency loss and multi-scale smoothening loss. This all-encompassing strategy tack led the challenges of high-resolution video-frame interpolation and generated computationally efficient results while maintaining visual quality.
- In addition to interpolating frames, S. Y. Kim et al. [44] acknowledged the significance of boosting spatiotemporal resolution in contemporary videos. Both objectives were sought after by their combined model, which provided a thorough response to the requirements of high-resolution video footage.
- W. Bao et al. [45] used flow vectors and convolutional kernels to build an adaptive warping layer that generated output pixels by using both flow vectors and motion-compensation kernels. This method not only made frame interpolation better, but it also made other video enhancement methods, such as super-resolution, possible.
- By taking depth information into account, techniques such as depth-aware flow projection (DAIN) [46] specifically addressed issues with occlusion. Using depth-aware frame synthesis networks, kernel estimation and context extraction, DAIN presented an effective approach. With less parameters and more effective performance, DAIN produced impressive results by highlighting the significance of depth in frame interpolation.
- The incorporation of meta-learning approaches by M. Choi et al. [47], as well as the usage of attention networks by J. Xiao et al. [39] and M. Choi et al. [48], improved the efficiency of frame-interpolation techniques. These methods increased performance and efficiency by concentrating on attention and adaptation within feature representations.

3. DATASETS

Several datasets are instrumental in the development and evaluation of T2V generation models, each

offering unique challenges and characteristics that cater to specific research goals. Table 4 presents the summary of some popular datasets used in T2V generative models.

Table 4. Datasets used in T2V.

Dataset	Description
UCF-101 [49]	13,320 videos with 101 action categories; Realistic action videos collected form YouTube.
MSR-VTT [50]	10,000 videos with 20 classes; annotation of 20 sentences per video clip.
WebVid-10M [51]	10.7M video-caption pairs. Short videos with textual descriptions; 2.5 M video subset with a total of 52K video hours.
HD-VILA-100M [52]	100M video-caption pairs. 720p videos ranging over a total of 371.5K video hours.
Kinetics-600 [53]	480K video clips with 600 action classes; video duration of 10-sec.
BAIR Robot Pushing [54]	64x64 images of a robot pushing objects on a tabletop; conditioned on 2 frames, predicting 14 frames.

4. METRICS AND RESULTS DISCUSSION

Text-to-video generative models are commonly evaluated using metrics, such as Frechet Inception Distance (FID) [55], Clip score [20], among others. However, the Fréchet Video Distance (FVD) [4] metric stands out as a superior choice. FVD incorporates both visual quality and temporal coherence, providing a more comprehensive assessment of generated videos [1]. In contrast to FID, which only looks at static-picture quality, FVD takes into account the dynamic elements that are important for video assessment. As a result, FVD becomes a more reliable tool for evaluating the general coherence and integrity of produced video sequences. Thus, we employed it as the standard metric for our testing. FVD works on trajectories that reflect the routes of moving objects in the movies, drawing inspiration from the Fréchet distance used in curve-similarity evaluations. Through the application of the Fréchet distance concept to video analysis, FVD allows researchers to evaluate the entire motion patterns holistically and identify subtle changes or similarities across different video sequences. Understanding the underlying motion dynamics is essential for good video interpretation in a number of computer-vision disciplines, such as action recognition, anomaly detection and content-similarity evaluation, where its application is widespread.

Trajectory representation, spatial point correspondence and trajectory-distance evaluation are laborious steps in the computation of FVD. By using this technique, FVD provides a sophisticated assessment of video content, making it possible to spot minute changes in motion patterns that could go unnoticed by conventional video-comparison criteria. FVD is a useful tool for researchers and practitioners trying to quantify and comprehend the nuances of motion dynamics in video data; the lower the value, the more comparable the films. FVD is still a crucial indicator in the ever-evolving field of video analysis, helping progress areas like automatic video-content classification, human-behavior analysis and surveillance.

Instead of using Google's initial implementation [56], we adopted FVD, which was developed by StyleGAN- V [57]. It uses approximations for faster FVD calculation and the errors are within the range of 1e-6, which is a reasonal trade off. We used MSR-VTT [50] dataset which has 10000 videos and used the standard test-train split. For each test, the video shortest prompt (caption) was selected, so as to reduce test size and ensure faster testing and all test videos were scaled to 256x256 resolution. For all models in Table, 5 2990 videos with 16 frames each were generated. These were further scaled to 256x256 resolution to have uniformity in tests. All videos were then converted into frames and respective FVD scores for 16 frames were calculated.

For experimentation, we utilized an Nvidia RTX Quadro A5000 (24 GB VRAM), 64 GB RAM and an Intel Xenon 20 core CPU. Table 5 shows the performance (FVD scores) of the various pre-trained models along with their inference time. Show-1 [12] yielded the best results for 16 frames; however, it employed 4 models (1 generation model, 2 SR models, 1 interpolation model) and took the longest time (10min-12min) compared to other examined models (15sec-20sec) per video. Hotshot-XL [6] yielded

good results for 8 frames, but when tested for 16 frames, it performed significantly worse. We evaluated Text2Video- zero with three base T2I models and observed that better performing T2I models produced better T2V outcomes. The training steps of text-to-video models vary due to differences in architectures and methodologies. Show1 follows a modular approach, with distinct modules undergoing specific numbers of training steps: Keyframe Module (120,000 steps), Interpolation Module (40,000 steps) and First and Second Super-resolution modules (40,000 and 120,000 steps, respectively). Zero-shot models leverage pre-existing T2I architectures, eliminating the need for a training phase. Stable-Diffusion-v1.5, a base T2I model, underwent training over 595,000 steps at a resolution of 512x512. Dreamlike-diffusion-1.0 and Dreamlike- photoreal-2.0 are derived from Stable-Diffusion-v1.5, thereby inheriting its training characteristics. Potat1, a text-to-video finetuning model, undergoes around 2,500 steps with a consistent learning rate of 5e-6, leveraging ModelsScope's architecture for rapid adaptation.

Table 5. Comparison of various open-source models and their FVD scores along with their inference time.

Model	**FVD Score (fvd2048_16f)**	**Inference Time(sec/video)**
Text2Video-zero [5] (dreamlike-photoreal-2.0)	1420.9068	18.7556
Text2Video-zero [5] (dreamlike-diffusion-1.0)	1519.5902	18.2612
Text2Video-zero [5] (stable-diffusion-v1-5)	1498.2528	18.6525
Show-1 [12]	1094.6304	654.8715
Text-to-video-finetuning [15] (camenduru/potat1)	2132.1784	15.1471
Hotshot-XL [6]	1421.3931	19.9149

5. OPEN CHALLENGES

The field of T2V generation confronts several prominent challenges that impede the seamless transition from textual descriptions to visually coherent and compelling video content. One significant hurdle is the lack of coherence in the generated videos, which necessitates the development of advanced methods to ensure smooth transitions between frames and scenes, preventing abrupt changes and disjointed visual elements.

Penalization is another critical aspect that demands attention, with the need to explore techniques that can tailor generated videos to individual preferences and contextual details, making the content more engaging and relevant to diverse audiences. The persistent issue of low resolution in generated videos calls for innovative approaches to enhance visual quality, involving sophisticated upscaling methods and the preservation of fine details. Frame interpolation, a fundamental process in video generation, faces challenges related to the lack of intricate details, requiring solutions that produce smoother and more realistic transitions between frames. Background-smoothening techniques must be developed to eliminate artifacts and inconsistencies, ensuring a natural flow in the visual elements of the generated videos. Moreover, the field lacks comprehensive study and survey papers, which hinders a thorough understanding of existing research and limits the identification of critical gaps and opportunities for advancement.

Additionally, the language dependency on English presents a significant limitation, urging the exploration of models and approaches that can accommodate multiple languages to enhance inclusivity and accessibility. Addressing these multifaceted challenges collectively will propel the field towards the development of more coherent, personalized and culturally-aware text-to-video generation systems.

6. CONCLUSION AND FUTURE SCOPE

In conclusion, this article has provided a comprehensive overview of the advancements in text-to-video generation leveraging GANs and stable diffusion models. The synthesis of these two powerful techniques has demonstrated promising results in overcoming challenges associated with coherence, personalization and visual quality. GANs have proven effective in capturing intricate details and generating realistic video frames, while Stable Diffusion models contribute to stable and coherent video synthesis over extended sequences. The synergistic integration of these approaches holds great potential for addressing the limitations identified in the existing literature. As we move forward, it is imperative

to continue exploring innovative combinations of GANs and stable diffusion models, pushing the boundaries of text-to-video synthesis to new heights. Moreover, future-research directions should prioritize scalability, real-time processing and ethical considerations, ensuring the responsible development and deployment of these advanced techniques in diverse applications. The amalgamation of GANs and stable diffusion models signifies a promising trajectory for the evolution of text-to-video synthesis, offering a rich landscape of possibilities for researchers, practitioners and industries invested in multimedia content generation.

In the future, improved temporal modeling techniques within GANs and stable diffusion frameworks can be used to overcome the coherence challenge. Personalization gaps can be bridged by integrating attention mechanisms and reinforcement learning for a more nuanced understanding of individual preferences. To tackle low resolution, exploring novel upscaling methods, incorporating perceptual loss functions and fine-tuning architectures can significantly enhance visual quality. Future studies should focus on creating versatile models capable of handling diverse content types through domain adaptation and cross- modal learning. In summary, the future scope lies in integrating cutting-edge technologies to create more coherent, personalized and high-resolution text-to-video generation systems.

REFERENCES

[1] A. Singh, "A Survey of AI Text-to-Image and AI Text-to-Video Generators," arXiv preprint, arXiv: 2311.06329, Nov. 2023.
[2] Z. Xing et al., "A Survey on Video Diffusion Models," arXiv preprint, arXiv: 2310.10647, October 2023.
[3] I. J. Goodfellow et al., "Generative Adversarial Nets," Advances in Neural Information Processing Systems, arXiv: 1406.2661, pp. 2672–2680, 2014.
[4] T. Unterthiner et al., "Towards Accurate Generative Models of Video: A New Metric & Challenges," arXiv preprint, arXiv: 1812.01717, 2018.
[5] L. Khachatryan et al., "Text2Video-zero: Text-to-image Diffusion Models are Zero-shot Video Generators," arXiv preprint, arXiv: 2303.13439, March 2023.
[6] J. Mullan et al., "Hotshot-XL," [Online], Available: https://github.com/hotshotco/Hotshot-XL, October 2023.
[7] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan and S. Ganguli, "Deep Unsupervised Learning Using Nonequilibrium Thermodynamics," arXiv preprint, arXiv:1503.03585, 2015.
[8] Y. Song and S. Ermon, "Generative Modeling by Estimating Gradients of the Data Distribution," arXiv preprint, arXiv: 1907.05600, July 2019.
[9] U. Singer et al., "Make-a-video: Text-to-video Generation without Text-video Data," arXiv preprint, arXiv: 2209.14792, 2022.
[10] J. Ho et al., "Imagen Video: High Definition Video Generation with Diffusion Models," arXiv preprint, arXiv: 2210.02303, 2022.
[11] C. Saharia, "Photorealistic Text-to-image Diffusion Models with Deep Language Understanding," Proc. of the 36th Conf. on Neural Information Processing Systems (NeurIPS 2022), arXiv: 2205.11487, 2022.
[12] D. Junhao Zhang et al., "Show-1: Marrying Pixel and Latent Diffusion Models for Text-to-video Generation," arXiv preprint, arXiv: 2309.15818, September 2023.
[13] Daquan Zhou et al., "MagicVideo: Efficient Video Generation with Latent Diffusion Models," arXiv preprint, arXiv: 2211.11018, November 2022.
[14] J. An et al., "Latent-Shift: Latent Diffusion with Temporal Shift for Efficient Text-to-video Generation," arXiv preprint, arXiv: 2304.08477, April 2023.
[15] J. Wang et al., "ModelScope Text-to-video Technical Report," arXiv preprint, arXiv: 2308.06571, August 2023.
[16] Z. Luo et al., "VideoFusion: Decomposed Diffusion Models for High-quality Video Generation," Proc. of the 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 10209-10218, 2023.
[17] S. Hong, J. Seo, S. Hong, H. Shin and S. Kim, "Large Language Models Are Frame-level Directors for Zero-shot Text-to-video Generation," arXiv preprint, arXiv: 2305.14330, May 2023.
[18] H. Huang, Y. Feng, C. Shi, L. Xu, J. Yu and S. Yang, "Free-Bloom: Zero-shot Text-to-video Generator with LLM Director and LDM Animator," arXiv preprint, arXiv: 2309.14494, September 2023.
[19] A. Sauer, T. Karras, S. Laine, A. Geiger and T. Aila, "Stylegan-t: Unlocking the Power of GANs for Fast Large-scale Text-to-image Synthesis," arXiv preprint, arXiv: 2301.09515, 2023.
[20] J. Hessel, A. Holtzman, M. Forbes, R. Le Bras and Y. Choi, "Clipscore: A Reference-free Evaluation Metric for Image Captioning," arXiv preprint, arXiv: 2104.08718, 2022.
[21] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu and M. Chen, "Hierarchical Text-conditional Image Generation with Clip Latents," arXiv preprint, arXiv: 2204.06125, 2022.
[22] Y. Tewel, R. Gal, G. Chechik and Y. Atzmon, "Key-locked Rank One Editing for Text-to-image Personalization," arXiv preprint, arXiv: 2305.01644, 2023.

[23] C. Vondrick, H. Pirsiavash and A. Torralba, "Generating Videos with Scene Dynamics," arXiv preprint, arXiv: 1609.02612, 2016.
[24] S. Tulyakov, M.-Y. Liu, X. Yang and J. Kautz, "MoCoGan: Decomposing Motion and Content for Video Generation," Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition, DOI: 10.1109/CVPR.2018.00165, Salt Lake City, USA, 6 2018.
[25] X. Sun, H. Xu and K Saenko, "TwoStreamVAN: Improving Motion Modeling in Video Generation," arXiv preprint, arXiv: 1812.01037, 2020.
[26] K. Kim et al., "Continuous-time Video Generation *via* Learning Motion Dynamics with Neural ODE," arXiv preprint, arXiv: 2112.10960, 2021.
[27] H. Fei, S. Wu, W. Ji, H. Zhang and T.-S. Chua, "Dysen-VDM: Empowering Dynamics-aware Text-to-video Diffusion with Large Language Models," arXiv preprint, arXiv: 2308.13812, 2023.
[28] J. Ho, T. Salimans, A. Gritsenko, W. Chan, M. Norouzi and D. J. Fleet, "Video Diffusion Models," arXiv preprint, arXiv: 2204.03458, 2022.
[29] A. Vaswani et al., "Attention Is All You Need," arXiv preprint, arXiv: 1706.03762, 2017.
[30] S Niklaus, L Mai and F. Liu, "Video Frame Interpolation *via* Adaptive Convolution," arXiv preprint, arXiv: 1703.07514, 2017.
[31] X. Cheng and Z. Chen, "Video Frame Interpolation *via* Deformable Separable Convolution," Proc. of the AAAI Conf. on Artificial Intelligence, vol. 34, pp. 10607–10614, 2020.
[32] X. Cheng and Z. Chen, "Multiple Video Frame Interpolation via Enhanced Deformable Separable Convolution," IEEE Trans. on Pattern Analysis And Machine Intell., vol. 44, no. 10, pp. 7029-7045, 2021.
[33] D. Mahajan, F.-C. Huang, W. Matusik, R. Ramamoorthi and P. Belhumeur, "Moving Gradients: A Path-based Method for Plausible Image Interpolation," ACM Transactions on Graphics, vol. 28, no. 3, Article no.: 42, pp 1–11, 2009.
[34] B. Yan and Y. Chen, "Low Complexity Image Interpolation Method Based on Path Selection," Journal of Visual Communication and Image Representation, vol. 24, pp. 661–668, 2013.
[35] Y. Fan, N. Yoda, T. Igarashi and H. Ma, "Path-based Image Sequence Interpolation Guided by Feature Points," Proc. of the 2016 IEEE Int. Conf. on Image Processing (ICIP), DOI: 10.1109/ICIP.2016.7532421, Phoenix, USA, 2016.
[36] T. Jayashankar, P. Moulin, T. Blu and C. Gilliam, "Lap-based Video Frame Interpolation," Proc. of the 2019 IEEE International Conference on Image Processing (ICIP), DOI: 10.1109/ICIP.2019.8803484, Taipei, Taiwan, 2019.
[37] J. van Amersfoort et al., "Frame Interpolation with Multi-scale Deep Loss Functions and Generative Adversarial Networks," arXiv preprint, arXiv: 1711.06045, 2019.
[38] S. Wen et al., "Generating Realistic Videos from Keyframes with Concatenated GANs," IEEE Trans. on Circuits and Systems for Video Tech., vol. 29, pp. 2337–2348, 2019.
[39] J. Xiao and X. Bi, "Multi-scale Attention Generative Adversarial Networks for Video Frame Interpolation," IEEE Access, vol. 8, pp. 94842–94851, 2020.
[40] P. Didyk, P. Sitthi-Amorn, W. Freeman, F. Durand and W. Matusik, "Joint View Expansion and Filtering for Automultiscopic 3D Displays 3D Stereo Content Multiview Content," ACM Trans. on Graphics, vol. 32, no. 6, Article no. 221, pp. 1–8, 2013.
[41] S. Meyer, O. Wang, H. Zimmer, M. Grosse and A. Sorkine-Hornung, "Phase-based Frame Interpolation for Video," Proc. of the 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), DOI: 10.1109/CVPR.2015.7298747, Boston, USA, 2015.
[42] S. Meyer, A. Djelouah, B. McWilliams, A. Sorkine-Hornung, M. Gross and C. Schroers, "Phasenet for Video Frame Interpolation," arXiv preprint, arXiv: 1804.00884, 2018.
[43] H. E. Ahn, J. Jeong, J. Woo Kim, S. Kwon and J. Yoo, "A Fast 4K Video Frame Interpolation Using a Multi-scale Optical Flow Reconstruction Network," Symmetry, vol. 11, no. 10, Article no. 1251, 2019.
[44] S. Ye Kim, J. Oh and M. Kim, "FISR: Deep Joint Frame Interpolation and Super-resolution with a Multi-scale Temporal Loss," Proc. of the 34th AAAI Conf. on Artificial Intelligence (AAAI-20), pp. 11278-11286, 2019.
[45] W. Bao, W.-S. Lai, X. Zhang, Z. Gao and M.-H. Yang, "MEMC-Net: Motion Estimation and Motion Compensation Driven Neural Network for Video Interpolation and Enhancement," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 3, 2018.
[46] W. Bao, W.-S. Lai, C. Ma, X. Zhang, Z. Gao and M.-H. Yang, "Depth-aware Video Frame Interpolation," arXiv preprint, arXiv: 1904.00830, 2019.
[47] M. Choi, J. Choi, S. Baik, T. H. Kim and K. Mu Lee, "Scene-adaptive Video Frame Interpolation *via* Meta-learning," arXiv preprint, arXiv: 2004.00779, pp.9444-9453, 2020.
[48] M. Choi, H. Kim, B. Han, N. Xu and K. Mu Lee, "Channel Attention Is All You Need for Video Frame Interpolation," Proc. of the 34th AAAI Conf. on Artificial Intelligence (AAAI-20), pp. 10663- 10671, 2020.
[49] K. Soomro, A. R. Zamir and M. Shah, "UCF101: A Dataset of 101 Human Actions Classes from Videos in the Wild," Proc. of the 1st Int. Workshop on Action Recognition with Large Number of Classes, arXiv: 1212.0402, 2012.

[50] J. Xu, T. Mei, T. Yao and Y. Rui, "MSR-VTT: A Large Video Description Dataset for Bridging Video and Language," Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), DOI: 10.1109/CVPR.2016.571, Las Vegas, USA, 2016.

[51] M. Bain, A. Nagrani, G. Varol and A. Zisserman, "Frozen in Time: A Joint Video and Image Encoder for End-to-end Retrieval," arXiv preprint, arXiv: 2104.00650, 2021.

[52] H. Xue et al., "Advancing High-resolution Video-language Representation with Large-scale Video Transcriptions," Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), DOI: 10.1109/CVPR52688.2022.00498, pp. 5026-5035, 2021.

[53] J. Carreira, E. Noland, A. Banki-Horvath, C. Hillier and A. Zisserman, "A Short Note about Kinetics-600," arXiv preprint, arXiv: 1808.01340, 2018.

[54] F. Ebert, C. Finn, A. X. Lee and S. Levine, "Self-supervised Visual Planning with Temporal Skip Connections," Proc. of the 1st Conf. on Robot Learning (CoRL 2017), Mountain View, USA, pp. 1-13, 2017.

[55] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler and S. Hochreiter, "Gans Trained by a Two Time-scale Update Rule Converge to a Local Nash Equilibrium," Proc. of the 31st Conf. on Neural Information Processing Systems (NIPS 2017), pp. 1-12, Long Beach, USA, 2018.

[56] Github, "Google-research," [Online], Available: https://github.com/google-research/google-research/blob/master/frechet_video_distance.

[57] I. Skorokhodov, S. Tulyakov and M. Elhoseiny, "StyleGAN-V: A Continuous Video Generator with the Price, Image Quality and Perks of StyleGAN2," Proc. of the IEEE CVPR 2022, pp. 3626-3636, 2021.

ملخص البحث:

يتطلّب التّحدي المتمثّل في السّعي إلى توليد الفيديو من النّصوص تحويل الأوصاف النّصّية إلى مقاطع فيديو حقيقية ومتماسكة. ولقد تعّرض هذا المجال البحثي إلى تطوّرات عديدة في السنوات الأخيرة، مع تطوّر النّماذج الاندماجية والشّبكات الاستدراكية التّوليدية (GANs).

تبحث هذه الورقة في أحدث نماذج التّحويل من نُصوص إلى صور فيديو، والخطوات الّتي يتضمّنها توليد صور الفيديو من النّصوص، بما فيها التّماسُك المؤقت، وتوليد الفيديو، إلى جانب ترميز النّصوص. كذلك نتناول التّحدّيات الّتي ينطوي عليها تحويل النّصوص إلى فيديو وأحدث ما تمّ التّوصّل إليه من الطّرق للتغلّب عليها. هذا إلى جانب تحليل ومراجعة مجموعات البيانات والمقاييس التي يكثر استخدامها في هذا المجال.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 02, June 2024.

A MODEL DRIVEN FRAMEWORK FOR COLLABORATIVE AND DYNAMIC DESIGN AND IMPLEMENTATION OF NOSQL-ORIENTED DATA WAREHOUSES

Khadija Letrache and Mohammed Ramdani

(Received: 10-Feb.-2024, Revised: 12-Apr.-2024, Accepted: 27-Apr.-2024)

ABSTRACT

Nowadays, modernizing the data warehouse ecosystem is a key challenge in decision-support systems. This modernization is crucial for ensuring scalability and meeting evolving business requirements, especially with the advent of big data. A promising solution involves implementing data warehouses with contemporary data stores, such as NoSQL. In this context, we introduce in this paper a framework that leverages Model-driven Architecture (MDA) to design and implement modern data warehouses across NoSQL data stores. Our MDA approach aims to offer a collaborative, dynamic and reusable process for developing NoSQL-oriented data warehouses tailored to specific project requirements. It facilitates the automatic and dynamic generation of a hybrid data-warehouse model from its conceptual model, which encompasses structural, domain and access parameters. Moreover, our framework includes the generation of implementation code for the data warehouse, along with a set of files to validate, document and illustrate the data-warehouse schema on a target platform. Finally, we present a detailed case study to highlight the effectiveness of our MDA framework.

KEYWORDS

Data warehouse, Model driven architecture, Metamodel, Dynamic transformation rule, NoSQL, Document, Key-value, Column-Family, Graph.

1. INTRODUCTION

A few years ago, traditional data warehouses were implemented on relational systems and utilized for analyzing operational data derived from relational databases [1]. However, with the advent of big data, numerous voices proclaimed the end of the data-warehousing era, asserting that such systems had become obsolete [2]-[3]. Meanwhile, according to a survey conducted by TDWI [4], it was found that a significant majority of enterprises, approximately 75%, continue to utilize on-premises or cloud-based data warehouses. The survey also revealed that more than a half of these enterprises have transitioned from analyzing only traditional structured data to exploring new types of data. This shift introduced a new generation of data-warehousing systems, deploying a new data stack that covers the entire decision-support ecosystem, from data storage to data visualization. This process is commonly known as data-warehouse modernization or data-warehouse augmentation [5].

Consequently, the focus for Business Intelligence (BI) developers often shifts towards mastering various technologies rather than addressing the analysis of actual requirements. In this context, we introduce a collaborative framework in this paper that enables the automatic generation of data-warehouse models across different NoSQL data stores. Our approach leverages the power of the Model-driven Architecture (MDA) paradigm [6] to automate the process of modeling and implementing data warehouses on selected platforms. The objective of the proposed framework is to guide BI developers in constructing their data warehouses on any desired platform, even with limited expertise in that particular platform. Furthermore, automation allows developers to seamlessly incorporate best practices from previous designs into future projects, enabling them to leverage valuable feedback and accumulate essential knowledge.

The proposed framework employs dynamic transformation rules to generate a data-mart model that is driven by the project's requirements. Traditionally, in classical MDA approaches, users define static transformation rules to automatically generate a specific model, either star or snowflake schema and its implementation code. However, when applied to NoSQL-oriented data warehouses, the obtained code provides limited metadata about the generated model. In contrast, our approach maximizes the

K. Letrache and M. Ramdani are with Informatics Department, LIM Laboratory, Faculty of Science and Techniques of Mohammedia, University Hassan II, Casablanca, Morocco Emails: khadija.letrache@fstm.ac.ma and mohammed.ramdani@fstm.ac.ma

utility of MDA by utilizing it as a channel for conveying design and tuning practices through predefined transformation rules. Consequently, the model generated by our framework is recommended based on the project's specific requirements, ensuring a more tailored and efficient design process.

To achieve this objective, this paper introduces four metamodels designed to effectively represent the target data stores: document, column-family, key-value and graph stores. Furthermore, we propose an extended conceptual metamodel that encompasses all the essential aspects required during the data warehouse-design process. Additionally, we introduce dynamic transformation rules that automatically derive destination models from the conceptual model.

Leveraging MDA model-to-text transformation rules, we then automatically generate the DW implementation code and documentation. These outputs are presented in three distinct files: The first contains the DW implementation code for a representative platform, which, in the case of NoSQL stores, offers limited metadata due to their schema-less nature. The second file is a data template, generated in JSON format, to provide guidance on data organization and storage. Lastly, a validation file is included to offer additional metadata and aid in verifying and validating the integrity of the DW data during the loading phase.

The remainder of this paper is organized as follows: Section 2 discusses the most relevant works related to NoSQL-oriented data warehouses. Section 3 outlines our approach and introduces the proposed conceptual metamodel, providing a background and context for our proposal. In Section 4, we delve into the metamodel and transformation rules for document-oriented data warehouses, while Section 5 focuses on key-value oriented DWs. The discussion on column-family DWs is presented in Section 6 and graph- oriented DWs are examined in Section 7. A case study that illustrates our approach is detailed in Section 8. In Section 9, we describe the generated code and documentation files and the used transformation rules. Finally, Section 10 concludes the paper, summarizing our findings and proposing directions for future research and perspectives.

2. RELATED WORKS

The primary objective of a data warehouse is to effectively store enterprise data, enabling data analysis and decision-making. Traditionally, DWs have been implemented using relational database management systems (RDBMS) as the foundational layer for data storage [1]. However, with the emergence of NoSQL databases, there has been a growing suggestion to utilize these data stores for the data-warehousing systems. In [7], the authors Chevalier et al. have studied the transition from a data warehouse conceptual model to NoSQL logical model; namely, column-family (CF) and document oriented stores. The authors provide the outline rules to model a data warehouse on document or column-family oriented stores. The same authors have proposed in [8] the implementation of OLAP cuboids in a document oriented data store designed as flat and shattered models and using materialized views. Boussahoua et al. [9] conducted a comprehensive study on implementing data warehouses in column-family oriented stores. The authors employed a k-means clustering method to effectively identify the necessary column families to group attributes. Other studies have investigated the implementation of data warehouses in graph- oriented databases, including the approaches proposed by Sellami [10] and Vaisman [11]. Additionally, Benhissen et al. [12] employed a shattered model, utilizing a distinct node for each attribute.

On the other hand, there has been research dealing with the development and automation of NoSQL-oriented data warehouses through model-driven approaches. In [13], the authors proposed an MDA approach for generating data warehouses in the four NoSQL data stores. The authors proposed a single metamodel that encompasses the basic concepts of document, column-family, key-value and graph rather than data-warehousing concepts. The proposed metamodel contains a "Value" class in all stores which is not a concept related to metamodeling. The authors also proposed four metamodels for specific database systems in each store type. These latter are generated from the generic PSM. The authors illustrated and provided transformation rules to obtain a column-family PSM from a relational PSM. In another related work focusing on Cassandra, the author proposed in [14] an approach for generating a data warehouse model within Cassandra, based on an information model that represents the DW conceptual model, but does not use data-warehousing concepts. Additionally, Yangui et al. [15] put forward an approach for mapping data from a multidimensional model of document and column-family

oriented databases as a flat model in both cases. More recently, Oukhouya et al. [16] proposed an MDA approach for DW design using a generic PIM (Platform Independent Model) metamodel for both NoSQL and relational platforms. The proposed metamodels present some drawbacks, such as the association between measures and dimensions, the cardinality of the primary key in the relational model or the generalization between identifiers, atomic fields and documents in the MongoDB metamodel. In a recent work, [17] deals with key-value oriented DWs using an MDA approach. Finally, another work proposed by Abdelhedi et al. [18] aims to create document-oriented data warehouses from relational data lakes using an MDA approach. However, the proposed transformation rules are related to models rather than metamodels, transforming data records into documents, for instance.

By analyzing the aforementioned works and their contributions to the modernization of the data warehouse ecosystem, we notice that all approaches are based on static transformation rules, where the user must independently choose a specific target model driven solely by the description of multidimensional concepts. Thus, our approach aims to guide users in designing a DW model tailored to each specific project, based on dynamic transformation rules. These rules are implemented through a collaborative process, driven by environmental parameters and developers' feedback. The objective of our approach is to leverage the model-driven paradigm to create a collaborative framework that facilitates the sharing and capitalization of developers' knowledge. Table 1 summarizes the state-of-the-art and outlines our contribution according to the following criteria: whether it is an MDA approach, the type of proposed transformation rules (static *vs.* dynamic, manual *vs*. automatic), the source model of the transformations, the target NoSQL store and model (flat using a single "table", star, snowflake, hybrid which is a combination of different models) and finally, the nature of the code generated by the proposed approach.

Table 1. Comparative analysis of related works on NoSQL-oriented data warehouses.

Paper	MDA	Trans. Rules	Source Model	Target NoSQL Store	Target Model	Generated code
[7]-[8]	Yes	Static	Multidimensional	CF/Document	Flat model/Star Model	-
[9]	No	Dynamic - Automatic (k-means)	-	CF	Flat	-
[10]	No	Static - Manual	Multidimensional	Graph	Snowflake	-
[12]	No	Static - Manual	Multidimensional	Graph	Shattered	-
[11]	No	-	Multidimensional	Graph	Star-Snowflake	-
[13]	Yes	Static - Automatic	Relational PSM	Document - CF - Key value- Graph	Generic NoSQL	-
[14]	Yes	Static	Information Model	CF	Flat	Impl. code
[15]	Yes	Static	Multidimensional	Document - CF	Flat	-
[16]	Yes	Static - Automatic	Generic	Relational - Document - CF	Star (relational, document) - Flat (CF)	Impl. code
[17]	Yes	Static - Automatic	Relational	Key value	Flat	-
[18]	Yes	Static	Relational	Document	-	Data file
Our Approach	Yes	Dynamic - Automatic	Extended Multidimensional	Document - CF - Key value- Graph	Hybrid (with vertical partitioning)	Impl. code- Validation file - Data template

3. OUR MDA APPROACH

3.1 Approach Description

As a reminder, the MDA is a modeling and automation approach based on metamodels. The MDA architecture distinguishes between three abstraction levels in any software application:

1. Computation-independent Model (CIM): This level captures the user requirements and high-level specifications of the application.
2. Platform-independent Model (PIM): This level represents the conceptual model of the application, independently of any specific target platform.
3. Platform-specific Model (PSM): This level represents the application's design tailored to a specific target platform.

The transition between these levels is performed automatically through the use of model-to-model transformation rules expressed using a transformation language, such as Atlas Transformation Language (ATL) [19] and Query/View/Transformation (QVT) [6]. Moreover, the MDA allows the generation of the implementation code from the PSM model through model-to-text transformation rules. The MDA offers numerous advantages, such as automation, communication and interoperability, while reducing development time and costs [20].

In this work, we employed the MDA paradigm to design a framework for data-warehouse modeling and implementation across diverse NoSQL stores, as illustrated in Figure 1. The objective of our approach is not only to automatically derive the data warehouse PSM model from the PIM model, but also to generate a model with a recommended design and configuration defined regarding developers' feedback and data stores' recommendations. Therefore, instead of choosing a predefined target model, flat, star or snowflake, our collaborative approach aims to provide the user with a suitable hybrid schema based on the defined project parameters. In this context, the transformation rules are implemented with a focus on the following aspects:

- Depending on the target platform, determining when to use normalization, denormalization and/or vertical partitioning.
- Identifying which project parameters should guide the DW modeling to obtain a tailored and efficient model.
- Determining which configuration parameters are crucial for the model's performance and that must be communicated through the transformation rules.

In fact, besides the DW structural model, our transformation rules address three main aspects: sharding, distribution and vertical partitioning. Indeed, in NoSQL systems, defining the appropriate sharding and distribution configuration is a crucial design consideration. Vertical partitioning is also an important design aspect in NoSQL, where splitting a "table" can provide better performance compared to using a single table. It's important to note that replication parameters are not included in our approach, as these are determined at the DW level, that is, for the database as a whole rather than for specific data marts. Thereby, we extended the PIM metamodel described in [21] to include essential project parameters needed to generate a tailored model of a data mart. This dynamic design enables us to formulate a transformation rule as follows:

$$S_p + \text{Dynamic Transformation rule (p)} = D_p \quad (1)$$

where S is the source model (PIM), p is a set of project's parameters and D_p is a fitted destination model (PSM) in terms of p.

However, to ensure coherent resulting models without missing or redundant elements, dynamic transformation rules must carry out the following constraints:

- Completeness: During the execution of transformation rules, all the necessary DW components and attributes are generated to avoid any omissions.
- Disjoint: During the execution of transformation rules, no DW components or attributes are generated more than once, thus avoiding redundancy.

It is essential to note that the objective of this paper is not to define specific design patterns, as these should be the outcome of numerous experiments tailored to each project's specific architecture and data characteristics.

Finally, from the obtained PSM, we generate three DW files using model-to-text transformation rules. The first one is the implementation code which in case of NoSQL databases because of their schemaless aspect [22] does not include all DW metadata. Therefore, our framework generates two additional files: data template and a schema validation file, the purpose of which is to enhance developers' comprehension of the derived model and to provide them with directives for the loading phase.

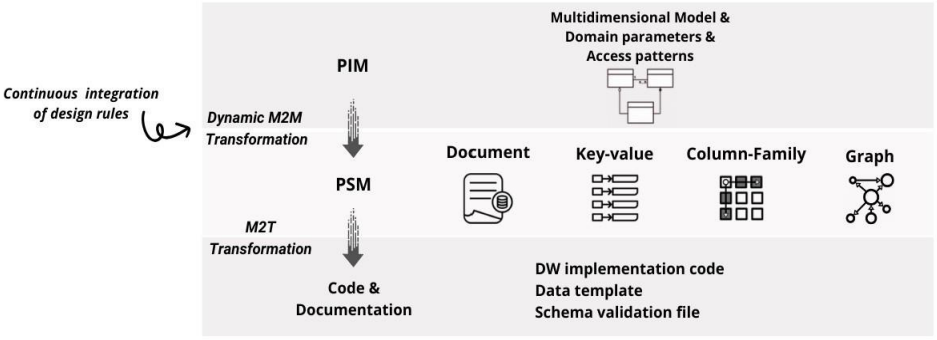


Figure 1. Our collaborative model-driven framework for NoSQL-oriented data warehouses, generated from the PIM model using collaborative dynamic transformation rules that are continuously enhanced with new design practices.

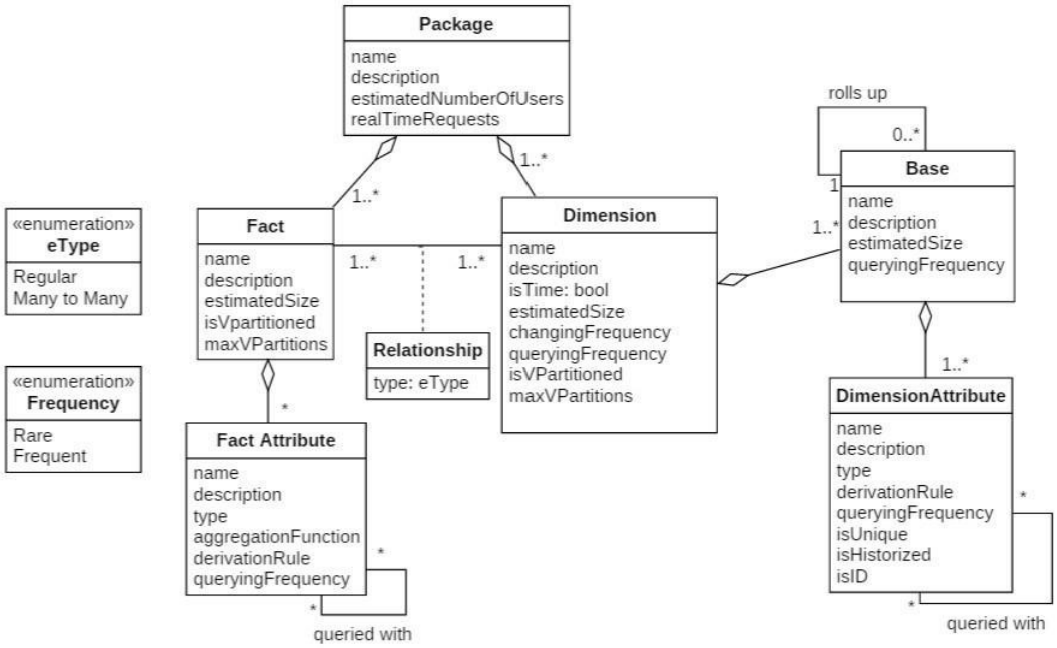


Figure 2. Our extended multidimensional metamodel (PIM) comprising domain and access parameters.

3.2 PIM Metamodel

To represent the data-warehouse conceptual model (PIM), we propose an extended multidimensional metamodel, as illustrated in Figure 2. This metamodel describes the data-warehouse structure and all aspects of the environment necessary for its design. Within this metamodel, the data warehouse is depicted as a package that includes facts and dimensions. Facts contain attributes (measures) characterized by a name, type and aggregation function. Furthermore, facts are linked to their related dimensions through relationships (either one-to-many or many-to-many relatioships). Additionally, dimensions are composed of base elements that represent hierarchy levels.

Besides the multidimensional aspects, our proposed metamodel is further extended to encompass domain parameters and access patterns, which play a crucial role in guiding data-warehouse design, especially within NoSQL platforms. This adaptation of the metamodel enables the definition of dynamic transformation rules based on these parameters. Consequently, our conceptual metamodel incorporates domain parameters related to data-size estimation, the number of users and a realTimeRequests boolean. For facts, in addition to estimated data size, attributes, such as isVpartitioned and maxPartitions are included. These allow users to specify whether facts could be vertically partitioned and to define the maximum number of partitions allowed. Furthermore, we have enhanced fact attributes by adding querying frequency attributes and the "queried with" association to capture access patterns among fact attributes. In regard to dimensions, similar to facts and fact attributes, we have incorporated attributes for estimated data size, is Vpartitioned, maxPartitions and the querying frequency of the dimension. Additionally, we introduced the changing frequency attribute for dimensions, which influences decisions related to dimension design. At the dimension attributes' level, we introduced is Historized to indicate decisions that must be made during the conception phase, impacting the attribute physical design. Attributes is Unique and is ID were also added to further define dimension characteristics. To capture access patterns between dimension attributes, the association "queried with" has been added. It is important to highlight that the environment parameters can be personalized and enriched by BI developers according to their specific needs.

4. DOCUMENT-ORIENTED DATA WAREHOUSE

4.1 Document-oriented Databases Metamodel

A document-oriented database consists of a collection of records stored in formats such as JSON, BSON, XML or YAML. Each record, referred to as a document, comprises key-value pairs Key:Value. The schema-less nature of these databases allows records within the same collection to possess different attributes, prompting us to represent a collection by a FieldSet (Figure 3). A FieldSet is defined as a group of documents that share the same fields. Each field, also known as a property, is characterized by a type (e.g. string, integer, array, list, …etc.), a description, a derivation rule and an isHistorized attribute. Additionally, a FieldSet may include other FieldSets, known as embedded documents. Collections in some document databases may have an optional identifier. These collections can also contain reference keys, similar to foreign keys in a relational model, which reference other collections using their URIs. It is possible to designate certain fields within a collection as required, ensuring their presence. Moreover, to improve model performance, parameters relating to sharding and distribution can be specified at the collection level. In our metamodel, these aspects are represented as dictionaries, allowing for the inclusion of numerous parameters as needed.

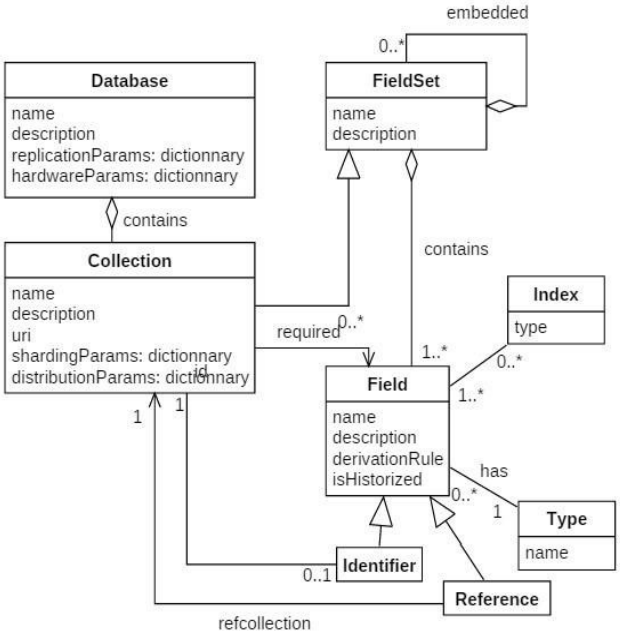


Figure 3. Our proposed document-oriented metamodel.

4.2 From PIM to Document-oriented PSM

In a document-oriented data warehouse, the transformation rules aim to determine the collections to be created for facts and dimensions within the data warehouse. They also dictate how data is distributed across these collections and define which attributes should be embedded or referenced. The configuration of these parameters is a critical task. It must be based on a combination of factors related to domain-specific requirements and access patterns, in addition to the experience and feedback of business-intelligence developers.

Basically, three fundamental models serve as a foundation for document-oriented data warehouses:

- Flat Model: In this model, a single collection is dedicated for each fact with its associated dimensions. Within this collection, all attributes of the fact are represented as fields, while the dimensions are incorporated as embedded documents.
- Star Model: In this model, an individual collection is created for each fact and for each dimension. Each dimension is then referenced within the corresponding fact collection. Hierarchies are embedded into the root or terminal dimension.
- Snowflake Model: This model employs separate collections to store facts, dimensions and hierarchy levels, joined through references.

In our approach, in addition to these classical models, we propose a hybrid model that combines embedded, referenced and vertically partitioned documents. This model is driven by project parameters defined in the PIM model through automatic and dynamic transformation rules. Below, we present possible mappings for each DW element.

- Fact F: Each fact F, defined by a set of fact attributes, can be mapped to a single collection CF with all the fact attributes represented as fields. Alternatively, it can be mapped to a set of collections, each containing a sub-set of the fact attributes and sharing the same dimensions. (Figure 4) illustrates an example of a dynamic rule for mapping a fact table to either a single collection or two collections. This decision is based on the fact table's estimated size, the querying frequency of its attributes and whether vertical partitioning of the facts is permitted by the user. It should be noted that the division of a fact into multiple collections can be executed based on more complex formulae. For example, using the "queried with" attribute allows for grouping fact attributes that are commonly queried together into the same collection.
- Regular Dimension: Each dimension *D*, defined by a set of dimension attributes, can be mapped to an embedded document within the fact collection, to a single collection C_d or to a set of collections.

```
rule Fact2Collection1{
    from f:PIM!Fact (f.estimatedSize < thisModule.FactSizeThreshold or f.isVpartitioned=false)
    to c:Doc!DocCollection (
    Name<- f.Name,
    Identifier <- thisModule.Fact2ID(f),
    shardingParams <- '(sharding,disabled)',
    contains <- c.Identifier,
    contains <- f.Attributes-> asSequence() -> collect(e|thisModule.FA2Field(e)),
    required <- c.contains,
    contains<-Doc!Reference.allInstances(),
    Embedded <-Doc!FieldSet.allInstances()->select (e|e.Description='Embedded Dimension')
    )
}
rule Fact2Collection2{
    from f:PIM!Fact (f.estimatedSize >= thisModule.FactSizeThreshold and f.isVpartitioned=true)
    to c1:Doc!DocCollection (
    Name<- f.Name,
    Identifier <- thisModule.Fact2ID(f),
    shardingParams <- '(sharding,enabled), (shardingKey, date)',
    contains <- c1.Identifier,
    contains <- f.Attributes-> asSequence() ->select (fa|fa.queryingFrequency='frequent') -> collect(e|thisModule.FA2Field(e)),
    required <- c1.contains,
    contains<-Doc!Reference.allInstances(),
    Embedded <-Doc!FieldSet.allInstances()->select (e|e.Description='Embedded Dimension')
    ),
    c2:Doc!DocCollection (
    Name<- f.Name,
    Identifier <- thisModule.Fact2ID(f),
    shardingParams <- '(sharding,enabled), (shardingKey, date)',
    contains <- c2.Identifier,
    contains <- f.Attributes-> asSequence() ->select (fa|fa.queryingFrequency='rare') -> collect(e|thisModule.FA2Field(e)),
    required <- c2.contains,
    contains<-Doc!Reference.allInstances(),
    Embedded <-Doc!FieldSet.allInstances()->select (e|e.Description='Embedded Dimension')
    )
}
```

Figure 4. ATL dynamic transformation rule for mapping a fact to collections based on its estimated size and the querying frequency of its attributes.

- Hierarchy: Each hierarchy level L within dimension D is mapped according to the mapping of D. If D is mapped as embedded, then L is also mapped as embedded. Otherwise, L can be mapped to an embedded document E_h within collection C_d or to a separate collection C_h.
- Many-to-many dimension: In traditional data warehouses, this kind of dimension is linked to the fact using a degenerate fact [21]. However, in the context of NoSQL data stores, which support a wide range of data types, the many-to-many dimensions D m2m are mapped to a field in the fact collection. This field utilizes collection data types, such as sets or lists, to store the dimension if it has only a single attribute. In cases where the dimension possesses multiple attributes, a distinct collection Cm2m is created to accommodate the dimension's attributes and a collection field is added to the fact to store references to this dimension's collection.

5. KEY VALUE-ORIENTED DATA WAREHOUSE

5.1 Key-value-oriented Databases Metamodel

Key-value stores, akin to hash tables [23]-[24], function by mapping values to specific keys. These values can range from primitive data types like integers and strings to more complex ones like lists, blobs and JSON documents. A distinguishing characteristic of key-value databases is their schema-less nature, which allows users to dynamically add or remove values (fields) at any time. This results in records that may possess different attributes within the same table.

To represent this flexible data model, we propose the following metamodel (Figure 5). The principal component in this metamodel is the table, which has a primary key, also known as a partition key and a set of attributes. Additionally, most key-value data stores offer the ability to define multiple indices on attributes, known as secondary, sort or local keys. These indices play a crucial role in enabling efficient data querying. Due to the flexibility of the key-value model, we utilize KeyValueSets to group key-value pairs that share the same attributes. Although the KeyValueSet concept is not employed during the implementation phase, it helps developers understand how data can conceptually be organized.

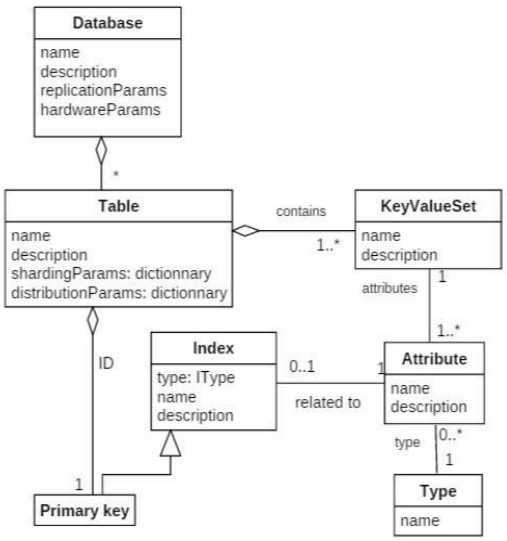


Figure 5. Our proposed key-value oriented metamodel.

5.2 From PIM to Key-value Oriented PSM

In the context of key-value oriented data warehouse and considering the lack of joins in most key-value stores, we adopt in this paper a flat model using a single table. Under this model, each fact is associated to a single table with all its corresponding dimensions. The table has a primary key formed by the concatenation of all dimensions keys, to allow and facilitate data querying. All attributes of the fact and dimensions are mapped to attributes within that single table. The same solution is also applicable for the many-to-many dimensions, but using collection data type. Additionally, an index is created for each dimension attribute required for querying data as illustrated by the (Figure 6). It is important to note that

defining indices is a design consideration requiring knowledge and expertise in the chosen database system and environment. Such expertise can be communicated through dynamic transformation rules.

```
rule Dim2KV{
    from d:PIM!Dimension
    to k:KV!KeyValueSet(
    Name<-d.Name,
    Description <- d.Description,
    composed <- d.hierarchy.contains_Att -> asSequence() -> collect(e|thisModule.DA2Attribute(e))
    )
}
lazy rule DA2Attribute{
    from da:PIM!DimensionAttribute, desc:PIM!Descriptor, oid:PIM!OID
    to f:KV!Attribute(
    Name<- da.Name,
    Type <- da.Type,
    Description <-  da.Description,
    Index <- if da.queryingFrequency='frequent' then da.thisModule.DA2Index(da) else '' endif
    )
}
lazy rule DA2Index{
    from a:PIM!DimensionAttribute
    to I: KV!Index (
    Name<-'I_'+a.Name,
    Type <-'Secondary'
    )
}
```

Figure 6. ATL dynamic transformation rule for mapping a dimension to a KeyValueSet and dimension attributes to attributes and generating an index for each frequently accessed attribute.

6. COLUMN-FAMILY ORIENTED DATA WAREHOUSE

6.1 Column-family Oriented Database Metamodel

NoSQL column or column-family stores, also referred as wide-column stores [25], closely resemble key-value stores [23], with data stored as keys mapped to values and rows or records having varying attributes. However, in column-family stores, data can be grouped into column families [24], where each column family represents a specific map of data [25]. Thus, to represent the column-family oriented data model, we propose the metamodel depicted in (Figure 7). In this metamodel, a database (also referred to as a keyspace in some data stores) is composed of tables. Each table is made up of columns that can be organized into column families translated by the 0,1 multiplicity. It is important to note that in some NoSQL column-family stores, like Cassandra, a column family is synonymous with a table. Our proposed metamodel accommodates all scenarios. Moreover, in some column-family stores, column families can be further grouped by super columns, which we represent in our metamodel with a reflexive association between column families. Additionally, in most column-family stores, each column is assigned a timestamp, which the DBMS uses to manage consistency conflicts. Each table has a row key, serving as its primary key. It is noteworthy that, unlike relational databases, NoSQL column-family stores do not support joins. Finally, tables in our metamodel include two dictionaries for configuration parameters related to sharding and distribution, which are key design considerations for such types of databases. Meanwhile, hardware and replication parameters are defined at the database level. These configuration parameters are generally determined based on the platform used, hardware capacity and expert feedback.

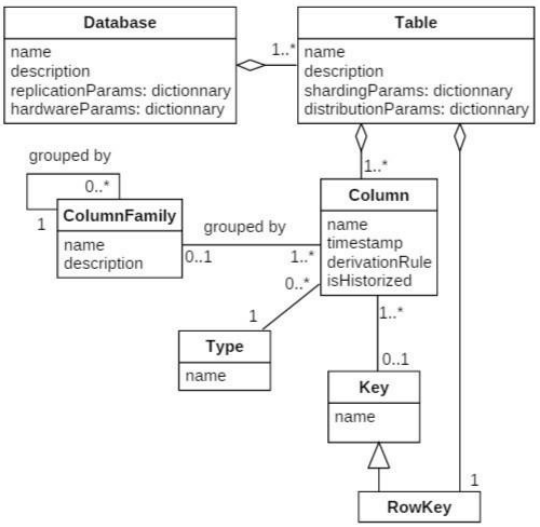


Figure 7. Our proposed column-family oriented metamodel.

6.2 From PIM to Column-Family oriented PSM

To design a column-family oriented data warehouse, similar to a key-value oriented data warehouse, the flat model is the only viable approach. This model utilizes a single table to manage both facts and dimensions, including both regular and many-to-many dimensions, with columns organized into column families.

Indeed, one crucial design consideration in column-family stores is how columns are grouped, which implicitly defines how data is stored and accessed. Therefore, the organization of columns must be carefully planned based on access patterns, as defined at the PIM level. The division of columns into column families can be guided by a straightforward transformation rule, such as using access frequency, or it can involve more complex criteria, like clustering columns into sub-groups according to the "queried with" association.

Figure 8 shows an example of a dynamic transformation rule: The rule generates a table containing all attributes of facts and dimensions, grouped into two column families, one for frequently queried attributes and the other for rarely accessed attributes. Additionally, the rule recommends two configuration parameters: the compaction strategy and the clustering key.

```
rule Fact2Table{
    from f: PIM!Fact
    to T: CL!Table (
    Name <- f.Name,
    distributionParams <- 'compaction, LeveledCompactionStrategy ',
    shardingParams <- 'clusteringkey, date',
    composed <- f.Attributes -> asSequence() -> collect(e|thisModule.FA2Column(e)),
    composed <- CL!Column.allInstances(),
    ID <- thisModule.Fact2RowKey(f),
    composed <- T.ID.attributes
    ),
    cf1:CL!ColumnFamily(
    Name <- 'FrequentlyAccessed'
),
    cf2:CL!ColumnFamily(
    Name <- 'RarelyAccessed'
)}
rule DA2Column{
    from da:PIM!DimensionAttribute
    to c:CL!Column(
    Name <- da.Name,
    Type <- da.Type,
    DerivationRule<- 'Derivation rule: ' + da.DerivationRule,
    IsHistorized <- da.isHistorized ,
    groupedBy <- if da.queryingFrequency='frequent' then CL!ColumnFamily.allInstances()-> select(cf|cf.Name='FrequentlyAccessed')->first()
                else
                CL!ColumnFamily.allInstances()->select(cf | cf.Name = 'RarelyAccessed')->first()
                endif
)}
rule FA2Column{
    from fa:PIM!Fact_Attribute
    to f:CL!Column(
    Name <- fa.Name,
    Type <- fa.Type,
    Formula<- 'Derivation rule: ' + fa.DerivationRule+ ' and Aggregation Function: ' + fa.Aggregation_Function,
    groupedBy <- if fa.queryingFrequency='frequent' then CL!ColumnFamily.allInstances()-> select(cf|cf.Name='FrequentlyAccessed')->first()
                else
                CL!ColumnFamily.allInstances()->select(cf | cf.Name = 'RarelyAccessed')->first()
                endif
)}
```

Figure 8. ATL dynamic transformation rule for converting a fact into a table with two column-families; one for frequently accessed attributes and one for rarely accessed attributes. The rule specifies configuration parameters essential for data mart table creation.

7. GRAPH-ORIENTED DATA WAREHOUSE

7.1 Graph-oriented Metamodel

Graph databases have many distinct characteristics when compared to other NoSQL databases and these characteristics can also vary when comparing different graph database platforms. In this paper, we introduce a metamodel designed to capture the essential concepts necessary for the implementation of a data warehouse in a graph-database environment. The primary components in a graph-oriented database are nodes, which store data and edges which represent relationships between nodes, as depicted in Figure 9. Each edge has two endpoints: a start and an end node. Furthermore, like nodes, edges can also possess attributes. They may be directed or undirected, a distinction represented in our metamodel by the "isDirected" attribute. The "type" attribute categorizes edges based on other characteristics, such as whether they are weighted, reflexive or composite. Additionally, attributes may include an "isUnique" property, allowing for the specification of unique identifiers at the node level, functioning like primary keys. Each attribute can also be assigned one or many indices of different types. Finally, unlike other

types of NoSQL stores, in graph-oriented databases, most configuration parameters are set at the database level rather than at the node level, as represented in our metamodel. However, we have retained a configuration dictionary at the node level for specific cases.

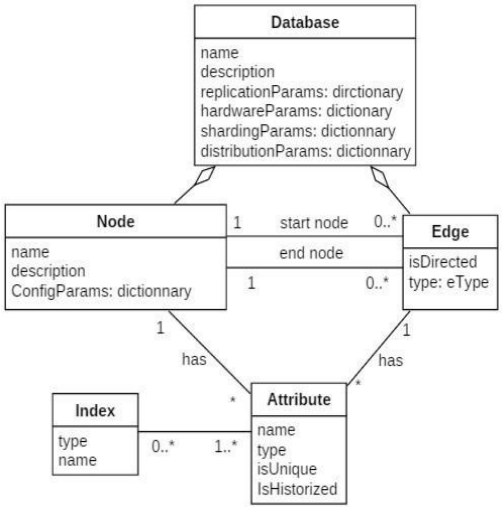


Figure 9. Graph-oriented metamodel (PSM).

```
helper  def: FactFAThreshold : Integer = 30;
rule Fact2SingleNode{
    from f: PIM!Fact (f.Attributes.count()<=thisModule.FactFAThreshold or f.isVpartitioned=false)
    to N: Graph!Node (
    Name <- f.Name,
    Description <- 'Fact Node',
    Attributes <- f.Attributes -> asSequence() -> collect(a|thisModule.FA2Attribute(a))
)
}
rule Fact2MultiNode{
    from f: PIM!Fact (f.Attributes.count()>thisModule.FactFAThreshold and f.isVpartitioned=true)
    to nodes : Sequence(Graph!Node) (
    nodes <- thisModule.SplitFA(f.Attributes)
    )
}
```

Figure 10. ATL dynamic transformation rule for mapping a fact into one or multiple nodes based on the number of fact attributes.

7.2 From PIM to Graph-oriented PSM

In designing a graph-oriented data warehouse, two primary models are typically considered; star and snowflake, while the flat model with one single node is not recommended in graph stores as it does not leverage the inherent advantages of graph databases.

- Star model: this model employs separate nodes to store facts and dimensions, joined through edges. Hierarchies are placed in the related dimension's node.
- Snowflake model: this model employs separate nodes to store facts, dimensions and hierarchies, joined through edges.

Building on these two fundamental models, our approach aims to dynamically generate a hybrid model that satisfies the DW requirements, rather than being a predetermined choice. Below, we present possible mappings for each DW component:

- Fact F: Each fact is mapped to a single node N containing all the fact attributes, which is then connected to its dimension nodes through edges. Alternatively, a fact can be mapped to several nodes that contain groups of fact attributes. This approach is used in scenarios such as a high number of fact attributes or a large size of the fact. Figure 10 illustrates an example of a dynamic transformation rule for mapping a fact based on the number of fact attributes.
- Regular Dimension *D*: A dimension can be either normalized or denormalized, depending on the transformation rule parameters. In the normalized case, each dimension is mapped to a node N_d, which is joined to the fact node *N* and contains all the dimension's attributes and hierarchies. In the denormalized case, the dimension and its hierarchies are mapped to distinct nodes. It is

important to note that, in both scenarios, the dimension nodes can be further vertically split into multiple nodes under certain conditions, such as an excessive number of attributes or if some attributes are rarely queried. Furthermore, to leverage the attributes of edges, we add two attributes to each edge to manage historical data: is Last and Modification Date. Figure 11 illustrates an example of a dynamic transformation rule for dimension mapping, driven by the changing frequency of the dimension.

- Many-to-many dimensions: These dimensions can be implemented similarly to regular dimensions, using a dedicated node connected to the fact node through edges.

```
rule Dim2Node{
    from d:PIM!Dimension (d.changingFrequency='rare' or d.changingFrequency='')
    to N: Graph!Node(
    Name<- d.Name,
    Description <- d.Description,
    Attributes <- d.Attributes -> asSequence() -> collect(a|thisModule.DA2Attribute(a))
    ),
    e:Graph!Edge (
    startNode <- Graph!Node.allInstances() ->select (n|n.Description='Fact Node'),
    endNode <- N
    )
}
rule Base2Node{
    from b:PIM!Base (b.Dimension.changingFrequency='frequent')
    to N: Graph!Node(
    Name<- b.Name,
    Description <- b.Description,
    Attributes <- b.Attributes -> asSequence() -> collect(a|thisModule.DA2Attribute(a))
    ),
    e:Graph!Edge (
    startNode <- Graph!Node.allInstances() ->select (n|n.Name=b.rolls_up_to.Name),
    endNode <- N
    )
}
lazy rule DA2Attribute{
    from da:PIM!DimensionAttribute
    to a:Graph!Attribute(
    Name<- da.Name,
    Type <- da.Type ,
    IsUnique<- da.isUnique,
    IsHistorized <- da.isHistorized
    )
}
```

Figure 11. ATL dynamic transformation rule for converting a dimension; slowly changing dimensions and their hierarchies are mapped to a single node. In contrast, frequently changing dimensions are mapped with a separate node for each hierarchy level.

8. FROM NOSQL PSM TO CODE

In addition to model-to-model transformations, the MDA offers the capability of generating text through model-to-text transformations. This feature is particularly useful when generating the implementation code for a specific platform or generating text in any desired format. In our approach, as previously mentioned, three different files are generated for each data-warehouse project:

- The data warehouse or data-mart implementation code for the target platform.
- Data-template file.
- Schema-validation file.

To ensure flexibility and clarity for the framework's users, the data-template and schema-validation files are generated in JSON format.

8.1 From PSM to Implementation Code

In NoSQL stores, due to the schemaless aspect of these databases, a predefined schema is not required. In document databases, the schema is dynamically derived from the documents inserted into the database. For other types of NoSQL stores, creating a data warehouse necessitates defining only the table names while the attributes are deduced at the data-loading time. For this reason, the implementation code generated by our framework is coupled with other files to provide developers with all metadata available in the PSM model.

In case of key-value databases, the table can be created with an initial schema. This can be performed through a JSON file in some platforms or using a programming language or a platform-specific script. To illustrate that, we generate in this work the implementation code for the DynamoDB [26] platform. In DynamoDB, only the primary key attribute is mandatory and an optional sort key can be specified.

The remaining attributes are defined during the data-loading process. Below, we present a conceptual template for creating a table in DynamoDB:

```
aws dynamodb create-table \
–table-name T-name \
–attribute-definitions \
AttributeName=PKA-name,
AttributeType=PKA-type \
AttributeName=SKA-name,
AttributeType=SKA-type \
–key-schema \
AttributeName=PKA-name,KeyType=HASH \
AttributeName=SKA-name,KeyType=RANGE \
```

In the context of a column-family data warehouses, various specific languages are available for generating the implementation code, with most of them being specific to each database system. In our framework, we generate the code to implement the data warehouse in Hbase. In this latter, only the table name and the names of its column-families are required. Below, we present a conceptual template for creating a table in Hbase:

CREATE HBASE TABLE T name,CF_1 name,CF_2 name, ..., CF_n name

Figure 12 illustrates the transformation rule used to generate the data-mart table for our case study.

```
query Hbase_File=Column!Database.allInstances()->
collect(d | d.Code_Generator().writeTo('/PIM2PSM/Hbase.sql'));

helper context Column!Database def : Code_Generator(): String=
 Column!Table.allInstances() -> select(d|not d.oclIsUndefined()) -> iterate (t;code:String=''|
code
+ 'CREATE '+'"'+ t.Name+'"'
+    Column!ColumnFamily.allInstances() -> select(c|not c.oclIsUndefined()) -> iterate (cf;columnsF:String=''|columnsF
 + ' , ' + '"'+ cf.Name +'"'
 ))
;
```

```
CREATE "StoreSales" , "FrequentlyAccessed" , "RarelyAccessed"
```

Figure 12. M2T transformation rule and the obtained implementation code to create the data-mart table in HBase for our case study.

In graph-oriented databases, many specific query languages exist, depending on the platform used, such as Gremlin, SPARQL and Cypher. In our framework, we generate a code for Neo4j, which utilizes the Cypher language. In this scenario, only the names of the nodes and the edges joining them are specified:

CREATE (f:N-name) CREATE (d_i : N_{d_i} -name)// foreach dimension
MATCH (f:N-name), (d_i : N_{d_i} name)
CREATE (f)-[r:Ed-name] > (d_i)
//foreach edge

8.2 From PSM to Schema-validation File

In data-warehousing systems, each field plays a crucial role in the analysis and visualization phases, making precise control over the DW's data essential. Such control ensures accurate calculations and prevents the occurrence of empty data in reports. However, in NoSQL data stores, being schemaless, the responsibility for managing the database schema and business rules is assigned to the application layer rather than to the database itself. Consequently, in NoSQL-oriented data-warehousing systems, schema control typically occurs during the loading phase.

In our MDA approach, the schema-validation file provides the user with all available metadata and outlines on how data should be organized, considering that the implementation code does not offer this information. This file can also be used in data-loading batches to control data structure. It serves to outline the required fields and to provide a detailed description of each field. In case of a document-oriented data warehouse, a schema-validation file is generated for each collection encompassing all its fields or embedded documents. In the case of a key-value data warehouse, all table attributes are directly listed in a single file. For a column-family data warehouse, a single schema-validation file is generated for the data mart, with column-family attributes grouped in embedded documents. In case of graph-oriented DW, a validation file is generated for each node. Figure 13 shows the schema-validation file

generated for our case study.

```
-- command to write in JSON file
query Schema_validation=Doc!DocCollection.allInstances()->
collect(d | d.Code_Generator().writeTo('/PIM2PSM/SchemaValidation.json'));

helper context Doc!DocCollection def : Code_Generator(): String=
 '{\n "$id":"/'+self.Name+' .schema.json",//schema URI \n' +
    '"title":"'+self.Name+'",\n'+
    '"description":"'+self.Description+'",\n'+
    '"type":"object",\n'+
    '"required":['+ self.required -> select(c|not c.oclIsUndefined()) ->
                   iterate (fa;required:String=''|
    required + fa.Name+
    if self.required.last()<>fa then ','
    else '],\n'
    endif
    )
+
'"Properties":{ \n'+
'//Measures Fields \n'+
self.contains -> select(c|not c.oclIsUndefined()) -> iterate (fa;contains:String=''|
    contains+
'"'+ fa.FName+'":{ \n'+
'"type":"' + fa.FType+ '", \n'+
'"description":"' + fa.FDescription +'"} \n'
    )
+self.Embedded -> select(c|not c.oclIsUndefined()) -> iterate (em;contains:String=''|
    contains+
    '//'+ em.Name+' Embedded Dimension \n'+
'"'+ em.Name +'":{'+
'type:"object",\n'+
'properties:{\n'+
em.contains -> select(c|not c.oclIsUndefined()) -> iterate (e;field:String=''|
    field+
'"'+ e.FName+'":{ \n'+
'"type":"' + e.FType+ '", \n'+
'"description":"' + e.FDescription +'"} \n'
    )
)
+' }'
;
```

```
{
 "$id":"/Store .schema.json",//schema URI
"title":"Store",
"description":"Sales measures",
"type":"object",
"required":[WholeSalesCost,ExTax,NetPaid,NetProfit,],
Properties":{
//Measures Fields
"WholeSalesCost":{
"type":"Float",
"description":"Calculates the total amount of sales"
} ,
"ExTax":{
"type":"Float",
"description":"Calculates the amount of taxes"
} ,
"NetPaid":{
"type":"Float",
"description":"Calculates the net paid by customers"
} ,
"NetProfit":{
"type":"Float",
"description":"Calculates the net profit by deducing the amount of ta
} ,
//Item Embedded Dimension
"Item":{
type:"object",
properties:{
"ItemDescription": {
type:"String",
description:"Represents the product description"},
"ProductName": {
type:"String",
description:""},
},
//Date Embedded Dimension
"Date":{
type:"object",
properties:{
"Date": {
```

Figure 13. M2T transformation rule to generate a schema-validation file.

8.3 From PSM to Data Template

To enhance developers' understanding of the data-warehouse model, we generate a data template for each destination model. This generated template is a JSON file that includes examples of automatically generated data, along with descriptions of each attribute as detailed in the PIM model and transferred to the target PSM model. Figure 14 illustrates the definition of the model-to-text transformation rule and the resulting data template in the case of a key-value PSM.

```
-- @path KV=/PIM2PSM/KeyValue_PSM.ecore
-- command to write in JSON file
query Template_File=KV!Table.allInstances()->
collect(d | d.Code_Generator().writeTo('/PIM2PSM/CompactDataTemplate.json'));
helper context KV!Table def : Code_Generator(): String=
 '{\n ' + self.PK.Name + ':"'+self.contains -> select(c|not c.oclIsUndefined()) -> iterate (k;pk:String=''|
    pk
    +'#'+ k.Name+':1'
)
+'",\n'
+self.contains -> select(c|not c.oclIsUndefined() and c.Name=self.Name) -> iterate (fact;f:String=''|
    f
    +fact.composed -> select(a|not a.oclIsUndefined()) -> iterate (fa;a:String=''|
    a+ fa.Name+':"'+fa.Type.TypeToData() +'"//'+fa.Description+
    if fact.composed.last()<>fa then ',\n'
    else '\n'
    endif
    )
)+
self.contains -> select(c|not c.oclIsUndefined() and c.Name<>self.Name) -> iterate (k;edge:String=''|
    edge
    +'ID_'+ k.Name+':"'+k.Name+'1" //'+  k.Description+
    if self.contains.last()<>k then ',\n'
    else '\n'
    endif
)
+' }\n '
+self.contains -> select(c|not c.oclIsUndefined()and c.Name<>self.Name) -> iterate (dim;d:String=''|
    d+ '{\n'+ self.PK.Name+ ':"'+dim.Name+'1",\n'
    +dim.composed -> select(a|not a.oclIsUndefined()) -> iterate (da;a:String=''|
    a+ da.Name+':"'+ dim.Name+'1 '+ da.Name+'"//'+ da.Description+
    if dim.composed.last()<>da then ',\n'
    else ' \n }\n'
    endif
    ));
helper context String def : TypeToData(): String=
if self = 'Integer' or self='Float' then '100'
    else if self='String' then 'A'
            else if self='Date' then '01/01/2000'
                else self
                endif
            endif
```

```
{
 PK:"#Item:1#Customer:1#Date:1#Store:1#StoreSales:1",
WholeSalesCost:"100"//Calculates the total amount of sales,
ExtTax:"100"//Calculates the amount of taxes,
NetPaid:"100"//Calculates the net paid by customers,
NetProfit:"100"//Calculates the net profit by deducing the amo
ID_Item:"Item1" //Represent the product description,
ID_Customer:"Customer1" //Represents the customer description,
ID_Date:"Date1",
ID_Store:"Store1" //Represents the store description,
 }
 {
PK:"Item1",
ItemDescription:"Item1 ItemDescription",
CurrentPrice:"Item1 CurrentPrice",
Category:"Item1 Category",
Color:"Item1 Color",
ProductName:"Item1 ProductName"
 }
{
PK:"Customer1",
FirstName:"Customer1 FirstName",
LastName:"Customer1 LastName",
Login:"Customer1 Login",
EmailAdresse:"Customer1 EmailAdresse"
 }
{
PK:"Date1",
Date:"Date1 Date",
Month:"Date1 Month",
Year:"Date1 Year"
 }
{
PK:"Store1",
Name:"Store1 Name",
NbEmployees:"Store1 NbEmployees",
FloorSpace:"Store1 FloorSpace"
 }
```

Figure 14. M2T transformation rule to generate a JSON data-template file.

9. CASE STUDY

To illustrate and validate our approach, we conducted experiments utilizing the benchmark database TPC- DS [27], which comprises a fact table named StoreSales and four dimensions; namely, Item, Customer, Date and Store. We used the Eclipse Modeling Framework (EMF) to create the PIM metamodel, along with our four proposed PSM metamodels; namely, document, key-value, column-family and graph. We then implemented the model-to-model and model-to-text transformation rules using the ATL language [19]. Figure 15 displays the metamodel instance and the properties of the Customer Login attribute, with the "querying frequency" set to "rare." This attribute serves as a parameter for the dynamic transformation rule related to dimensions and their attributes, as illustrated in Figure 4. The same applies to the attributes First Name, Last Name, Email Address, Current Price, Color, NbEmployees and FloorSpace. This is the reason why they have been placed in a separate column family, as explained by the rule displayed in Figure 8. The Customer dimension, defined as rarely accessed, has been placed in a separate collection in the document store, while the Date and Store dimensions were embedded within the fact collection due to their frequent access. In contrast, due to its estimated size, the Item dimension was divided: frequently queried attributes were placed in an embedded document and the remaining attributes were stored in a separate collection. In the graph scenario, the Item dimension, identified as frequently changing, was normalized by creating a dedicated node for the hierarchy-level category, as defined by the transformation rule displayed in Figure 11. In the key-value scenario, a single table was created to hold all facts and dimension attributes. These were grouped by KeyValueSet to show data organization, even though this concept is not directly applied during the implementation phase. Similarly, we have implemented model-to-text transformation rules, facilitated by the ATL writeTo function, as previously demonstrated.

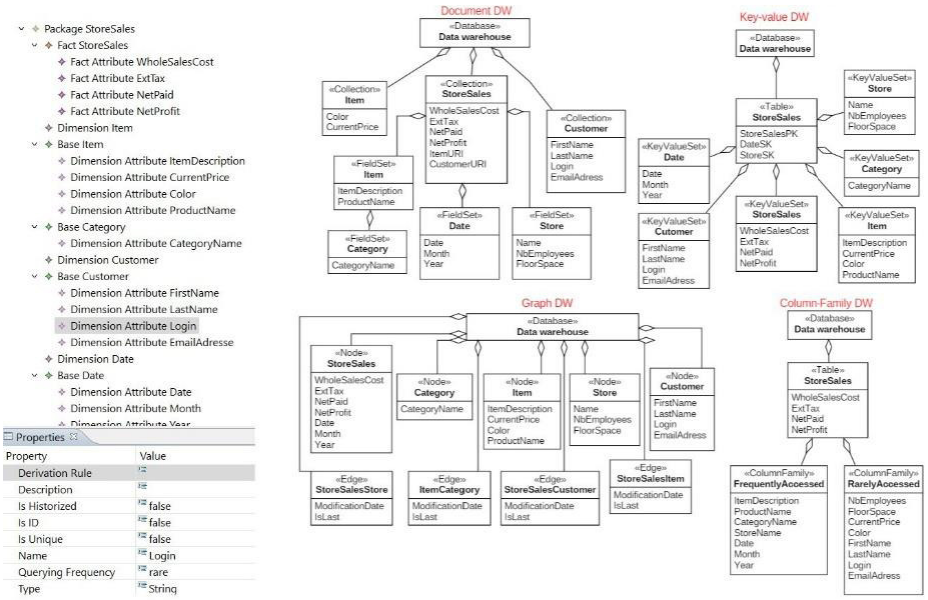


Figure 15. On the left, the PIM of our case study, showing the 'Login' attribute defined as rarely queried. On the right, the obtained physical models for the four NoSQL stores, with attributes organized according to the transformation rules previously presented.

10. CONCLUSION

In this paper, we introduced an MDA-based framework for the design and implementation of NoSQL-oriented data warehouses. We proposed a conceptual model that captures all the essential concepts needed to design a data warehouse and facilitates the transition to specific models. Additionally, we presented four metamodels to represent the logical model of a data warehouse related to document, column-family, key-value and graph data stores.

Furthermore, we proposed possible designs for a data warehouse in each type of data store. These designs are supported by dynamic transformation rules, enabling the automatic and dynamic derivation of target models. These models are tailored based on the metadata provided at the conceptual-model level, involving not just structural concepts, but also domain and access parameters. By employing

model-to-text transformations, our framework generates three critical files to implement and document the obtained model. To validate our approach, we have presented a case study demonstrating the practical implementation of dynamic transformation rules in the ATL language, showcasing the resulting models and files. Our proposal is driven by our conviction that MDA serves not only to model and automate data-warehouse creation, but also to foster a collaborative environment. This environment enables developers to consolidate their design patterns and feedback through the definition of transformation rules.

In our future work, we plan to delve deeper into each individual model, defining specialized design patterns for each targeted platform and comparing their performances. Additionally, we aim to incorporate more data platforms into our MDA framework, thereby expanding the scope and applicability of our approach.

REFERENCES

[1] W. H. Inmon and D. Linstedt, Data Architecture:A Primer for the Data Scientist, Elsevier Kaufman, 2014.

[2] C. Costa and M. Y. Santos, "Evaluating Several Design Patterns and Trends in Big Data Warehousing Systems," Proc. of the Int. Conf. on Advanced Information Systems Engineering, pp. 459-473, Springer, Cham, June 2018.

[3] F. Halper, "Modernizing the Organization to Support Data and Analytics," TDWI, Best Practices Report, [Online], Available: https://tdwi.org/research/2022/06/ppm-all-best-practices-report-modernizing-the-organization-support-data-analytics.aspx?tc=assetpg, 2022.

[4] D. Stodder, "Modernizing Data and Information Integration for Business Innovation," TDWI, [Online], Available: https://f.hubspotusercontent30.net/hubfs/6618383/Report%20-%20TDWI%20Best%20Practices%20-%20Q4-2021.pdf, Q4 2021.

[5] S. Chowdhury, [Online], Available: https://www.ibm.com/developerworks/analytics/library/baaugment-data-warehouse4/ba-augment-data-warehouse4-pdf.pdf.

[6] OMG, "MDA Guide Rev. 2.0," Object Management Group Model Driven Architecture (MDA), OMG Document ormsc/2014-06-01, [Online], Available: https://www.omg.org/cgi-bin/doc?ormsc/14-06-01, 2014.

[7] M. Chevalier, M. E. Malki, A. Kopliku, O. Teste and R. Tournier, "How Can We Implement a Multi-dimensional Data Warehouse Using NoSQL?," Proc. of the Int. Conf. on Enterprise Information Systems, LNBIP, vol. 241, pp. 108-130, Springer, Cham, April 2015.

[8] M. Chevalier, M. El Malki, A. Kopliku, O. Teste and R. Tournier, "Document-oriented Data Warehouses: Models and Extended Cuboids," Proc. of the 2016 IEEE 10th Int. Conf. on Research Challenges in Information Science (RCIS), DOI: 10.1109/RCIS.2016.7549351, Grenoble, France, 2016.

[9] M. Boussahoua, O. Boussaid and F. Bentayeb, "Logical Schema for Data Warehouse on Column-oriented NoSQL Databases," Proc. of the Int. Conf. on Database and Expert Systems Applications, LNISA, vol. 10439, pp. 247-256, Springer, Cham, August 2017.

[10] A. Sellami, A. Nabli and F. Gargouri, "Transformation of data warehouse schema to NoSQL graph data base," Proc. of the 18th Int. Conf. on Intelligent Systems Design and Applications (ISDA 2018), vol. 2, pp. 410-420, Vellore, India, December 6-8, 2018, Springer International Publishing, 2020.

[11] A. Vaisman, F. Besteiro and M. Valverde, "Modeling and Querying Star and Snowflake Warehouses Using Graph Databases," Proc. of New Trends in Databases and Information Systems: ADBIS 2019 Short Papers, Workshops BBIGAP, QAUCA, SemBDM, SIMPDA, M2P, MADEISD and Doctoral Consortium, Proceedings 23, pp. 144-152, Bled, Slovenia, Springer International Publishing, September 8–11, 2019.

[12] R. Benhissen, F. Bentayeb and O. Boussaid, "GAMM: Graph-based Agile Multidimensional Model," CEUR, [Online], Available: https://ceur-ws.org/Vol-3369/paper2.pdf, 2023.

[13] F. Kalna, A. Belangour, M. Banane and A. Erraissi, "MDA Transformation Process of a PIM Logical Decision-making from NoSQL Database to Big Data NoSQL PSM," Int. J. of Engineering and Advanced Technology, vol. 9, no. 1, pp. 4208-4215, 2019.

[14] D. Prakash, "NOSOLAP: Moving from Data Warehouse Requirements to NoSQL Databases," Proc. of the 14th Int. Conf. on Evaluation of Novel Approaches to Software Engineering, vol. 1: ENASE, pp. 452-458, DOI: 10.5220/0007748304520458, May 2019.

[15] R. Yangui, A. Nabli and F. Gargouri, "Automatic Transformation of Data Warehouse Schema to NoSQL Data Base: Comparative Study," Procedia Computer Science, vol. 96, pp. 255-264, 2016.

[16] L. Oukhouya, A. El Haddadi, B. Er-Raha and A. Sbai, "Automating Data Warehouse Design With MDA Approach Using NoSQL and Relational Systems," J. of Theoretical and Applied Information Technology, vol. 101, no. 23, pp. 7941-7957, 2023.

[17] A. Srai and F. Guerouate, "MDA Approach for Generating the PSM Model for the NoSQL Key-value Database, Application on Redis," Proc. of the 2023 3rd Int. Conf. on Innovative Research in Applied

Science, Engineering and Technology (IRASET), pp. 1-5, Mohammedia, Morocco, 2023.
[18] F. Abdelhedi, R. Jemmali and G. Zurfluh, "Relational Databases Ingestion into a NoSQL Data Warehouse," arXiv preprint, arXiv: 2203.06949, 2022.
[19] Eclipse, "ATL Documentation," [Online], Available: https://www.eclipse.org/atl/documentation.
[20] OMG, "MDA - The Architecture of Choice for a Changing World," [Online], Available: https://www.omg.org/mda/
[21] K. Letrache, O. El Beggar and M. Ramdani, "The Automatic Creation of OLAP Cube Using an MDA Approach," Software: Practice and Experience, vol. 47, no 12, pp. 1887-1903, 2017.
[22] W. Khan, T. Kumar, C. Zhang, K. Raj, A. M. Roy and B. Luo, "SQL and NoSQL Database Software Architecture Performance Analysis and Assessments: A Systematic Literature Review," Big Data and Cognitive Computing, vol. 7, no. 2, Atricle no. 97, DOI: 10.3390/bdcc7020097, 2023.
[23] P. J. Sadalage and M. Fowler, NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, 1st Edition, ISBN-10: 0321826620, Pearson Education, 2013.
[24] A. Meier and M. Kaufmann, SQL & NoSQL Databases, ISBN-10: 3658245484, Springer Fachmedien Wiesbaden, 2019.
[25] A. Vaisman and E. Zimányi, Data Warehouse Systems: Design and Implementation, 2nd Edition, ISBN-10: 3642546544, 2022.
[26] G. DeCandia et al., "Dynamo: Amazon's Highly Available Key-value Store," ACM SIGOPS Operating Systems Review, vol. 41, no. 6, pp. 205-220, 2007.
[27] TPC BENCHMARK, StandardSpecification, Version 3.2.0, pp. 1-141, [Online], Available: http://tpc.org/tpc_documents_current_versions/pdf/tpc-ds_v3.2.0.pdf, June 2021.

ملخص البحث:

في هذه الأيّام، يعدّ تحديث البيئات الخاصّة بمستودعات البيانات من التّحدّيات الأساسية في أنظمة دعْم القرارات. وهذا التّحديث ضروري لضمان قابلية الأنظمة للتّوسيع وتلبية متطلّبات عملها، لا سيّما بعد ظهور "البيانات الضّخمة". وهناك حلّ واعد لهذه المسألة يتمثّل في تنفيذ مستودعات بياناتٍ مع مخازن للبيانات على غرار (NoSQL).

في هذه الورقة، نقدّم إطارَ عملٍ معزَّزاً بنموذجٍ لتصميم وتنفيذ مستودعات حديثة للبيانات. وتهدف المنهجية المتّبعة إلى تقديم طريقةٍ تعاونيةٍ وديناميكيةٍ وقابلةٍ لإعادة الاستخدام لتطوير مستودعات بياناتٍ مخصَّصة لمشاريع ذات متطلّبات معينة. وتسهّل الطّريقة المقترحة التّوليد الأوتوماتيكي والدّيناميكي لنموذج مستودع بياناتٍ هجينٍ من نموذجه المفاهيمي، يشمل المتغيّرات البنيوية، ومتغيّرات الحقول، ومتغيّرات الوصول.

كذلك يتضمّن إطار العمل المقترح توليد الشّيفرة الخاصّة بالتّنفيذ لمستودع البيانات المقترح، إلى جانب مجموعةٍ من الملفّات للتحقُّق من النّموذج المقترح وتوثيقه ورسم مخطط مستودع البيانات على منصّة هَدَف. من جانب آخر، نقدّم دراسةَ حالةٍ مفصّلةً لتسليطِ الضّوء على فاعلية إطار العمل المقترح.

المجلة الأردنية للحاسوب وتكنولوجيا المعلومات (JJCIT) مجلة علمية عالمية متخصصة محكمة تنشر الأوراق البحثية الأصيلة عالية المستوى في جميع الجوانب والتقنيات المتعلقة بمجالات تكنولوجيا وهندسة الحاسوب والاتصالات وتكنولوجيا المعلومات.

تحتضن وتنشر جامعة الأميرة سمية للتكنولوجيا (PSUT) المجلة الأردنية للحاسوب وتكنولوجيا المعلومات، وهي تصدر بدعم من صندوق دعم البحث العلمي في الأردن. وللباحثين الحق في قراءة كامل نصوص الأوراق البحثية المنشورة في المجلة وطباعتها وتوزيعها والبحث عنها وتنزيلها وتصويرها والوصول إليها. وتسمح المجلة بالنسخ من الأوراق المنشورة، لكن مع الإشارة إلى المصدر.

الأهداف والمجال

تهدف المجلة الأردنية للحاسوب وتكنولوجيا المعلومات (JJCIT) إلى نشر آخر التطورات في شكل أوراق بحثية أصيلة وبحوث مراجعة في جميع المجالات المتعلقة بالاتصالات وهندسة الحاسوب وتكنولوجيا المعلومات وجعلها متاحة للباحثين في شتى أرجاء العالم. وتركز المجلة على موضوعات تشمل على سبيل المثال لا الحصر: هندسة الحاسوب وشبكات الاتصالات وعلوم الحاسوب ونظم المعلومات وتكنولوجيا المعلومات وتطبيقاتها.

الفهرسة

المجلة الأردنية للحاسوب وتكنولوجيا المعلومات مفهرسة في كل من:

فريق دعم هيئة التحرير

عنوان المجلة

الموقع الإلكتروني: www.jjcit.org

البريد الإلكتروني: jjcit@psut.edu.jo

العنوان: جامعة الاميرة سمية للتكنولوجيا، شارع خليل السالم، الجبيهة، عمان، الأردن.

صندوق بريد: 1438 عمان 11941 الأردن

هاتف: +962-6-5359949

فاكس: +962-6-7295534

جامعة الأميرة سميّة للتكنولوجيا
Princess Sumaya University for Technology

صندوق دعم البحث العلمي والابتكار
Scientific Research and Innovation Support Fund

المجلة الأردنية للحاسوب وتكنولوجيا المعلومات

ISSN 2415 - 1076 (Online)
ISSN 2413 - 9351 (Print)

حزيران ٢٠٢٤ المجلد ١٠ العدد ٢

JJCIT

الصفحات	عنوان البحث

www.jjcit.org

jjcit@psut.edu.jo

مجلة علمية عالمية متخصصة تصدر
بدعم من صندوق دعم البحث العلمي والابتكار