International Journal of Multidisciplinary Studies and Innovative Technologies

IJMSIT

Year : 2024 | Volume : 8 | Issue : 2

SETECH
science-engineering-technology

e-ISSN
2602-4888

International Journal of Multidisciplinary Studies and Innovative Technologies

Year :2024 **Volume : 8** **Issue : 2**

International Journal of Multidisciplinary Studies and Innovative Technologies is an online, open access, double-blind, peer-reviewed, international research journal. The language of the journal is English and Turkish. Authors should only submit original work that has not been published and is not currently considered for publication elsewhere.

An expert editor is assigned to a submitted manuscript. The editor appoints reviewers to evaluate the manuscript. As a result of the evaluation of the manuscript by reviewers, the editor decides about the acceptance, modification, or rejection of the manuscript.

International Journal of Multidisciplinary Studies and Innovative Technologies

Year: 2024, Volume: 8, Issue: 2

CONTENTS

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 52-58
DOI: 10.36287/ijmsit.8.2.1
Received: July 3, 2024; Accepted: August 30, 2024

Comparison of Different Optimization Algorithms in the Fashion MNIST Dataset

Umut Saray [1*], Uğur Çavdar [2]

[1*] *Department of Electronic Automation, Turhal Vocational School, Tokat Gaziosmanpaşa University, Tokat, Türkiye,(umutsaray@gmail.com) (ORCID: 0000-0003-3339-6876)*
[2] *Department of Mechanical Engineering, Faculty of Engineering, Izmir Democracy University, İzmir, Türkiye,(ugur.cavdar@idu.edu.tr) (ORCID:0000-0002-3434-6670)*

Abstract – This study examines the effects of various optimization algorithms used in deep learning models to classify fashion-oriented clothing items. The Fashion MNIST dataset has been chosen as a rich data source. Models developed using Convolutional Neural Networks (CNN) have been trained with various optimization algorithms such as Nadam, Adadelta, Adamax, Adam, Adagrad, SGD, and RMSprop. Understanding the impact of these algorithms on the model's performance during the training process forms the basis of the study. The findings of the research reveal that optimization algorithms have a significant effect on the accuracy rates of the model. While the Nadam and Adadelta algorithms achieved the highest accuracy rates, the RMSprop algorithm displayed relatively lower performance. These results indicate that different optimization techniques can significantly influence the performance of deep learning-based classification systems.

Keywords *– Convolutional Neural Networks (CNN),Fashion MNIST, Optimization Algorithms, Adam, RMSprop*

Citation: Saray, U., Çavdar, U. (2024). Comparison of Different Optimization Algorithms in the Fashion MNIST Dataset. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 52-58.

I. INTRODUCTION

Deep learning has become one of the most attractive areas in artificial intelligence and machine learning over the past decade, undergoing significant evolution. This method offers the ability to learn from comprehensive and voluminous datasets through models composed of multi-layered neural networks. Particularly, revolutionary results have been achieved in fields such as visual and auditory recognition, forecasting apps, natural language processing, and various pattern recognitions [1],[2]-[5].

Optimization algorithms play an indispensable role in the training process of these models [6]. The success of a deep learning model is largely dependent on the effectiveness of the chosen optimization algorithm. Stochastic Gradient Descent (SGD) [7-8] and its variants enable the model to demonstrate superior performance on the dataset by adjusting its weights and bias values. In literature, comparing different optimization algorithms and the identification of the most suitable one have become important research topics, especially for models operating on large and complex datasets [9].

Recent studies on deep learning and optimization algorithms have examined the performance of various algorithms on different datasets. For instance, Ö. Dolma [1] classified COVID-19 and non-COVID-19 lung CT scan images using deep convolutional neural networks. E. Avuçlu [2-3] evaluated the classification performance of COVID-19 images using deep learning methods. In another study, M. C. Bingöl and G. Bilgin [4] investigated the prediction of chicken diseases using transfer learning methods. Comparing optimization algorithms, especially for large and complex datasets, is crucial to determining which algorithm is more suitable. Stochastic Gradient Descent (SGD) and its variants enable the model to adjust its weights and biases to perform optimally on the dataset [6], [7-8]. Algorithms with adaptive learning rates, such as Adam [9], [10-11], Nadam [12], [13-14], RMSprop [15], [16-17], and Adagrad [18], [19-20], are widely used to achieve strong results in the training process of deep learning models. In this context, the Fashion MNIST dataset is frequently preferred as a rich data source for classifying fashion-oriented clothing items. For example, R. Sirisha and colleagues [23] compared the performance of different optimization algorithms on the Fashion MNIST dataset; A. S. Henrique and his team [24] developed CNN models using this dataset. Khanday and colleagues [25] examined the effect of filter sizes on classification accuracy. Other studies include those by Tang et al. [26], Kayed et al. [27], Zhu et al. [28], and Hur et al. [29], who have all utilized the Fashion MNIST dataset for various purposes, such as optimizing deep residual networks, using CNN LeNet-5 architecture, space-efficient optical computing, and quantum convolutional neural networks, respectively. These studies examined the impact of different optimization algorithms on the accuracy rates of deep learning-based classification models and identified the most effective algorithms [21], [22-29].

Researchers and practitioners have closely examined the algorithms used in optimizing deep learning models in recent years. Among these algorithms, methods with adaptive learning rates such as Adam, Nadam, RMSprop, and Adagrad have become popular for achieving strong results in the training process of deep learning models [21].

This study aims to examine the effects of these algorithms on the Fashion MNIST dataset [22], [23-29], a widely used dataset for training and testing contemporary artificial intelligence and machine learning systems. This dataset contains grayscale images of various clothing items and offers an excellent test ground for algorithmic classification [30].

The purpose of this research is to understand the impact of different optimization algorithms on the accuracy rates of deep learning-based classification models and to use this knowledge to enhance the effectiveness of classification systems. The findings highlight the importance of selecting optimization strategies in AI applications and guide future research in this direction.

II. Materials and Method

Convolutional Neural Networks (CNNs)[31] are frequently utilized in image processing and visual recognition tasks. Essentially, they employ convolutional layers to detect local features in an image, such as edges, textures, and shapes. These layers, through a specific learning process, automatically learn to extract useful features from different parts of the image. CNNs are capable of recognizing complex visual patterns by combining and interpreting these features in subsequent layers.

The fundamental components of CNNs include convolutional layers, activation functions, pooling layers, and fully connected layers. Convolutional layers apply filters to the input image to create feature maps, effectively extracting information from different sections of the image to identify important features.

Activation functions enhance the network's non-linear learning capability. One of the most commonly used activation functions is ReLU, which speeds up the model's training process by setting negative values to zero and helping to address the gradient vanishing problem.

Pooling layers reduce the dimensionality of feature maps, lightening the network's computational load. This is achieved by taking the maximum or average value of certain sections of the image. Pooling ensures the network's robustness against translational variances, such as changes in the position of an object within the image [32].

Fully connected layers are located at the end of the network and use the learned features to perform tasks such as classification or regression, producing the final output. These layers associate each input with probabilities for each class in the output.

Due to their ability to successfully recognize complex visual patterns, CNNs are effectively used in various application areas such as face recognition, vehicle license plate recognition, medical image analysis, and object detection from satellite images. Recent advancements in deep learning have further improved the performance and applicability of CNNs, making them an indispensable component of artificial intelligence applications [33].

In this study, Convolutional Neural Networks (CNNs) were used. Various optimization algorithms within the CNN have been compared for their success rates on the MNIST dataset. The optimization algorithms used are explained in sequence. The algorithms employed include Stochastic Gradient Descent (SGD), Adagrad, RMSprop, Adadelta, Adamax, Nadam, and Adam.

A. Stochastic Gradient Descent (SGD)

Stochastic Gradient Descent (SGD) is a method that calculates the gradient using a single training example at each step to update the model parameters. This approach enables quick parameter updates based on randomly selected samples, eliminating the need to process the entire dataset in each iteration. This efficiency makes SGD particularly effective for large datasets. However, the optimization path of SGD can be somewhat erratic, leading towards the target through a fluctuating route, which necessitates precise hyperparameter tuning for optimal performance.

The core of SGD's methodology is encapsulated in its update rule, where the parameter θ at any given iteration t+1 is adjusted according to the formula:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L(\theta_t; x_i, y_i) \quad (1)$$

In this equation, θ_t represents the parameter vector at iteration t, η denotes the learning rate, and $\nabla_\theta L(\theta_t; x_i, y_i)$ signifies the gradient of the loss function L with respect to θ, evaluated for the i th training example at the t th iteration. This process underscores the iterative nature of SGD, where each step is calculated to steer the parameters closer to the optimum by leveraging the gradient information from a single, randomly selected training example [7-8].

B. Adagrad

Adagrad is an optimization algorithm that adaptively adjusts the learning rate for each parameter, making it particularly well-suited for dealing with sparse datasets. Unlike conventional methods that use a single learning rate for all parameters throughout the training process, Adagrad modifies the learning rate individually for each parameter based on the historical gradient information. This approach lowers the learning rate for parameters corresponding to frequently occurring features, while ensuring a higher learning rate for rare features. As a result, Adagrad can significantly improve the efficiency of model training, especially in scenarios where the data is sparse.The key to Adagrad's adaptive learning rate adjustment lies in its update rule, which is mathematically formulated as follows:

For each parameter θ_t, the update at iteration t is given by;

$$\theta_{i,t+1} = \theta_{i,t} - \frac{\eta}{\sqrt{G_{i,t}+\epsilon}} \cdot g_{i,t} \quad (2)$$

Here $g_{i,t}$, represents the gradient of the loss with respect to the parameter θ_i at iteration t, $G_{i,t}$ is the sum of the squares of the past gradients with respect to θ_i up to time t, η is a global learning rate, and ϵ is a smoothing term added to improve numerical stability (often set to a small constant like 1^{e-8}), preventing division by zero.

This formula ensures that parameters with large gradients have their learning rate decreased over time, which helps in honing in on the minimum more efficiently. However, a notable downside of Adagrad is its tendency for the learning rate to decrease continually throughout training, potentially leading to premature convergence and the model stopping early in long training processes. Despite this limitation, Adagrad's ability to adapt the learning rate to the parameters has made it a foundational algorithm for further developments in adaptive learning rate techniques[18], [34-35].

C. RMSprop

RMSprop, short for Root Mean Square Propagation, is an optimization algorithm designed to overcome the challenge of the excessively decreasing learning rate that Adagrad faces. By focusing on the magnitude of gradients in only the most recent iterations, RMSprop dynamically adjusts the learning rate. This method ensures that the learning rate does not diminish too quickly, maintaining a level that is conducive to continued learning and optimization over time. RMSprop is particularly effective in scenarios involving recurrent neural networks and non-stationary targets, where the landscape of the optimization problem changes over time.

The mathematical foundation of RMSprop is expressed through its update rule, which modifies the learning rate for each parameter based on the recent gradients. The update for a parameter θ at iteration t is given by:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{v_t + \epsilon}} . g_t \quad (3)$$

In this equation, g_t is the gradient of the loss with respect to the parameter θ at iteration t, η is the initial learning rate, and ϵ is a small constant (like 1^{e-8}) to prevent division by zero. The term v_t represents the exponentially weighted moving average of the squares of the gradients, calculated as:

$$v_t = \beta v_{t-1} + (1 + \beta)\, g_t^2 \quad (4)$$

Here, β is a decay rate that determines the extent to which the moving average considers the most recent gradient magnitudes, typically set to a value like 0.9. This mechanism of adjusting $v_t$ensures that RMSprop considers the magnitude of recent gradients, enabling adaptive learning rates that respond to the current state of the optimization process.

By employing this strategy, RMSprop effectively prevents the learning rate from dropping too low, a significant improvement over Adagrad's approach. This adaptability makes RMSprop a robust choice for training deep neural networks, particularly in the challenging environments presented by recurrent neural networks and tasks with non-stationary objectives. [36],[18].

D. Adadelta

Adadelta is an optimization algorithm that extends the principles of RMSprop to enhance stability in the learning rate throughout the training process. It achieves this by employing a unit measure for weight updates, which allows for continuous model improvement without the explicit need to adjust the learning rate manually. This approach addresses one of the key challenges in optimization algorithms - the sensitivity to the choice of learning rate. By eliminating the dependence on a global learning rate, Adadelta simplifies the optimization process, making it more robust and easier to use, especially in environments where parameter tuning can be laborious.

The foundation of Adadelta is grounded in the modification of the RMSprop update rule, incorporating the use of the moving average of squared gradients to adjust the learning rate dynamically, but it also introduces the concept of accumulating updates over time to determine the step size. The update rule in Adadelta for a parameter θ at iteration t can be expressed as follows:

$$\Delta\theta_{t+1} = -\frac{\sqrt{\sum_{i=1}^{t-1} \Delta\theta_{i-1}^2 + \epsilon}}{\sqrt{E[g^2]_t + \epsilon}} . g_t \quad (5)$$

Here, g_t represents the gradient of the loss with respect to the parameter θ at iteration *t*, $E[g^2]_t$ is the exponentially decaying average of squared gradients up to time *t*, and ϵ is a small constant (similar to RMSprop) added for numerical stability. The term $\Delta\theta_t$denotes the change in θ at iteration t, and the numerator $\sqrt{\sum_{i=1}^{t-1} \Delta\theta_{i-1}^2 + \epsilon}$ represents the root mean square of previous parameter updates, which serves to scale the gradient in proportion to the historical update magnitudes.

The key innovation of Adadelta is that it does not require an explicit learning rate. Instead, it adapts the parameter updates based on the moving averages of the squared gradients and the squared updates, thus regulating the step size based on the history of changes. This self-adjusting mechanism ensures more stable and consistent learning progress, mitigating the risk of drastic updates that could potentially derail the optimization process.

By combining the adaptive gradient approach of RMSprop with the innovative update adjustment mechanism, Adadelta offers a sophisticated solution to the challenge of learning rate selection and stability, making it an attractive choice for training deep neural networks where tuning hyperparameters can be particularly challenging [36-37].

E. Adamax

Adamax is a variation of the Adam optimization algorithm, designed to enhance stability in scenarios characterized by extreme gradient values. While Adam employs adaptive moment estimation to adjust learning rates based on the first and second moments of gradients (mean and uncentered variance), Adamax introduces an alternative approach by utilizing a different norm, making it potentially more robust in the face of extreme updates. This characteristic of Adamax stems from its adaptation of the ∞-norm, which provides a theoretical upper bound on the updates, hence its name. The update rules for Adamax at iteration t for a parameter θ can be summarized as follows:

Update the first moment (the mean) of the gradient:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\, g_t \quad (6)$$

where $g_t$is the gradient of the loss with respect to θ at iteration t, and m_t is the first moment vector.

Update the ∞-norm of the gradients rather than the second moment:

$$u_t = \max(\beta_2\, u_{t-1}, |g_t|) \quad (7)$$

Here, u_t represents the ∞-norm of the gradients, which is updated to be the maximum of the previous ∞-norm scaled by β_2 and the absolute value of the current gradient. This replaces the second moment estimation used in the original Adam.

Compute the parameter update using the first moment and the ∞-norm:

$$\theta_{t+1} = \theta_t - \frac{\eta}{u_t + \epsilon}\, m_t \quad (8)$$

In this equation, η is the step size (learning rate), and ϵ is a small constant added for numerical stability.The introduction of the ∞-norm in Adamax, as opposed to the squared gradients norm in Adam, aims to provide a more stable and less aggressive adaptation of the learning rates, especially in the presence of large gradients. This makes Adamax an appealing alternative for optimization in machine learning tasks where gradients can vary significantly in magnitude, potentially

leading to more consistent and reliable convergence over the course of training [38-39].

F. Nadam

Nadam, short for Nesterov-accelerated Adaptive Moment Estimation, merges the Adam optimization algorithm with Nesterov momentum, harnessing the strengths of both methodologies to achieve more efficient optimization. By integrating Adam's adaptive learning rate features with the anticipatory updates of Nesterov momentum, Nadam facilitates faster convergence and improved performance, particularly in the context of deep learning and complex optimization tasks. This combination allows Nadam to navigate the optimization landscape more effectively, making it a powerful tool for training neural networks.

The mathematical formulation of Nadam incorporates elements from both Adam and Nesterov momentum, resulting in an update rule that looks as follows:

Update the first moment (mean) and the second moment (uncentered variance) of the gradients, similar to Adam:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \ , v_t = \beta v_{t-1} + (1 + \beta)\, g_t^2 \quad (9)$$

where g_t is the gradient of the loss with respect to the parameter θ at iteration *t*, m_t is the first moment vector, and v_t is the second moment vector.

Incorporate Nesterov momentum into the moment update by adjusting the first moment before the parameter update:

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} + \frac{(1-\beta_1) g_t}{(1-\beta_1^t)(1-\beta_1)} \quad (10)$$

$$\hat{v}_t = \frac{v_t}{(1-\beta_2^t)} \quad (11)$$

Compute the parameter update using the adjusted first moment and the second moment:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \quad (12)$$

In this equation, η is the learning rate, and ϵ is a small constant added for numerical stability.

By leveraging the lookahead nature of Nesterov momentum, which essentially incorporates information about the future gradient, Nadam ensures that each update is more informed and precise. This results in a more aggressive and effective approach to finding the minimum of the loss function, reducing the number of iterations needed to achieve convergence. Nadam's unique blend of Adam's adaptiveness and Nesterov's accelerated updates provides a significant advantage in training deep learning models, offering a balance between speed and accuracy in the optimization process [38-39].

G. Adam

The Adam optimization algorithm, standing for Adaptive Moment Estimation, is widely recognized for its ability to adaptively adjust both the learning rate and momentum for each parameter, making it a popular and effective method for deep learning tasks. By calculating exponential moving averages of both the gradients and the squared gradients, Adam maintains separate learning rates for each parameter, which are adjusted as learning progresses. This adaptability allows Adam to perform well across a wide range of deep-learning tasks, from simple to complex models. The mathematical formulation of Adam involves several key steps, as outlined below:

First and Second Moment Estimation: For each parameter θ, Adam computes the first moment (the mean) and the second moment (the uncentered variance) of the gradients:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \ , v_t = \beta v_{t-1} + (1 + \beta_2)\, g_t^2 \quad (13)$$

Here, g_t represents the gradient of the loss with respect to θ at iteration t, m_t and v_t are the estimates of the first and second moments respectively, and β_1 and β_2 are the decay rates for these moments.

Bias Correction: To counteract the biases introduced by initializing the moments as zeros, Adam applies bias corrections:

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \ , \ \hat{v}_t = \frac{v_t}{(1-\beta_2^t)} \quad (14)$$

This step ensures that the moment estimates are unbiased towards zero at the start of optimization.

Parameter Update: The parameters are updated using the bias-corrected moments:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \quad (15)$$

In this equation, η is the step size (learning rate), and ϵ is a small constant added for numerical stability. Adam's approach to adjusting the learning rate based on the first and second moments of the gradients allows for more effective and efficient optimization, especially in the context of deep learning. This adaptability, combined with its straightforward implementation and robust performance across various tasks, has cemented Adam's status as a go-to optimization algorithm for many deep learning practitioners.[9],[40].

III.IMPLEMENTATION

Dataset and Preprocessing; The Fashion MNIST dataset consists of 60,000 training and 10,000 test images. Each image is a grayscale image of a garment with a resolution of 28x28 pixels.

In terms of data preprocessing steps, the images have been normalized between 0 and 1 and reshaped to the appropriate input size for the model (28x28x1), facilitating more effective learning by the model.

Model Architecture; First Convolutional Layer: Equipped with 32 filters, each having a kernel size of (3,3), and utilizes the ReLU activation function. This layer accepts input data of 28x28 pixel resolution with 1 color channel (grayscale images).

First Max Pooling Layer: Has a pool size of 2x2, aiming to halve the spatial dimensions.

Second Convolutional Layer: Contains 64 filters, also using the ReLU activation function, with each filter having a kernel size of (3,3).

Second Max Pooling Layer: Applies a 2x2 pooling operation again to further reduce spatial dimensions.

Flatten Layer: Transforms the outputs from the convolutional and pooling layers into a single, long feature vector.

First Dense Layer: Comprises 128 neurons and employs the ReLU activation function.

Output Dense Layer: Contains 10 neurons corresponding to the ten different garment classes in the dataset, using the

softmax activation function to output probability distributions for classification.

This study aims to compare the performance of different optimization algorithms: SGD, Adagrad, RMSprop, Adadelta, Adamax, Nadam, and Adam. Each algorithm is evaluated separately using the same model architecture.

The models are trained over twenty epochs, and the accuracy rates of each optimization algorithm are assessed on the test set. Performance evaluations are conducted using loss and accuracy metrics on the test set. This methodology allows the research to be conducted within a concrete and reproducible framework, contributing to the reliability of the results obtained.

IV. Results

This research has yielded significant findings by examining the impact of different optimization algorithms on the Fashion MNIST dataset. The study reveals that optimization algorithms have substantial effects on the training process and accuracy rates of the model.

The graphs depicted in the figures demonstrate the variations in loss and accuracy values during the training process for different optimization algorithms. Figure 1 presents the comparative training and validation loss across training epochs, while Figure 2 illustrates the training and validation accuracy throughout the training iterations for various optimization algorithms on the Fashion MNIST dataset.

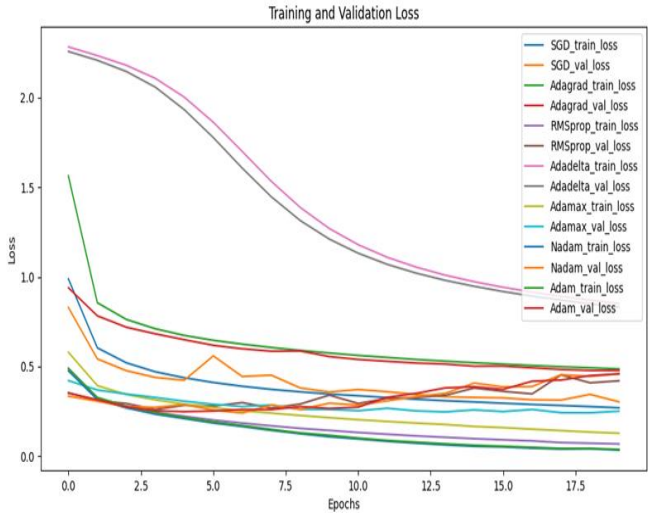


Figure 1. Optimization Algorithm Comparison: Training and Validation Loss Across Epochs for Fashion MNIST Dataset

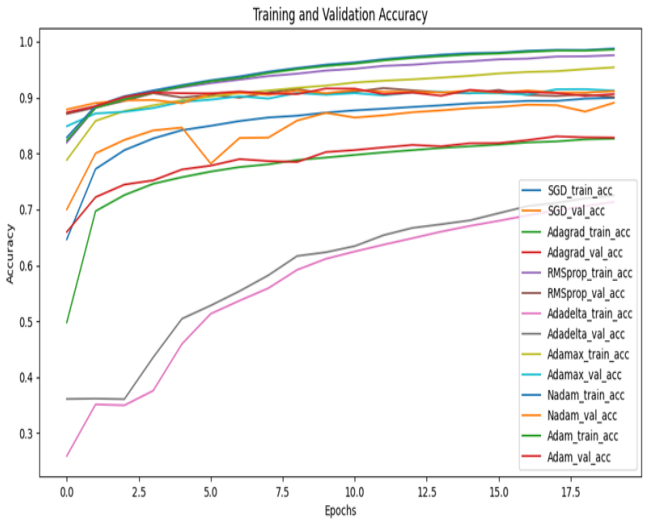


Figure 2. Comparative Analysis of Training and Validation Accuracy Over Epochs Among Various Optimization Algorithms on Fashion MNIST Dataset

Overall, all algorithms manage to reduce training loss, but some are notably more effective in reducing validation loss. The Nadam and Adam algorithms, in particular, maintain low and stable validation losses, indicating their strong generalization capabilities. On the other hand, the Adadelta algorithm, despite rapidly decreasing high initial loss values, appears to be less effective than other algorithms. The validation losses for SGD and Adagrad are also observed to be higher, which could suggest inferior generalization performance.

As seen in Figure 2, Nadam and Adam algorithms also display high validation accuracy, signifying robust performance. This high accuracy suggests that the model is well-generalized to new data. In contrast, SGD and Adagrad have lower validation accuracy, hinting that they may be less effective for training the model on this dataset. RMSprop, although exhibiting high training accuracy, has a validation accuracy that falls short of expectations, a possible indication of overfitting.

These insights reveal that the efficacy of an algorithm can vary significantly based on the dataset and specific problem at hand. They also highlight the critical role of model selection and hyperparameter tuning in machine learning. When selecting the best model, the performance on the validation set should be carefully considered.

Table.1. Performance Metrics of Different Optimization Algorithms on Fashion MNIST Dataset

Optimization Algorithm	Final Training Accuracy	Final Training Loss	Final Validation Accuracy	Final Validation Loss
SGD	0.9019	0.2651	0.8765	0.3398
Adagrad	0.8325	0.4681	0.8352	0.4603
RMSprop	0.9751	0.0678	0.8998	0.5287
Adadelta	0.7214	0.8193	0.7305	0.8004
Adamax	0.9557	0.1269	0.9115	0.2569
Nadam	0.9876	0.0331	0.9040	0.4831
Adam	0.9839	0.0436	0.9098	0.4788

According to Table 1; SGD has shown lower training accuracy and higher training loss compared to other algorithms. These findings suggest that SGD is less applicable to this particular problem than other alternatives.

Adagrad Optimization Algorithm: Adagrad has shown a moderate performance similar to SGD. Its accuracy and loss values were found to be at an average level. While Adagrad's dynamic adjustment of the learning rate can be advantageous in some problems, it has not achieved the best performance in this study.

RMSprop Optimization Algorithm: The RMSprop algorithm achieved high training accuracy and low training loss but maintained a high validation loss. This could indicate that the model did not generalize well to the validation data and might be an indication of overfitting.

Adadelta Optimization Algorithm: Adadelta's performance was lower compared to other algorithms. Both its training and validation accuracy, as well as loss values, were found to be high, indicating that Adadelta's generalization capability and training performance are lower than other alternatives.

Adamax Optimization Algorithm: Adamax exhibited good performance with high training and validation accuracy. Its low validation loss indicates that the model generalizes well to new data.

Nadam Optimization Algorithm: Nadam overall showed the best performance with the highest validation accuracy and the lowest validation loss. This indicates that the model generalizes very well to new data and that this optimization algorithm is effective for the chosen dataset.

Adam Optimization Algorithm: Adam showed a performance very close to Nadam. Its high training and validation accuracy and low validation loss indicate that this algorithm generalizes effectively.

V. DISCUSSION

This study evaluates the effectiveness of various optimization algorithms on the Fashion MNIST dataset, revealing significant differences in overall model performance. Specifically, the Nadam and Adam algorithms demonstrate superior generalization capabilities with their low validation losses and high validation accuracies, indicating their resilience against overfitting due to adaptive learning rates. On the other hand, the weaker performance exhibited by algorithms such as SGD and Adagrad, particularly in terms of high training and validation losses, highlights their limitations in developing effective learning strategies for high-dimensional datasets. These findings emphasize the critical importance of selecting and tuning optimization algorithms in machine learning projects and underscore the significance of further research to improve these algorithms. Additionally, a better understanding of performance variations among algorithms can enhance the applicability of models across broader datasets and ensure more successful implementations in practical applications.

VI. CONCLUSION

This study has demonstrated that the most effective optimization algorithms for the Fashion MNIST dataset are Nadam and Adam. The performance of other alternatives was found to be lower compared to these two algorithms. However, since the performance of each algorithm can vary depending on the dataset and the problem, conducting more comprehensive tests such as cross-validation is recommended to select the most suitable algorithm.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics

REFERENCES

[1] Ö. Dolma, "COVID-19 and Non-COVID-19 Classification from Lung CT-Scan Images Using Deep Convolutional Neural Networks," *Int. J. Multidiscip. Stud. Innov. Technol.*, vol. 7, no. 2, p. 53, 2023, doi: 10.36287/ijmsit.7.2.3.

[2] E. Avuçlu, "Examining The Effect of Pre-processed Covid-19 Images On Classification Performance Using Deep Learning Method," *Int. Sci. Vocat. Stud. J.*, vol. 7, no. 2, pp. 94–102, Dec. 2023, doi: 10.47897/bilmes.1359954.

[3] E. Avuçlu, "Classification of Pistachio Images Using VGG16 and VGG19 Deep Learning Models," *Int. Sci. Vocat. Stud. J.*, vol. 7, no. 2, pp. 79–86, Dec. 2023, doi: 10.47897/bilmes.1328313.

[4] M. C. Bıngol and G. Bilgin, "Prediction of Chicken Diseases by Transfer Learning Method," *Int. Sci. Vocat. Stud. J.*, vol. 7, no. 2, pp. 170–175, Dec. 2023, doi: 10.47897/bilmes.1396890.

[5] Y. Durgun, "Classification of Starch Adulteration in Milk Using Spectroscopic Data and Machine Learning," *Int. J. Eng. Res. Dev*., vol. 16, no. 1, pp. 221-226, 2024, doi: 10.29137/umagd.1379171.

[6] A. Williams, N. Walton, A. Maryanski, S. Bogetic, W. Hines, and V. Sobes, "Stochastic gradient descent for optimization for nuclear systems," *Sci. Rep.*, vol. 13, no. 1, p. 8474, May 2023, doi: 10.1038/s41598-023-32112-7.

[7] S. Nagendram *et al.*, "Stochastic gradient descent optimisation for convolutional neural network for medical image segmentation," *Open Life Sci.*, vol. 18, no. 1, Aug. 2023, doi: 10.1515/biol-2022-0665.

[8] C. Song, A. Pons, and K. Yen, "AG-SGD: Angle-Based Stochastic Gradient Descent," *IEEE Access*, vol. 9, pp. 23007–23024, 2021, doi: 10.1109/ACCESS.2021.3055993.

[9] C. Milovic *et al.*, "Comparison of parameter optimization methods for quantitative susceptibility mapping," *Magn. Reson. Med.*, vol. 85, no. 1, pp. 480–494, Jan. 2021, doi: 10.1002/mrm.28435.

[10] M. Reyad, A. M. Sarhan, and M. Arafa, "A modified Adam algorithm for deep neural network optimization," *Neural Comput. Appl.*, vol. 35, no. 23, pp. 17095–17112, 2023, doi: 10.1007/s00521-023-08568-z.

[11] I. K. M. Jais, A. R. Ismail, and S. Q. Nisa, "Adam Optimization Algorithm for Wide and Deep Neural Network," *Knowl. Eng. Data Sci.*, vol. 2, no. 1, p. 41, 2019, doi: 10.17977/um018v2i12019p41-46.

[12] B. Cortiñas-Lorenzo and F. Pérez-González, "Adam and the Ants: On the Influence of the Optimization Algorithm on the Detectability of DNN Watermarks," *Entropy*, vol. 22, no. 12, p. 1379, Dec. 2020, doi: 10.3390/e22121379.

[13] P. Ramachandran, T. Eswarlal, M. Lehman, and Z. Colbert, "Assessment of optimizers and their performance in autosegmenting lung tumors," *J. Med. Phys.*, vol. 48, no. 2, pp. 129–135, 2023, doi: 10.4103/jmp.jmp_54_23.

[14] P. Podder *et al.*, "LDDNet: A Deep Learning Framework for the Diagnosis of Infectious Lung Diseases," *Sensors*, vol. 23, no. 1, 2023, doi: 10.3390/s23010480.

[15] C. Annamalai, C. Vijayakumaran, V. Ponnusamy, and H. Kim, "Optimal ElGamal Encryption with Hybrid Deep-Learning-Based Classification on Secure Internet of Things Environment," *Sensors*, vol. 23, no. 12, p. 5596, Jun. 2023, doi: 10.3390/s23125596.

[16] R. Elshamy, O. Abu-Elnasr, M. Elhoseny, and S. Elmougy, "Improving the efficiency of RMSProp optimizer by utilizing Nestrove in deep learning," *Sci. Rep.*, vol. 13, no. 1, p. 8814, May 2023, doi: 10.1038/s41598-023-35663-x.

[17] X. Jiang, B. Hu, S. Chandra Satapathy, S. H. Wang, and Y. D. Zhang, "Fingerspelling Identification for Chinese Sign Language via AlexNet-Based Transfer Learning and Adam Optimizer," *Sci. Program.*, vol. 2020, 2020, doi: 10.1155/2020/3291426.

[18] A. Daneshvar, M. Ebrahimi, F. Salahi, M. Rahmaty, and M. Homayounfar, "Brent Crude Oil Price Forecast Utilizing Deep Neural Network Architectures," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–13, May 2022, doi: 10.1155/2022/6140796.

[19] V. Ojha and G. Nicosia, "Backpropagation Neural Tree," *Neural*

Networks, vol. 149, pp. 66–83, May 2022, doi: 10.1016/j.neunet.2022.02.003.

[20] B. Zhu, Y. Shi, J. Hao, and G. Fu, "Prediction of Coal Mine Pressure Hazard Based on Logistic Regression and Adagrad Algorithm—A Case Study of C Coal Mine," *Appl. Sci.*, vol. 13, no. 22, 2023, doi: 10.3390/app132212227.

[21] F. Aamir, I. Aslam, M. Arshad, and H. Omer, "Accelerated Diffusion-Weighted MR Image Reconstruction Using Deep Neural Networks," *J. Digit. Imaging*, vol. 36, no. 1, pp. 276–288, Nov. 2022, doi: 10.1007/s10278-022-00709-5.

[22] G. Ayana, J. Park, J.-W. Jeong, and S. Choe, "A Novel Multistage Transfer Learning for Ultrasound Breast Cancer Image Classification," *Diagnostics*, vol. 12, no. 1, p. 135, Jan. 2022, doi: 10.3390/diagnostics12010135.

[23] R. Sirisha, N. Anjum, and K. Vaidehi, "INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY Implementation of CNN and ANN for Fashion-MNIST-Dataset using Different Optimizers," *Indian J. Sci. Technol.*, vol. 15, no. 47, pp. 2639–2645, 2022, [Online]. Available: https://www.indjst.org/

[24] A. S. Henrique *et al.*, "Classifying Garments from Fashion-MNIST Dataset Through CNNs," *Adv. Sci. Technol. Eng. Syst. J.*, vol. 6, no. 1, pp. 989–994, 2021, doi: 10.25046/aj0601109.

[25] O. M. Khanday, S. Dadvandipour, and M. A. Lone, "Effect of filter sizes on image classification in CNN: A case study on CFIR10 and fashion-MNIST datasets," *IAES Int. J. Artif. Intell.*, vol. 10, no. 4, pp. 872–878, 2021, doi: 10.11591/ijai.v10.i4.pp872-878.

[26] Y. Tang, H. Cui, and S. Liu, "Optimal Design of Deep Residual Network Based on Image Classification of Fashion-MNIST Dataset," *J. Phys. Conf. Ser.*, vol. 1624, no. 5, pp. 0–7, 2020, doi: 10.1088/1742-6596/1624/5/052011.

[27] M. Kayed, A. Anter, and H. Mohamed, "Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture," *Proc. 2020 Int. Conf. Innov. Trends Commun. Comput. Eng. ITCE 2020*, no. June, pp. 238–243, 2020, doi: 10.1109/ITCE48509.2020.9047776.

[28] H. H. Zhu *et al.*, "Space-efficient optical computing with an integrated chip diffractive neural network," *Nat. Commun.*, vol. 13, no. 1, pp. 1–9, 2022, doi: 10.1038/s41467-022-28702-0.

[29] T. Hur, L. Kim, and D. K. Park, "Quantum convolutional neural network for classical data classification," *Quantum Mach. Intell.*, vol. 4, no. 1, pp. 1–18, 2022, doi: 10.1007/s42484-021-00061-x.

[30] O. Nocentini, J. Kim, M. Z. Bashir, and F. Cavallo, "Image Classification Using Multiple Convolutional Neural Networks on the Fashion-MNIST Dataset," *Sensors*, vol. 22, no. 23, p. 9544, Dec. 2022, doi: 10.3390/s22239544.

[31] S. Yang, S. Hoque, and F. Deravi, "Adaptive Template Reconstruction for Effective Pattern Classification," *Sensors*, vol. 23, no. 15, p. 6707, Jul. 2023, doi: 10.3390/s23156707.

[32] S. Coleman, D. Kerr, and Y. Zhang, "Image Sensing and Processing with Convolutional Neural Networks," *Sensors*, vol. 22, no. 10, p. 3612, May 2022, doi: 10.3390/s22103612.

[33] V. Terziyan, D. Malyk, M. Golovianko, and V. Branytskyi, "Hyper-flexible Convolutional Neural Networks based on Generalized Lehmer and Power Means," *Neural Networks*, vol. 155, pp. 177–203, Nov. 2022, doi: 10.1016/j.neunet.2022.08.017.

[34] K. Wang, C. Xu, G. Li, Y. Zhang, Y. Zheng, and C. Sun, "Combining convolutional neural networks and self-attention for fundus diseases identification," *Sci. Rep.*, vol. 13, no. 1, p. 76, Jan. 2023, doi: 10.1038/s41598-022-27358-6.

[35] E. Chu, D. Li, and Y. Tong, "Optimized federated learning based on Adagrad algorithm and algorithm optimization," *Appl. Comput. Eng.*, vol. 19, no. 1, pp. 9–17, Oct. 2023, doi: 10.54254/2755-2721/19/20231000.

[36] I. Naseer, S. Akram, T. Masood, A. Jaffar, M. A. Khan, and A. Mosavi, "Performance Analysis of State-of-the-Art CNN Architectures for LUNA16," *Sensors*, vol. 22, no. 12, p. 4426, Jun. 2022, doi: 10.3390/s22124426.

[37] Y. S. Saboo, S. Kapse, and P. Prasanna, "Convolutional Neural Networks (CNNs) for Pneumonia Classification on Pediatric Chest Radiographs," *Cureus*, Aug. 2023, doi: 10.7759/cureus.44130.

[38] M. Uppal *et al.*, "Enhancing accuracy in brain stroke detection: Multi-layer perceptron with Adadelta, RMSProp and AdaMax optimizers," *Front. Bioeng. Biotechnol.*, vol. 11, Sep. 2023, doi: 10.3389/fbioe.2023.1257591.

[39] R. Liang, X. Chang, P. Jia, and C. Xu, "Mine Gas Concentration Forecasting Model Based on an Optimized BiGRU Network," *ACS Omega*, vol. 5, no. 44, pp. 28579–28586, Nov. 2020, doi: 10.1021/acsomega.0c03417.

[40] S. B. ud din Tahir, A. Jalal, and K. Kim, "Wearable Inertial Sensors for Daily Activity Analysis Based on Adam Optimization and the Maximum Entropy Markov Model," *Entropy*, vol. 22, no. 5, p. 579, May 2020, doi: 10.3390/e22050579.

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 59-62
DOI: 10.36287/ijmsit.8.2.2
Received: October 9, 2024; Accepted: November 28, 2024

Thyroid Disease Diagnosis: A Study on the Efficacy of Feature Reduction and Biomarker Selection in Artificial Neural Network Models

Erman ÖZER[1*]

[1*] *Computer Engineering Department, Recep Tayyip Erdoğan University, Rize, Turkey (erman.ozer@erdogan.edu.tr) (ORCID: 0000-0002-9638-0233)*

Abstract **–**This study employs ANN to enhance thyroid disease diagnosis while minimizing features and choosing the most biomarkers. The data were analyzed focusing on three key indicators of thyroid function: TSH, TT4, and FTI. All of these biomarkers are vital signs that reflect thyroid activity and are incorporated in ANN models. This is achievable by minimizing the number of features and there by the Billboard ANN models deliver high diagnostic accuracy and high computational effectiveness. Computing with this simplified dataset results in faster computation times while at the same time, maintaining a high degree of diagnostic accuracy. Thus, the profound features of TSH, TT4, and FTI as indices of thyroid disorders, as well as the introduction of these markers into simple diagnostic algorithms, are discussed. The regression values achieved with the complete dataset were 0.66 for the training phase, 0.62 for validation, and 0.61 for testing. Conversely, utilizing the reduced dataset resulted in regression values of 0.67 during training, 0.99 in validation, and 0.80 in testing. Hence this study supports the application of ANN models in medical diagnosis by adding to the existing proof to the strategy. The data suggest that the exclusion of features can enhance the speed and boost the time to obtain a precise result. These improvements could have significant implications for clinical practice, especially in enhancing the management and treatment of thyroid diseases, where precise and prompt diagnosis is essential.

Keywords *–Thyroid; Disease Diagnosis; Artificial Neural Networks; Feature Reduction*

Citation: Özer, E. Thyroid Disease Diagnosis: A Study on the Efficacy of Feature Reduction and Biomarker Selection in Artificial Neural Network Models. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 59-62.

I. Introduction

Diseases affecting the thyroid, such as hypothyroidism and hyperthyroidism, are influenced by a complex interplay of physiological processes. Accurate diagnosis and effective treatment of these conditions are challenging without robust mathematical models to capture the intricacies of thyroid function. In recent years, Artificial Neural Networks (ANNs) have emerged as a powerful tool to model these complex systems, providing valuable insights into disease diagnosis and management [1-3].

Inspired by the structure and functioning of the brain's network of neurons, ANNs emulate biological learning by creating interconnected networks capable of identifying patterns, behaviors, and pathways in data. These networks are widely applied across various domains, including classification, prediction, and complex system modeling. In the context of thyroid disease, ANNs are particularly useful for predicting outcomes based on patient information, enabling early detection and personalized treatment planning [4,5].

An ANN typically comprises three main components: an input layer, one or more hidden layers, and an output layer. The input layer processes raw data such as hormone levels (e.g., TSH, T4, FT3) and imaging data (e.g., ultrasound images). This information is then transmitted through the hidden layers, where patterns and relationships relevant to thyroid regulation are identified. Finally, the output layer provides predictions or classifications, such as the likelihood of a specific thyroid disorder. Despite their utility, ANNs have limitations, particularly their "black box" nature, which makes it challenging to interpret their decision-making processes. This lack of transparency poses challenges in medical applications, where reproducibility and explainability are critical for clinical acceptance.

To address these challenges, researchers are exploring methods to improve ANN interpretability, such as integrating explainable AI techniques and incorporating domain-specific knowledge into model architectures. By overcoming these barriers, ANNs can become more reliable tools for advancing the diagnosis and treatment of thyroid disorders while maintaining the trust of healthcare professionals.

II. Literature Review

Research on diagnosing thyroid disorders using advanced computational techniques has shown significant progress. Irina and Liviu proposed a data mining technique to classify thyroid disorders, specifically hyperthyroidism and

hypothyroidism[6]. Their research aims to assess the effectiveness of data mining in classifying thyroid diseases.

Similarly, Geetha and Baboo developed a Hybrid Differential Evolution Kernel-Based Naive Bayes algorithm, which reduced the dimensionality of thyroid data, achieving an impressive classification accuracy of 97.97% [7]. These findings underscore the role of sophisticated algorithms in enhancing diagnostic outcomes.

Further studies have validated the effectiveness of machine learning in thyroid disease diagnostics. Doğantekin et al. introduced the ADSTG (Automatic Diagnosis System based on the Thyroid Gland), which employed Principal Component Analysis (PCA) for dimensionality reduction and Least Squares Support Vector Machines (LS-SVM) for binary classification. This hybrid approach achieved a notable accuracy of 97.67%, illustrating the benefits of combining dimensionality reduction techniques with advanced classifiers [8].

Recent advances in deep learning have further pushed the boundaries of diagnostic performance. Yadav et al. reported 99.95% accuracy in detecting thyroid diseases, demonstrating the success of hybrid methods when deep-learning with intricate medical datasets[9]. Usman et al. noted the potential of artificial intelligence algorithms and multiple linear regression for predicting thyroid hormone balance, indicating a greater implication of AI in clinical diagnostics with this regard[10].

Despite these advancements, challenges such as the interpretability of complex models remain. This study addresses these gaps by employing ANNs in a novel configuration. While ANNs excel in identifying intricate patterns, their "black box" nature often limits their interpretability. To overcome this, the current research emphasizes improving both diagnostic accuracy and computational efficiency through advanced techniques.

Building on prior studies, this work utilizes ANN technology to classify thyroid disorders with enhanced accuracy. By leveraging the latest datasets and incorporating a larger sample size, the proposed model achieves significant improvements in computational efficiency. Although the diagnostic accuracy remains comparable to existing methods, the reduction in computational complexity marks a substantial step forward. This innovation contributes to early detection and better management of thyroid diseases, highlighting the transformative potential of ANNs in medical diagnostics.

Our study stands out from previous research by using the latest data and including a larger number of patients. Although the accuracy rate using similar feature reduction techniques stayed the same, our method significantly improved computational efficiency.

III. Dataset and Methodology

The thyroid disease dataset, sourced from the UCI Machine Learning Repository, is a valuable resource for machine learning tasks related to thyroid health. The databases from the Garvan Institute include medical data from patients with thyroid disease, each containing around 2,800 training instances and 972 test instances[11]. In total, there are approximately 29 features in each database, classified as Boolean or continuous variables.

Quinlan offers databases that may have missing data. The Aeberhard database has 215 instances, 5 features, and 3 classes of medical data. The Turney database has 3,772 training instances, 3,428 test instances, 5 features, and 3 classes of both medical and cost data. This versatile dataset supports tasks such as diagnosing thyroid disease, assessing treatment efficacy, and cost-effectiveness analysis, making it valuable for clinical practice and research.

Table 1. Thyroid Disease Dataset

Feature	Type	Description
age	Continuous	Patient's age
sex	Binary	Patient's gender
on thyroxine	Binary	Whether the patient is taking thyroid hormone thyroxine
query on thyroxine	Binary	Whether the patient has inquired about thyroxine
on antithyroid medication	Binary	Whether the patient is taking medication to treat hyperthyroidism
sick	Binary	Whether the patient is currently sick
pregnant	Binary	Whether the patient is pregnant
thyroid surgery	Binary	Whether the patient has undergone thyroid surgery
I131 treatment	Binary	Whether the patient has received radioactive iodine treatment
query hypothyroid	Binary	Whether the patient has inquired about hypothyroidism
query hyperthyroid	Binary	Whether the patient has inquired about hyperthyroidism
lithium	Binary	Whether the patient is taking lithium medication
goiter	Binary	Whether the patient has a goiter
tumor	Binary	Whether the patient has a thyroid tumor
hypopituitary	Binary	Whether the patient has hypopituitarism
psych	Binary	Whether the patient has a psychiatric condition
referral source	Categorical	Source of referral for the patient
binary class	Binary	Whether the patient has thyroid disease

Thyroid disease is evaluated using key biomarkers such as Thyroid Stimulating Hormone (TSH), thyroxine (T4), and free triiodothyronine (FT3). Elevated TSH points to hypothyroidism, low TSH indicates hyperthyroidism, and high T4 and FT3 levels are associated with increased thyroid function. Refer to Table 2 in this study for more information on these specific biomarkers.

Table 2. Key Biomarkers and Their Clinical Significance İn Thyroid Disease Diagnosis

Feature	Type	Description
TSH measured	Measured TSH value	Input
T3 measured	Measured T3 value	Input
TT4 measured	Measured TT4 value	Input
T4U measured	Measured T4U value	Input
FTI measured	Measured FTI value	Input
binaryClass	Thyroid disease status: 0 = normal, 1 = disease present	Output

The formula for regression is provided as follows:

$$\beta_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

- X_i represents the individual values of the predictor variable X,
- $\bar{X}$ denotes the mean of X
- Y_i represents the individual values of the target variable Y
- $\bar{Y}$ denotes the mean of Y

IV. Results

TSH, total TT4, and FTI are important biomarkers for diagnosing thyroid disease because they accurately reflect thyroid gland activity. A model using these values demonstrated high accuracy in diagnosing thyroid disease, confirming the effectiveness of these biomarkers for clinical use. These results highlight the critical role of TSH, TT4, and FTI in evaluating thyroid function, offering valuable insights to enhance diagnostic accuracy.

The pituitary gland releases TSH to signal the thyroid to release T4 and T3. T4 is converted to the active form, FT3. The FTI measurement reflects levels of TT4 and T3. The best validation performance of the entire dataset (Fig. 1) and the ANN results (Fig. 2) were both obtained using the complete dataset.

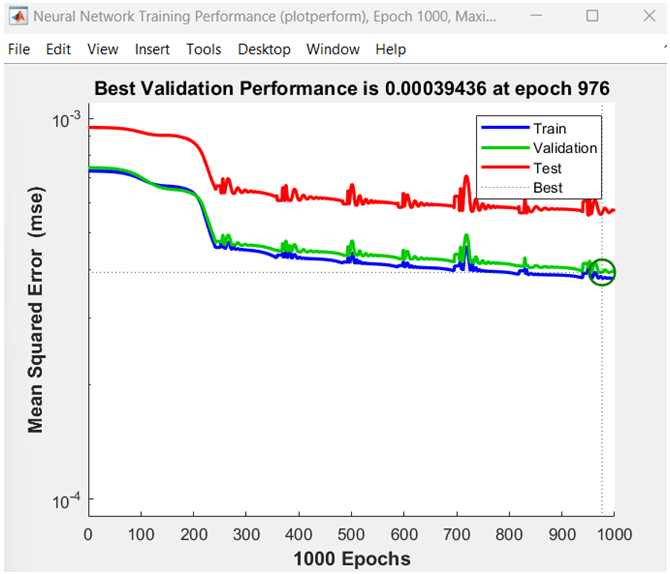


Fig. 1. Best Validation Performance of the Whole Dataset

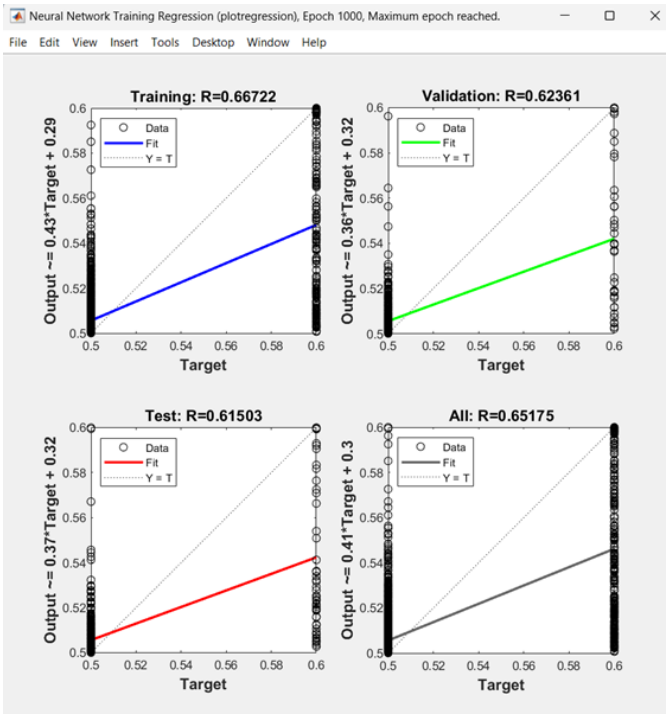


Fig. 2. ANN results of the whole dataset

Data selection was based on the correlation coefficient, which measures the relationship between variables. Strong correlations were found among TSH, TT4, and FTI values, suggesting a close connection among these biomarkers. This indicates that these variables collectively offer a thorough evaluation of thyroid function.

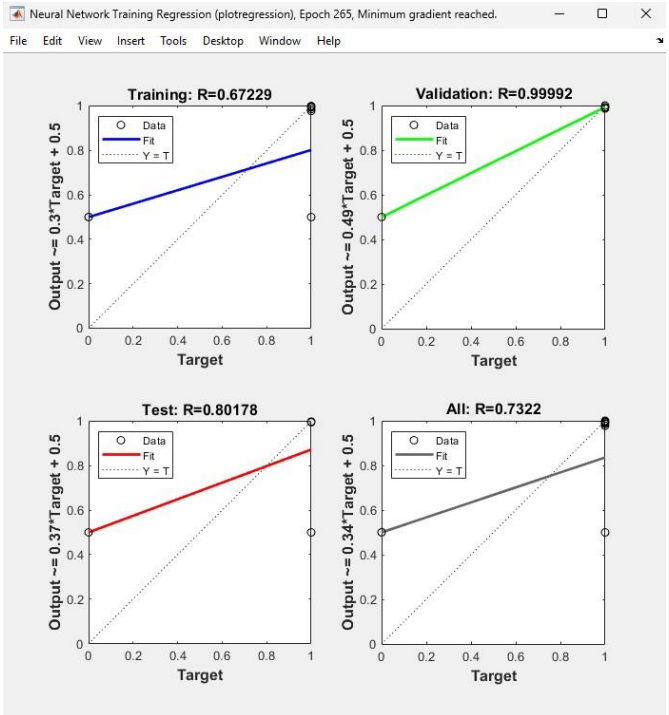


Fig. 3. ANN results of the reduced dataset

Reducing the dataset features led to improved results and saved significant time. The study achieved approximately 7% faster results.

V. CONCLUSION

This study highlights the effectiveness of ANNs in diagnosing thyroid diseases with improved computational efficiency and diagnostic accuracy. By leveraging key biomarkers such as TSH, TT4, and FTI, the proposed model demonstrates a significant enhancement in both performance and speed. Utilizing a reduced dataset further improved validation results, achieving a regression value of 0.99, and reduced computational time by approximately 7%, underscoring the practical benefits of feature reduction.

ANNs, inspired by the structure and functionality of the human brain, offer robust capabilities for modeling complex medical systems. They are invaluable tools in medical diagnostics, particularly in tasks such as classification and prediction, enabling early detection and effective management of conditions like hyperthyroidism and hypothyroidism. The strengths of ANNs include their ability to handle incomplete data, tolerate faults, and process information in parallel. However, their reliance on specialized hardware, unpredictable behavior, and the challenge of optimizing network architecture present areas for further research.

This study contributes to the growing body of evidence supporting the use of ANN models in clinical applications, with a focus on achieving a balance between computational efficiency and diagnostic accuracy. While ANN models exhibit limitations in interpretability, integrating explainable AI techniques in future work could address these challenges, further enhancing their clinical relevance. Overall, the findings of this study pave the way for more efficient and precise diagnostic tools, offering significant implications for improving the management and treatment of thyroid diseases.

VI. LIMITATIONS AND FUTURE WORK

This study is limited by the size and diversity of the dataset used, as it may not fully represent the broad range of thyroid disease cases encountered in clinical practice. Additionally, the demographic distribution of the dataset could introduce biases, potentially affecting the generalizability of the results. Another limitation is the "black box" nature of the ANN model, which restricts its interpretability and could hinder its acceptance in critical clinical settings where decision transparency is vital.

Future work could address these limitations by incorporating larger, more diverse datasets to improve model robustness and applicability. Exploring advanced techniques, such as explainable AI (XAI), could enhance the interpretability of the ANN model, making it more suitable for clinical decision-making. Real-time data integration, such as wearable device outputs or continuous monitoring systems, could further refine diagnostic capabilities and expand the model's utility. Lastly, extending the methodology to include multi-disease diagnostic capabilities would make the model more versatile and impactful for broader healthcare applications.

ACKNOWLEDGMENT

The heading of the Acknowledgment section and the References section must not be numbered.

Authors' Contributions

The authors' contributions to the paper are equal.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics

REFERENCES

[1] M. A. -A. -R. Asif et al., "Computer Aided Diagnosis of Thyroid Disease Using Machine Learning Algorithms," 2020 11th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 2020, pp. 222-225, doi: 10.1109/ICECE51571.2020.9393054.

[2] A. R. Rao and B. S. Renuka, "A Machine Learning Approach to Predict Thyroid Disease at Early Stages of Diagnosis," 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 2020, pp. 1-4, doi: 10.1109/INOCON50539.2020.9298252.

[3] M. Riajuliislam, K. Z. Rahim and A. Mahmud, "Prediction of Thyroid Disease(Hypothyroid) in Early Stage Using Feature Selection and Classification Techniques," 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, 2021, pp. 60-64, doi: 10.1109/ICICT4SD50815.2021.9397052.

[4] A. Begum and A. Parkavi, "Prediction of thyroid Disease Using Data Mining Techniques," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 342-345, doi: 10.1109/ICACCS.2019.8728320.

[5] E. Özer, N. Sevinçkan and E. Demiroğlu, "Comparative Analysis of Computational Intelligence Techniques in Financial Forecasting: A Case Study on ANN and ANFIS Models," 2024 32nd Signal Processing and Communications Applications Conference (SIU), Mersin, Turkiye, 2024, pp. 1-4, doi: 10.1109/SIU61531.2024.10600769.

[6] A. Begum and A. Parkavi, "Prediction of thyroid Disease Using Data Mining Techniques," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 342-345, doi: 10.1109/ICACCS.2019.8728320.

[7] K. Geetha and C. S. S. Baboo, "An Empirical Model for Thyroid Disease Classification using Evolutionary Multivariate Bayesian Prediction Method", Glob. J. Comput. Sci. Technol. E Network, Web Secur., 16:1, 242-250.

[8] Esin Dogantekin, Akif Dogantekin, Derya Avci, An automatic diagnosis system based on thyroid gland: ADSTG, Expert Systems with Applications, Volume 37, Issue 9, 2010, Pages 6368-6372, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2010.02.083.

[9] Yadav, D.C., Pal, S. Prediction of thyroid disease using decision tree ensemble method. Hum.-Intell. Syst. Integr. 2, 89–95 (2020). https://doi.org/10.1007/s42454-020-00006-y

[10] Usman, Abdullahi & Alhosen, Mohamed & Degm, Ali & Alsharksi, Ahmed & Muhammed Naibi, Aishat & Abba, Sani & Muhammad, Umar Ghali. (2020). Applications of Artificial Intelligence-Based Models and Multi- Linear Regression for the Prediction of Thyroid Stimulating Hormone Level in the Human Body.

[11] Quinlan,Ross. Thyroid Disease. UCI Machine Learning Repository. https://doi.org/10.24432/C5D010.

[12] Sheehan MT. Biochemical Testing of the Thyroid: TSH is the Best and, Oftentimes, - A Review for Primary Care. Clin Med Res. 2016 Jun;14(2):83-92. doi: 10.3121/cmr.2016.1309. Epub 2016 May 26. PMID: 27231117; PMCID: PMC5321289.

[13] Feldt-Rasmussen U, Klose M. Clinical Strategies in the Testing of Thyroid Function. [Updated 2020 Nov 20]. In: Feingold KR, Anawalt B, Blackman MR, et al., editors. South Dartmouth (MA): ncbi.nlm.nih.gov/books/NBK285558/

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 63-68
DOI: 10.36287/ijmsit.8.2.3
Received: November 6, 2024; Accepted: November 28, 2024

Feasibility Analysis of Wind-Battery Energy Storage Hybrid Systems in Türkiye

Busra Eyupoglu[1], Kubra Nur Akpinar [2*]

[1]*Electrical&Energy, Marmara University, Istanbul, Türkiye (busra_eyupoglu@outlook.com), (ORCID: 0009-0000-1048-5441)*
[2]*Electrical&Energy, Marmara University, Istanbul, Türkiye (knur@marmara.edu.tr), (ORCID: 0000-0003-4579-4070)*

Abstract – This study investigates the financial viability of integrating wind turbines and battery energy storage systems (BESS) in hybrid power plants within Türkiye's renewable energy framework. The analysis focuses on the economic performance of the Enercon E58 wind turbine with a 30 MW capacity, a standalone BESS with a 30 MW/30 MWh capacity, and a hybrid system combining both technologies. Key financial metrics such as Net Present Value (NPV) and Internal Rate of Return (IRR) are employed to evaluate the investment potential of each scenario. The results indicate that while the wind turbine scenario yields a positive NPV, its IRR is relatively low due to high capital (CAPEX) and operational (OPEX) expenditures. Conversely, the BESS scenario demonstrates a lower CAPEX and minimal OPEX, resulting in moderate NPV and IRR. The hybrid configuration, leveraging both wind generation and energy storage, shows a balanced investment profile with improved NPV and IRR values, suggesting an enhanced financial outlook. This study provides valuable insights for investors, policymakers, and stakeholders, emphasizing the strategic benefits of hybrid renewable energy systems in achieving Türkiye's carbon-neutral energy targets.

Keywords *– BESS, Economic analysis, Feasibility, Hybrid energy systems, Wind energy.*

Citation: Eyupoglu, B., Akpinar, K. (2024). Feasibility Analysis of Wind-Battery Energy Storage Hybrid Systems in Türkiye. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 63-68.

I. INTRODUCTION

Energy security and climate change mitigation are pivotal concerns in contemporary global energy policies. In response to the European Green Deal's ambitious carbon-neutral (Net-Zero) targets, Türkiye is making substantial strategic advancements within its energy sector [1]. To meet these targets, there is an urgent need to promote the adoption of renewable energy sources and to integrate advanced energy storage systems [2]. The combined utilization of battery energy storage systems (BESS) and wind energy plants is proving to be instrumental in enhancing energy security and fostering greater grid flexibility [3].

Recent years have seen Türkiye implementing significant regulatory measures to encourage the deployment of BESS. Notably, the legislative changes published in the Official Gazette on November 19, 2022, have led the Energy Market Regulatory Authority (EMRA) to begin accepting applications for electricity generation with integrated storage [4]. These regulations are designed to facilitate the construction of wind power plants equivalent to the storage capacities pledged by investors. By mitigating the fluctuations inherent in energy production, these policies aim to increase energy security and stability.

The variability of wind energy presents challenges in maintaining consistent energy production, potentially jeopardizing energy security [5]. Battery energy storage systems offer a solution to this problem by balancing these fluctuations and enhancing overall grid stability [6]. BESS can be utilized for a range of functions within electrical grids, including primary frequency support, peak demand management, energy arbitrage, reserve capacity provision, and advanced frequency control. Furthermore, when integrated with solar and wind energy plants, BESS can support and stabilize these variable renewable energy sources [7].

In the literature, several studies have explored wind farm integration with BESS, such as [8], which highlights that behind the meter BESS reduces curtailments and improves resource adequacy but relies on capacity value monetization for economic viability. Similarly, [9] proposes an economic assessment tool to evaluate BESS viability in renewable power plants for various market applications, demonstrating that balancing market participation can yield positive internal rates of return, although combining functionalities provides limited additional benefits.

This paper investigates the development of energy management algorithms for wind-BESS hybrid power plants, which are currently being planned and initiated in specific regions of Türkiye. The study involves a detailed analysis based on real wind speed data to calculate the energy output from wind turbines. Additionally, it evaluates factors such as the state of charge of the batteries, the demands of grid operators, participation in ancillary services, and potential earnings from the day-ahead market. The paper presents the development of rule-based control algorithms tailored to optimize system performance and conducts comprehensive profitability analyses.

This paper lays a solid foundation for future investments by advancing the development of sustainable and efficient management strategies for both standalone wind power and

wind-battery energy storage hybrid systems. For standalone wind power, the focus is on optimizing wind turbine operations to maximize returns despite high CAPEX and OPEX. In contrast, the wind-battery hybrid approach leverages the complementary benefits of combining wind power with battery energy storage, aiming for enhanced financial performance through balanced investment and improved revenue potential. It provides critical insights into strategic deployment, offering guidance that benefits stakeholders, including investors, policymakers, and industry professionals. This research supports Türkiye's broader renewable energy objectives and energy transition goals, contributing to more informed decision-making and strategic planning within the renewable energy sector.

II. Materials and Method

The study used 30 Enercon E58 wind turbines, each with a 1 MW capacity, and a Battery Energy Storage System (BESS) with a 30 MWh capacity, installed in Istanbul's Catalca region. Wind and atmospheric data were analyzed to calculate wind power potential and assess system performance using MATLAB for graphical representation.

A. Wind Turbine

For the wind turbine to be used in Istanbul, 30 units of the Enercon E58 1 MW (1000 kW) model have been selected. The technical specifications and power curve of turbine are shown in the Table-1 and Fig.1 below.

Table 1. Technical Specifications of Enercon E58 Wind Turbine [10]

Specification	Value
Nominal Power	1000kW
Rotor Diameter	58m
Tower Height	70m
Cut-in Wind Speed	4 m/s
Cut-out Wind Speed	25 m/s

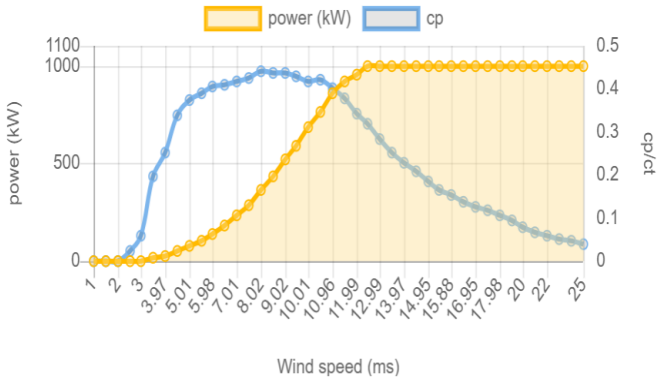


Fig. 1. Wind turbine power and capacity factor (cp) curve [10]

B. Battery Energy Storage System

Generally, the energy capacity of Battery Energy Storage Systems (BESS) varies depending on the discharge duration. If a discharge duration is specified (for example, 1 hour, 2 hours, 4 hours), the energy capacity can be calculated based on this duration. In this study, for a BESS with a 1-hour discharge duration, the maximum capacity will be 30 MW * 1 hour = 30 MWh.

C. Location and Data Acquisition

For the wind turbines' installation, the Çatalca region in Istanbul was selected. Wind speed data for this region were obtained from the Meteoblue website [11]. This data is crucial for accurately assessing wind power potential. A map indicating the installation area which has an ideal wind speed for a wind farm with a 30 MWe capacity was shown in Fig 2 [12].

Fig. 2. Wind turbine instalment map

D. Data Analysis

Atmospheric data specific to the Çatalca region, including wind speed, temperature (in Celsius), and pressure (in hPa) at a height of 10 meters, were analysed graphically. This analysis aimed to determine the wind power generated by the turbines and to understand the wind speed variations throughout the year.

E. Wind Power Calculation

To estimate wind speeds at various elevations, the Hellman coefficient (α) was used. The coefficient was selected based on surface roughness and is detailed in Table 2. The Hellman coefficient is essential for accurately predicting wind speeds at different heights.

Table 2. Hellman Coefficient

Location	α
Unstable air over open water	0.06
Neutral air over open water	0.10
Unstable air over flat open coast	0.11
Neutral air over flat open coast	0.16
Stable air over open water	0.27
Unstable air over populated areas	0.27
Neutral air over populated areas	0.34
Stable air over flat open coast	0.40
Stable air over populated areas	0.60

Annual wind power calculations were performed using the following equations:

$$\rho = 353\left(\frac{P}{T}\right) = \left(\frac{353}{T}\right) e^{\left(\frac{-0.0341}{T}\right)\cdot h} \quad (1)$$

$$v = v_{known}\left(\frac{h}{h_{known}}\right)^{\alpha} \quad (2)$$

$$P_W = \frac{1}{2}\rho A v^3 C_p \eta \quad (3)$$

Equation 1 represents the air density (ρ [kg/m³]), pressure (P [atm]), and temperature (T [Kelvin]) at the measurement site. Equation 2 calculates the wind speed (v [m/s]) at a specific location and height using the known height (h_{known}), wind speed (v_{known}), and the Hellman coefficient. Area (A [m²]) represents the area through which the wind passes over the

turbine blades, which is used in calculating the wind power (P_{wind} [MW]). The capacity factor (C_p) is included in Equation 3, which determines the total possible wind power by multiplying the wind power with the capacity factor.

Using Equations 1, 2, and 3, annual plots of wind speed versus time and wind power versus time over a 5-year period, along with annual plots of day-ahead market trading prices versus time over a 5-year period, have been plotted using MATLAB.

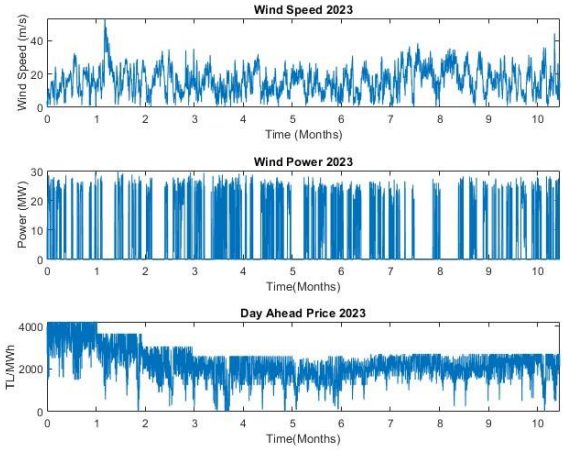


Fig. 3. Wind speed, power produced by wind turbine and DAP in 2023

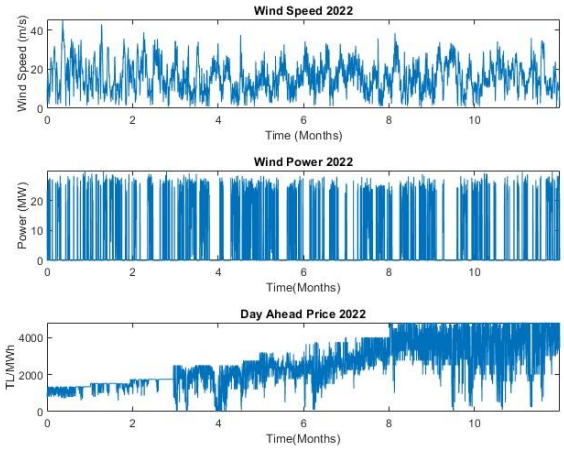


Fig. 4. Wind speed, power produced by wind turbine and DAP in 2022

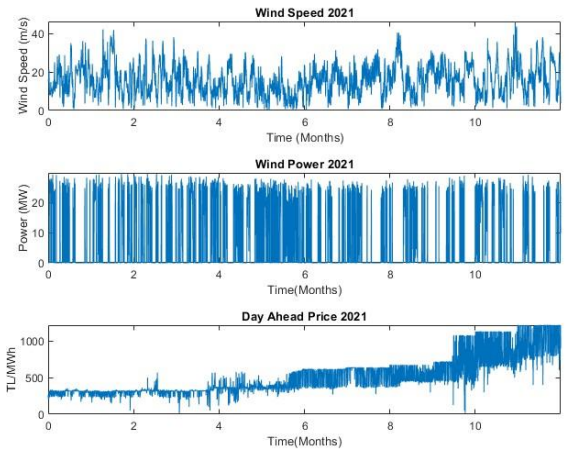


Fig. 5. Wind speed, power produced by wind turbine and DAP in 2021

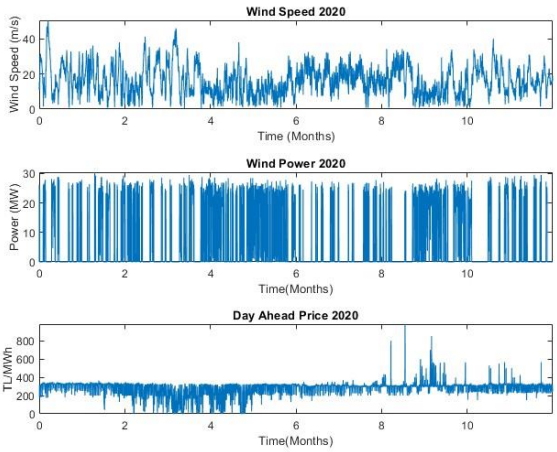


Fig. 6. Wind speed, power produced by wind turbine and DAP in 2020

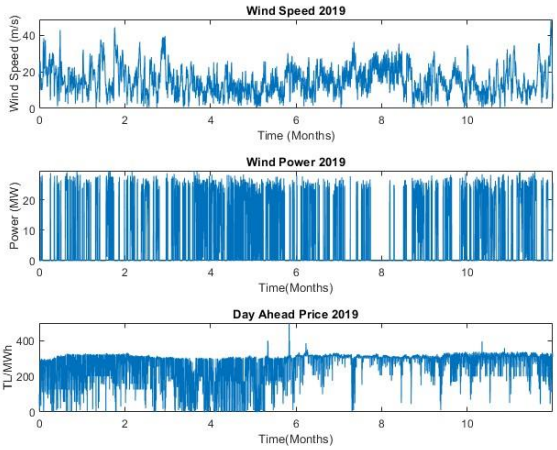


Fig. 7. Wind speed, power produced by wind turbine and DAP in 2019

F. Scenarios

Since wind speed and power prediction is not the focus of this paper, it is assumed that wind speed is forecasted accurately in the following scenarios, with power output predictions made with near 100% accuracy. All algorithms were developed based on this assumption, and the resulting graphs are interpreted accordingly.

G. Wind Turbines Only (Scenario-1)

To better observe the impact of the battery energy storage system, the profit obtained from a 30 MW installed capacity wind farm participating only in the day-ahead market will be calculated without integrating the battery energy storage system. The investment cost is assumed to be 22,450,000 Turkish Lira (TL) for installation (CAPEX) per 1 MW turbine and 1,000,000 Turkish Lira (TL) for operation (OPEX) [13]. The lifespan of the wind turbine is assumed to be 25 years [14].

The annual energy production is calculated using Equation (4), assuming that wind is present during the periods when it is within the turbine's minimum and maximum operational limits, and the turbine operates for those hours.

$$E = Power \times Hours \quad (4)$$

Here, E represents the annual energy production. Since power values are calculated on an hourly basis, they are multiplied by 8760 (total hours in a year). However, because

wind does not continuously blow at the desired optimal value, the power produced during wind periods is multiplied by the number of hours of wind to estimate total hourly energy production.

The sum of the calculated hourly power production values will provide the total annual energy production. Details of the technology used are shown in Table 1. If it is assumed that the entire produced energy is sold to the grid at day ahead price (DAP) rates, the annual profit amounts are shown in Table 3. Graphs of hourly total revenue by year are presented below.

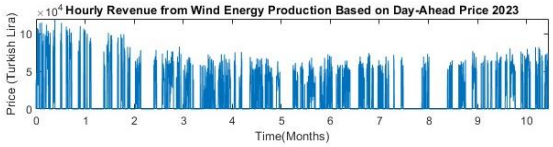


Fig. 8. Hourly revenue of wind power plant based on DAP in 2023

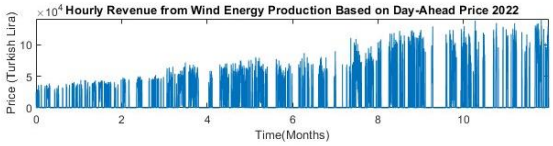


Fig. 9. Hourly revenue of wind power plant based on DAP in 2022

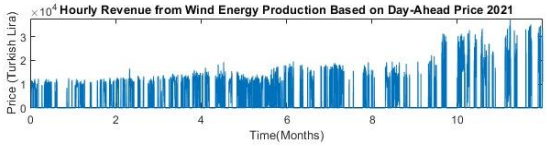


Fig. 10. Hourly revenue of wind power plant based on DAP in 2021

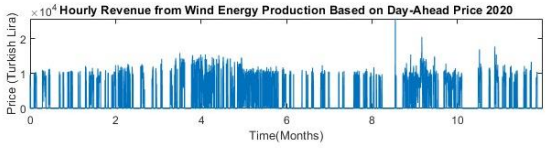


Fig. 11. Hourly revenue of wind power plant based on DAP in 2020

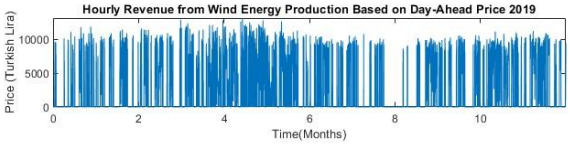


Fig. 12. Hourly revenue of wind power plant based on DAP in 2019

Net Present Value (NPV) and Internal Rate of Return (IRR) are calculated using Equations (5), (6), and (7), with the payback period (PP) derived accordingly. NPV is calculated by subtracting the initial investment (CAPEX) from the total present value of future cash flows. IRR is the discount rate that sets the NPV to zero. The payback period (PP) is the time required to recover the invested capital. It is calculated by finding the period when the accumulated cash flow becomes zero or positive.

$$\mathrm{NPV} = \sum_{t=1}^{n} \frac{R_t - C_t}{(1+r)^t} - \mathrm{CAPEX} \tag{5}$$

$$0 = \sum_{t=1}^{n} \frac{R_t - C_t}{(1+r_{IRR})^t} - \mathrm{CAPEX} \tag{6}$$

$$\mathrm{PP} = \min\left\{ t : \sum_{i=1}^{t} (R_i - C_i) \geq \mathrm{CAPEX} \right\} \tag{7}$$

Here:

- R_t: Revenue in period t
- C_t: Cost in period t
- r: Discount rate
- t: Period (from 1 to n)
- n: Total number of periods
- *CAPEX*: Initial capital expenditure

Table 3. Unit Energy Cost for Wind Farm

Technology	Installed Capacity (MW)	Lifespan	CAPEX (TL/kWh)	OPEX (TL/year)
Enercon E58	30	25 years	673,500,000	30,000,000

- Economic Discount Rate (r): 10%
- Annual Revenue: 150745969,9 (based on the year 2022)
- Installed Capacity: 30 x 1 MW = 30 MW
- OPEX (TL/year): 30 x 1,000,000 = 30,000,000
- Lifespan (n): 25 years
- CAPEX: 30 x 22,450,000 = 673,500,000 TL

Net Present Value (NPV): 422,516,000.7093 TL
Payback Period (PP): 6 years
Internal Rate of Return (IRR): 17.6179%

H. Wind-BESS Hybrid Participation in Day-Ahead Market and BESS Charge/Discharge Control (Scenario-2)

In this scenario, the economic performance of the wind-BESS hybrid system is analysed, considering participation in the Day-Ahead Market (DAM) and the control of battery charge/discharge cycles. The hybrid system integrates a 30 MW wind farm with a BESS, which is used to balance fluctuations in wind energy and optimize revenue through market participation. As seen in the flowchart of the scenario in the Figure 13, BESS provides energy to the grid during periods when wind power is unavailable. When wind speeds reach optimal levels, the BESS begins charging. Once the battery is sufficiently charged, it remains idle until the wind-generated energy drops to zero, at which point it again supplies energy to the grid according to its capacity. The study evaluates the impact of incorporating BESS on the overall system performance, including the financial returns and operational efficiency, by implementing charging and discharging strategies based on DAM prices and grid demands.

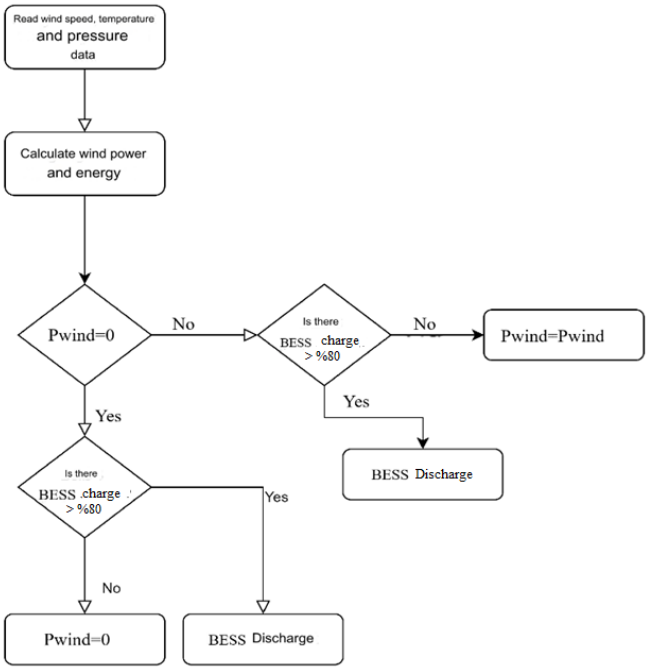


Fig. 13. Flowchart for Scenario-2

Table 4. Wind Farm-BESS Unit Energy Cost

Technology	Installed Capacity (MW)	Life-span (years)	CAPEX (TL/kWh)	OPEX (TL/year)
Enercon E58	30	25	673,500,000	30,000,000
BESS	30 MW/30 MWh	20	225,720,000	141 ,075
Hybrid Power Plant	30MW+30MWh	20	899,220,000	30,141,075

- Economic Discount Rate (r): 10%
- Revenue: 150,540,020 TL (Based on the year 2022)
- Installed Capacity: 30 MW Wind Power Plant +30MW/30MWh BESS
- OPEX (TL/year): 30,141,075 TL
- Lifetime (n): 20 years
- CAPEX: 899,220,000 TL

Net Present Value (NPV): 139,955,702.9353TL
Payback Period (PP): 8 years
Internal Rate of Return (IRR): 12.0017%

III. RESULTS

Table 5 highlights the key differences between the two scenarios in terms of financial metrics, investment costs, and overall system performance.

Table 5. Comparison of Financial Metrics for Scenario-1 vs. Scenario-2

Metric	Scenario-1: Wind Turbines Only	Scenario-2: Wind Turbines with BESS
NPV	422,516,000.7093 TL	139,955,702.9353 TL
PP	6 years	8 years
IRR	17.6179%	12.0017%
Energy Management Strategy	Wind energy sales only	Enhanced with BESS
CAPEX	673,500,000 TL	899,220,000 TL
OPEX	30,000,000 TL	30,141,075 TL

The financial analysis of Scenario-1 versus Scenario-2, which integrates a BESS, reveals several key differences in economic performance. In Scenario-2, the NPV benefits from enhanced energy management capabilities provided by the BESS. However, the initial costs associated with integrating the BESS result in a higher overall investment, which extends the PP compared to Scenario-1. This longer payback period suggests that the hybrid system takes more time to recover its costs, reflecting the increased complexity and expense of incorporating energy storage solutions.

Despite the advantages of improved energy management and grid stability with Scenario-2, the IRR is lower than that of Scenario-1. The additional costs and risks associated with the BESS contribute to this reduced IRR. While Scenario-1 relies solely on wind energy sales, Scenario-2 offers optimized energy usage and potential revenue enhancement through better grid integration. Furthermore, BESS integration enhances the flexibility of renewable energy sources, facilitating their integration into broader energy systems, which not only improves economic performance but also supports long-term environmental sustainability. This increased stability and cleaner energy production may align with sustainable energy development goals. Nevertheless, the trade-off between higher initial costs and longer payback periods versus the potential for increased long-term revenue and stability must be carefully considered when evaluating the overall financial viability of each scenario.

IV. DISCUSSION

The comparative analysis of two renewable energy investment scenarios—one involving wind turbines alone and the other combining wind turbines with a BESS—reveals distinct differences in their economic and operational performances. Scenario 1, which solely utilizes wind turbines, demonstrates superior economic feasibility with a NPV of 422,516,000.71 TL, an IRR of 17.62%, and a relatively short PP of 6 years. The lower CAPEX of 673,500,000 TL and annual OPEX of 30,000,000 TL contribute to this favorable outcome. The financial performance in this scenario is largely driven by the direct sale of wind-generated electricity, which enables quicker recovery of the initial investment and higher overall returns.

Conversely, Scenario 2, which integrates a 30 MW/30 MWh BESS with the wind power system, exhibits a more complex economic profile. Although the inclusion of BESS enhances the system's operational flexibility by allowing energy storage and optimized power dispatch, it results in a lower NPV of 139,955,702.94 TL and a reduced IRR of 12.00%. Additionally, the Payback Period extends to 8 years, reflecting the impact of a higher CAPEX of 899,220,000 TL and a slight increase in OPEX to 30,141,075 TL. The incorporation of BESS introduces benefits such as increased grid stability, energy arbitrage potential, and improved reliability of power supply. However, these operational advantages are offset by the substantial initial costs and ongoing operational expenses, leading to a less attractive financial outcome compared to the wind-only scenario.

V. CONCLUSION

In conclusion, while the integration of BESS with wind turbines offers enhanced operational benefits, particularly in

terms of energy management and grid stability, it does not provide the same level of economic return as a standalone wind power system under the current economic parameters. The decision to invest in either configuration should be informed by specific investment goals, risk assessments, and future market conditions, including potential regulatory incentives and advancements in storage technologies. These findings highlight the need for a balanced approach when considering the trade-offs between capital investment, operational flexibility, and economic returns in renewable energy projects, especially in the context of evolving electricity markets and policy frameworks. Additionally, it is essential to consider cost reduction, revenue-enhancing strategies, potential government incentives and improving metrics through strategies such as increasing battery capacity or efficiency, optimizing energy sale prices, and implementing additional revenue models could be beneficial.

ACKNOWLEDGMENT

This study has been financially supported by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) 2209-A - Research Project Support Programme for Undergraduate Students.

Authors' Contributions

The authors' contributions to the paper are equal.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics

REFERENCES

[1] G., Şahin, M. A., Taksim, & B. Yitgin,"Effects of the European Green Deal on Turkey's electricity market". *İşletme Ekonomi ve Yönetim Araştırmaları Dergisi*, *4*(1), pp. 40-58. 2021.

[2] M. Y. Worku, "Recent advances in energy storage systems for renewable source grid integration: a comprehensive review". *Sustainability*, *14*(10), 5985. 2022.

[3] R. D., Filho, A. C., Monteiro, T., Costa, A., Vasconcelos, A. C., Rode, & M. Marinho,"Strategic Guidelines for Battery Energy Storage System Deployment: Regulatory Framework, Incentives, and Market Planning". *Energies*, *16*(21), 7272. 2023.

[4] https://officialgazette.gov.gy/index.php/publications/1964-official-gazettes-19th-november-2022

[5] T., Kara, & A. D. Şahin, "Implications of Climate Change on Wind Energy Potential". *Sustainability*, *15*(20), 14822. 2023.

[6] X., Li, & S. Wang, "Energy management and operational control methods for grid battery energy storage systems". *CSEE Journal of Power and Energy Systems*, *7*(5), 1026-1040. 2019.

[7] Zhao, C., Andersen, P. B., Træholt, C., & Hashemi, S., "Grid-connected battery energy storage system: a review on application and integration". *Renewable and Sustainable Energy Reviews*, *182*, 113400. 2023.

[8] Dratsas, P. A., Psarros, G. N., & Papathanassiou, S. A. "Feasibility of behind-the-meter battery storage in wind farms operating on small islands". *Batteries*, *8*(12), 275, 2022.

[9] E., Lobato, L., Sigrist, A., Ortega, A., González, & J. M. Fernández, "Battery energy storage integration in wind farms: Economic viability in the Spanish market". *Sustainable Energy, Grids and Networks*, *32*, 100854, 2022.

[10] https://en.wind-turbine-models.com/turbines/114-enercon-e-58-10.58

[11] https://www.meteoblue.com/tr/hava/archive/export

[12] S. Şahin, M. Türkeş, "Assessing wind energy potential of Turkey via vectoral map of prevailing wind and mean wind of Turkey", Theor Appl Climatol 141, 1351–1366, 2020.

[13] https://atb.nrel.gov/electricity/2023/land-based_wind

[14] M., Adnan, J., Ahmad, S. F., Ali, & M. Imran, "A techno-economic analysis for power generation through wind energy: A case study of Pakistan". *Energy Reports*, *7*, 1424-1443. 2021.

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 69-75
DOI: 10.36287/ijmsit.8.2.4
Received: July 31, 2024; Accepted: September 13, 2024

Türkçe Günlük Kelime ve İfadeler Kullanarak CNN ve LSTM ile Görsel Konuşma Tanıma

Nergis Pervan Akman[1*], Talya Tümer Sivri[2], Ali Berkol[3], Hamit Erdem[4]

[1*] *Defence and Information Systems, BITES, Ankara, TURKEY (nergis.pervan@bites.com.tr) (ORCID: 0000-0003-3241-6812)*
[2] *Informatics Institute, Middle East Technical University, Ankara, TURKEY (talya.tumer@gmail.com) (ORCID: 0000-0003-1813-5539)*
[3] *Defence and Information Systems, BITES, Ankara, TURKEY (ali.berkol@bites.com.tr) (ORCID: 0000-0002-3056-1226)*
[4] *Electrics and Electronics Department, Başkent University, Ankara, TURKEY (herdem@baskent.edu.tr) (ORCID: 0000-0003-1704-1581)*

Türkçe Özet – Dudak okuma; el hareketleri, jestler ve yüz ifadeleri gibi konuşma örüntülerini, hareketlerini ve mimiklerini değerlendirmek amacıyla bir konuşmacının yüzünü incelemek olarak tanımlanmaktadır. Bilgisayarlara dudak okuma yeteneği kazandırma çalışmaları, derin öğrenmede sınıflandırma ve örüntü tanıma alanında büyüyen bir araştırma alanıdır ve günümüzde hâlâ çözülmesi gereken açık problemler barındırmaktadır. Son yıllarda, farklı dillerde konuşmayı metne dönüştürmek ve sınıflandırmak için çeşitli yöntemler geliştirilmiş ve uygulanmıştır. Ayrıca, çoğu yöntemde çok modlu veriler, yani konuşma ve görüntü verileri birleştirilmiştir. Bu çalışma, görüntülerle yeni Türkçe dudak okuma verileri sağlamayı ve Türkçe günlük kelimeler için yüksek doğrulukta bir sınıflandırma yöntemi sunmayı amaçlamaktadır. Kullanılan veriler, YouTube platformundan toplanmıştır. Bu zorlu verilerle, günlük kelimeleri ve ifadeleri sınıflandırmak için Evrişimli Sinir Ağı (Convolutional Neural Network – CNN) ve Uzun Kısa-Süreli Bellek (Long Short-Term Memory – LSTM) eğitilmiştir. Birçok deney sonucuna göre, CNN modeli daha iyi performans göstermiştir. Çoklu model verileri kullanmadan yalnızca görüntüler kullanmak, belleğin yorgunluğunu önler ve hesaplama süresini azaltır. Ayrıca, literatürde sınırlı bir çeşitlilik olduğundan, bu çalışma çok sınıflı Türkçe bir veri kümesi sunmaktadır.

Anahtar Kelimeler *– dudak okuma, çoklu sınıf sınıflandırma, Türkçe veri kümesi, derin öğrenme, konuşma tanıma*

Atıf: Pervan Akman, N., Tümer Sivri, T., Berkol A., Erdem H., (2024). Türkçe Günlük Kelime ve İfadeler Kullanarak CNN ve LSTM ile Görsel Konuşma Tanıma. International Journal of Multidisciplinary Studies and Innovative Technologies, 6(2): 69-75.

Visual Speech Recognition Using CNN and LSTM with Turkish Daily Words and Phrases

Abstract – Contemplating a speaker's face to evaluate speech patterns, movements, gestures, and expressions can be described as lip reading. Gaining the ability to lip reading to computers is a growing research area and has open problems for classification and pattern recognition in deep learning. In the last years, various methods have been developed and applied in different languages to classify and convert speech to text. Moreover, most methods have combined multi-model data, i.e., speech and image. This study aims to provide new Turkish lip-reading data with only images and provide a high-accuracy classification method for Turkish daily words. Data was collected from the YouTube platform. Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models were trained to classify daily words and phrases with this challenging data. According to numerous experiment results, the CNN model worked better. Using only images, not multi-modal data, prevents the memory from fatigue and decreases the computation time. Furthermore, we provide a multiclass dataset in Turkish since there is a narrow variety in the literature.

Keywords *– lip reading, multiclass classification, Turkish dataset, deep learning, speech recognition*

Citation: Pervan Akman, N., Tümer Sivri, T., Berkol A., Erdem H., (2024). Visual Speech Recognition Using CNN and LSTM with Turkish Daily Words and Phrases. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 69-75

I. GİRİŞ

Beyin-bilgisayar arayüzleri (BBA), kullanıcı ve bilgisayar arasındaki yazılım bileşenlerine odaklanarak en işlevsel tasarım ve teknoloji uygulamalarını geliştirmeyi amaçlayan bir araştırma alanıdır. İnsan beyni ve bilgisayarlar, sinyaller aracılığıyla görsel örüntüleri yakalayabilir, öğrenebilir ve bu örüntüleri önceki deneyimlere dayalı olarak anlamlı sonuçlar çıkarmak için işleyebilir. Görsel konuşma tanıma, diğer adıyla dudak okuma, ses ve görsellerin BBA sistemlerinde veri olarak

kullanıldığı popüler bir araştırma alanıdır. Bir kişinin ne söylediğini sadece ağız hareketlerine bakarak anlamak, insanlar için oldukça karmaşıktır [1]. Dahası, insanların dudak okuma performansı oldukça düşüktür. Örneğin, işitme engelli ve işitme güçlüğü çeken yetişkinler, 30 tek heceli kelimeden oluşan küçük bir alt küme için sadece %17±12 ve 30 karmaşık kelime için %21±11 doğruluk oranına ulaşmaktadır [2]. Ayrıca, dudak okumanın verimli bir şekilde gerçekleştirilebilmesi için önemli olan bir başka konu da konuşmacılar arasındaki mesafedir. Tejedor'e göre [3], önerilen mesafe 50 santimetre ile 3 metre arasındadır.

Bazı çalışmalar [6], [9], çok modlu verilerle dudak okuma üzerine yoğunlaşmaktadır. Çok modlu verilerle çalışmanın avantajları olsa da, önemli dezavantajları da bulunmaktadır. Özellikle birçok kişinin bulunduğu kalabalık günlük yaşam ortamlarından gelen ses kaynakları söz konusu olduğunda, veriden gürültüyü ayırmak zor bir problemdir. Ses verisini devre dışı bırakmak, günlük yaşam uygulamalarında dudak okuma için daha doğru modellerin geliştirilmesine yardımcı olacaktır. Ayrıca, hem görsel hem de ses verilerini kullanmak, aşırı veri kullanımı ve daha uzun eğitim süresi gerektirir. Derin öğrenme modellerini eğitirken bellek kullanımını dikkate almak önemlidir.

Ses-görüntü tabanlı dudak okuma, dikkate değer derecede iyi sonuçlar göstermiş olsa da, yalnızca görüntü tabanlı dudak okuma da etkinliğini kanıtlamıştır [10], [12], [17]. Tüm derin öğrenme uygulamaları gibi, bunun da bazı zorlukları bulunmaktadır. Yalnızca görüntü verisi içerdiğinden, benzer dudak hareketlerine sahip sesleri ayırt etmedeki zorluklar önemli bir problemdir. Ayrıca, birden fazla kişi varsa, algoritma çoğu uygulamada yalnızca bir kişinin verisini işleyebildiği için, kimin konuştuğunu ve algoritmanın kimi dikkate alacağını ayırt etmek gerçek dünya uygulamalarında zor olacaktır. Ancak, yukarıda belirttiğimiz gibi, görüntülerdeki kişilerin bilgilerini ayırmak ses verilerine göre nispeten daha kolaydır. Ayrıca, gerçek dünya problemlerinde, beyaz gürültüyü iptal etmek bir diğer önemli sorundur. Benzer şekilde, bu durum ses için nispeten zordur.

Bu çalışmada, sınıflandırma başarı oranını artırmak amacıyla yalnızca görüntü tabanlı dudak okuma modeli sunulmakta ve Ural-Altay dil ailesinin bir parçası olan Türkçe için yeni bir görüntü dudak okuma verisi literatüre kazandırılmaktadır. İlerleyen bölümlerde, veri ön işleme aşamalarına, Evrişimli Sinir Ağı ve Uzun Kısa-Süreli Bellek kullanılarak yapılan modelleme deneyine ilişkin problemleri ele alıyoruz.

Çalışmanın literatüre katkıları 1) Yeni bir dudak okuma verisi ile CNN ve LSTM gibi sık kullanılan yaklaşımların ele alınması, 2) Ural-Altay dil ailesinin bir parçası olan Türkçe görüntü veri kümesinin literatüre kazandırılması, şeklindedir.

II. LİTERATÜR ARAŞTIRMASI

Son yıllarda, dudak okuma problemi, sadece engelli bireyler ve onların yakınları değil, aynı zamanda yapay zeka araştırmacılarının da ilgisini çekmektedir. Bu bağlamda, ilk olarak ana diller üzerine gerçekleştirilmiş birçok çalışmadan bahsedilebilir [11], [7]. Daha genel uygulamalar geliştirmek için literatürdeki dil çeşitliliğini genişletmek çok önemlidir. İkinci olarak, veri türleri ve dillere göre son teknolojiye sahip birçok ileri düzey makale bulunmaktadır. Daha iyi doğruluk için Haar Feature-Based Cascade sınıflandırıcı ve CNN ağı kullanılmıştır [4]. Doğruluğu artırmak için geniş bir çalışma yelpazesi bulunmaktadır. Chitu ve Rothkrantz [7] ağız ve açıklığın yükseklikleri, genişlikleri ve alanları gibi görsel özelliklerin geometrik bilgilerini vurgulamışlardır. Tanıma problemi için Hidden Markov Model (HMM) kullanmışlardır. Articulated Feature Extraction yöntemlerinin kullanıldığı başka bir uygulamada da kısa cümlelerin tanınması için Dynamic Bayesian ağı ve sınıflandırma için Destek Vektör Makinesi (Support Vector Machine – SVM) kullanılmıştır [8]. Yan yüzeyden geometrik bilgi kullanan başka bir uygulama da HMM kullanmıştır [9]. Üst ve alt dudak konumlarından çıkarılan iki çizgi arasındaki açı, Lip Contour Features (LCGFs) olarak adlandırılmıştır. Yazarlar, dudak alanını tespit eder, dudakların merkez noktasını çıkarır ve LCGF adımları olarak dudak çizgilerini ve dudak açısını belirler.

Fenghour vd. [10], sinir ağları ve özellik çıkarma üzerine odaklanan çeşitli yöntemleri karşılaştırmak için iyi bir derlemedir. Yazarların en önemli çıkarımı Dikkat Dönüştürücülerinin (Attention-Transformers) ve Zamansal Evrişim Ağlarının (Temporal Convolutional Networks) Tekrarlayan Sinir Ağlarına (Recurrent Neural Networks – RNN) karşı avantajlarıdır. Çalışmada hem görsel-işitsel verilere hem de yalnızca görsel verilere odaklanmışlardır. Ayrıca, harf tabanlı, kelime tabanlı ve cümle tabanlı yöntemlerin İngilizce, Arapça, Çince ve Almanca gibi çeşitli dilleri kapsadığını belirtmişlerdir. Ozcan ve Basturk [5] AvLetters veri kümesinde AlexNet ve GoogleNet'in önceden eğitilmiş CNN modelini kullanmışlardır. Çalışmada veri boyutunu artırmak için veri artırma teknikleri kullanılmıştır. Makalede kullanılan teknikler, "gaussian", "salt and pepper" ve "speckle" ile gürültü ekleme, "unsharp" ile keskinleştirme ve "median" filtreleme ile yumuşatma şeklindedir. Lu ve Li [13] rakamların sınıflandırılması için yeni bir ağ önermişlerdir. Veriler, 3 kadın ve 3 erkek konuşmacının 100 defaya kadar telaffuz ettiği 0'dan 9'a kadar sayıları içermektedir. Uzamsal özellikleri çıkarmak için VGG19 ağı kullanılırken, zaman özelliklerini çıkarmak için Dikkat Tabanlı (Attention Based) LSTM kullanılmıştır.

Zamansal Konvolüsyonel Ağlar, LSTM'ye bir alternatiftir [14]. Martinez vd. [15] kelime düzeyinde sınıflandırma için Çok Ölçekli Zamansal Evrişim (Multi-Scale Temporal Convolution) yöntemini sunmuşlardır. Yalnızca işitsel, görsel-işitsel ve yalnızca görsel veriler üzerinde deneyler yapmışlardır. Amit vd. [16] sınıflandırma için CNN ve LSTM kullanılmış ve IMDB ve Google Görseller'den aldıkları ünlü insan yüzleri üzerinde önceden eğitilmiş VGGNet'i uygulamışlardır. Görüntüleri birleştirmek ve LSTM'den zamansal bilgi çıkarma işlemi, onların katkısı olmuştur.

Chung vd. [6] ağız hareketlerinin videolarını karakterlere dönüştürmeyi öğrenmek için Watch, Listen, Attend and Spell (WLAS) ağı geliştirilmiştir. Yalnızca görseller için çalışan WAS, WLAS modelinin bir parçasıdır. Ayrıca, eğitim süresini azaltmak ve aşırı uyumu önlemek için bir "curriculum learning strategy" önermişlerdir. Ek olarak İngiliz televizyonundan alınan 100.000'den fazla doğal cümle içeren Lip Reading Sentences (LRS) veri kümesi, görsel konuşma tanıma uygulamaları için yayınlanmıştır. LipNet [17], uçtan uca cümle ve ifade düzeyinde tahminler yapmak üzere geliştirilmiş ve eğitilmiştir. Model, karakter düzeyinde çalışmakta olup, uzay - zamansal CNN'ler, RNN'ler ve bağlantısal zamansal sınıflandırma (Connectionist Temporal Classification – CTC) kaybını kullanmaktadır [18]. Yazarlar, halka açık cümle düzeyinde bir veri kümesi olan GRID corpus üzerinde deneyler yapmışlardır [19].

LipType [12] ileri hız ve doğruluk için geliştirilmiş bir diğer modeldir. Yazarlar ayrıca zayıf ışık koşulları altında model sonuçlarının iyileştirilmesine de katkıda bulunmuşlardır. İlk aşama olarak yüz hatlarının Kalman Filtrelemesi, 3D-CNN ve 2D SE-ResNet ile düzeltilmesini içeren uzay-zamansal özellik çıkarma yöntemi (spatiotemporal feature extraction method) kullanılmış olup daha sonra CTC'li Çift Yönlü Geçitli Tekrarlayan Sinir Ağları (Bidirectional Gated Recurrent Neural Networks with CTC) uygulanmıştır.

Jittakoti ve Phumeechanya [23] CNN ve LSTM kullanarak Temporal Keyframe tekniği yoluyla dudak okuma performansını iyileştirmeye yönelik bir yöntem sunmaktır. Çalışmada kullanılan veri seti, tam ve yarım dudak görüntü verilerini içermektedir ve çalışma sırasında 3 kare, 5 kare ve 10 kare olmak üzere 3 gruba ayrılmıştır. Görülmemiş test seti değerlendirildiğinde, 10 kare, tam dudak görüntü veri seti için %87.9 doğruluk ve yarım dudak görüntü veri seti için %86.8 doğruluk ile en iyi tanıma oranını sağlamış ve karşılaştırılabilir bir performans sergilemiştir.

Shashidhar vd. [24] MIRACL VC1 veri seti için görsel konuşma tanıma amacıyla LSTM ve 3D CNN hibrit modeli önermiştir. Önerilen çalışma, hibrit modeli içermesiyle önceki çalışmalardan ayrılmaktadır. Sadece 3D CNN modelinin test doğruluğu %79 ve LSTM modelinin doğruluğu %85 iken, hibrit modelin eğitim, test ve doğrulama setlerinin doğruluğu sırasıyla %98, %85 ve %86 olmuştur.

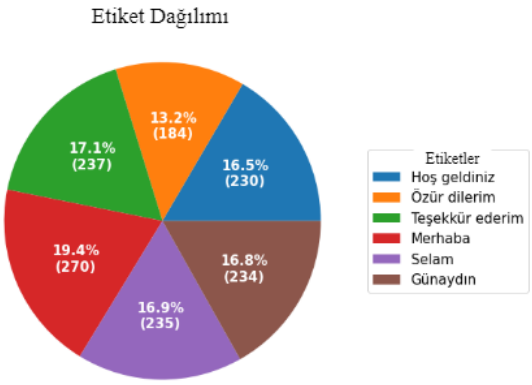
Pourmousa ve Özen [25] evrişimli sinir ağları kullanılarak Türkçe dilinde dudak okuma işlemi gerçekleştirilmiştir. Türkçe 20 sayının söylendiği video verileri kullanılmıştır. Bu videolar, sayıların dudak hareketlerini içerir. Veri seti, dudak hareketlerinin ve yüz ifadelerinin detaylı bir şekilde gösterildiği videoları kapsar. Bu kapsamda kişilerden sayıların videosunu (61 video) çekip göndermeleri istenmiş ve onun yanı sıra YouTube'tan 9 video toplanmıştır. CNN tabanlı yöntem, dudak okuma sistemlerinde yüksek doğruluk sağlar ve Türkçe sayıların tanınmasını geliştirir. Türkçe sayıların dudak hareketlerinden tanınması için CNN tabanlı yaklaşımlar üzerine yenilikçi bir bakış açısı sunar.

Exarchos vd. **Error! Reference source not found.** 3D Evrişimli Sinir Ağları ve Uzun Kısa Vadeli Hafıza ağlarını birleştiren yenilikçi bir yaklaşım önerilmektedir. Araştırmada; farklı konuşma kalıplarını, konuşmacıları ve çevresel koşulları kapsayan titizlikle oluşturulmuş "MobLip" adlı bir veri setinden yararlanmaktadır. Çalışmada kullanılan veri seti, dudak hareketlerini içeren video verileri içermektedir. 3D CNN'ler tarafından çıkarılan mekansal bilgiler ile LSTM'ler tarafından yakalanan zamansal dinamikler arasındaki sinerji, aydınlatma değişikliklerine ve konuşmacı çeşitliliğine karşı dayanıklılık sergileyerek %87,5'e varan bir doğruluk oranına ulaşarak etkileyici sonuçlar vermektedir.

III. Materyal ve Metot

A. *Veri Toplama*

Dudak okuma uygulamalarına yönelik literatürde, İngilizce, Almanca gibi çeşitli dillerde yalnızca görüntü, yalnızca ses ve ses-görüntü verileri için çok sayıda çalışma bulunmaktadır. Bu çalışmada Türkçe için yeni bir kelime düzeyinde ve ifade düzeyinde çok sınıflı bir veri kümesi [20] öneriyoruz. Veri kümesi, "selam", "merhaba", "günaydın" olmak üzere üç kelime sınıfı ve "hoş geldiniz", "özür dilerim", "teşekkür ederim" olmak üzere üç kelime öbeği ile toplam altı sınıf içermektedir. Her sınıf yaklaşık olarak aynı sayıda veri içermektedir; bkz. Şekil 1. Veriler Youtube platformu üzerinden oluşturulmuştur. İlgili kelimeler söylenirken kısa videolar kaydedilerek daha sonra, kelimenin başladığı ve bittiği dudak hareketlerine göre çerçeveleri belirlenmiştir. Veri kümesini toplarken konuşmacı sayısı, konuşmacı ile kamera arasındaki mesafe ve dudaklar ile kamera arasındaki açı gibi açılardan veri çeşitliliğine özellikle dikkat edilmiştir. Sentetik verilerle eğitilen modeller iyi tahmin ve sınıflandırma sonuçlarına sahip olsa da, eğitilen modelde kullanılan verilerin toplandığı ortam fazlasıyla kontrollüdür. Bu nedenle mümkün olduğunca gerçek dünyaya benzetmeye çalışılmıştır. Her sınıfın çerçeve sayılarına ilişkin dağılımlarına bakmak önemlidir, çünkü bu hem konuşmacının konuşma hızı hem de kelimenin telaffuz edilme uzunluğu göz önüne alındığında, çerçeve sayıları açısından zaman zaman dengeli olabilirken zaman zaman çerçeve sayıları farklılaşmaktadır. Her sözcükteki çerçeve sayıları önemlidir ve model çerçeve sayılarını etkileyebileceğinden sözcüğün ne kadar hızlı telaffuz edildiğine bağlıdır. Şekil 2'de bazı sınıfların dengeli veri sağlayan normal dağılımlara sahip olduğu görülmektedir. "selam", "özür dilerim" ve "tüm sınıflar" sağa eğik dağılımlardır, bkz. Şekil 2e,2d,2g.



Şekil. 1. Sınıfların Etiket Dağılımı

Yani verilerin çoğu ortalamanın altında çerçeve sayısına sahiptir. Diğerleri tahmin edilebilir normal dağılımlardır.

B. *Veri Ön İşleme*

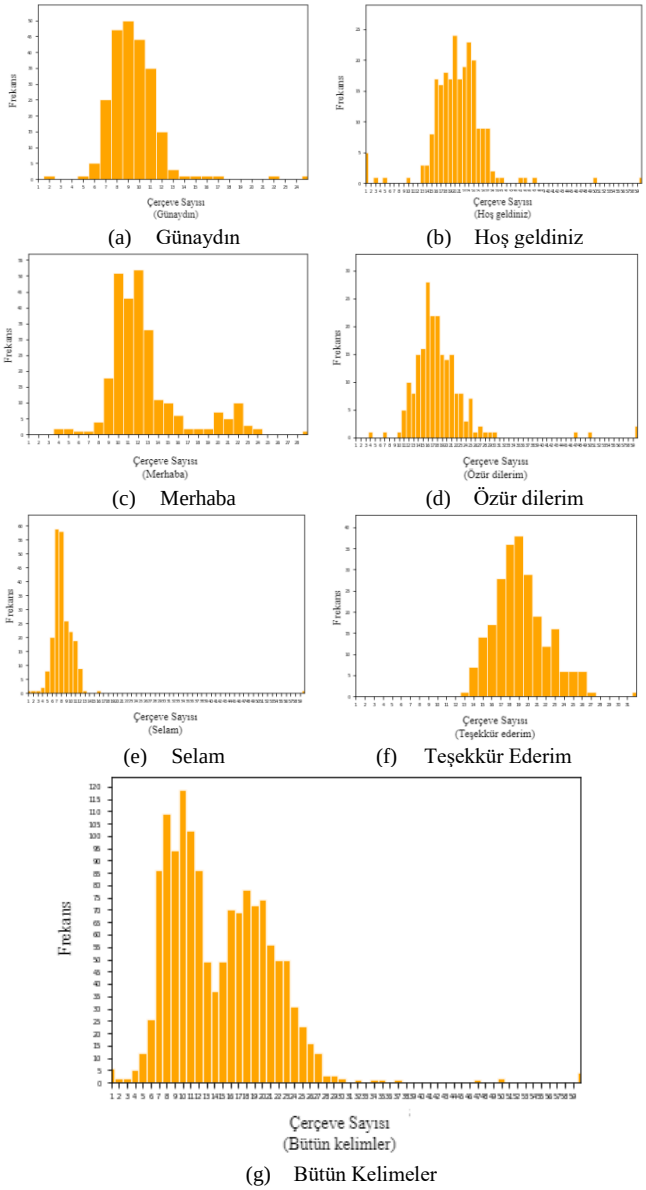
Toplanan veriler oldukça ham görüntüler olduğundan, bazı görüntü düzenlemelerine ihtiyaç duyulmuştur. İlk ve en temel işlem, görüntülerin gri tonlamaya dönüştürülmesidir. Görseldeki kişinin gözleri, burnu veya görüntüdeki diğer kısımları dudak okuma problemi için gerekli olmadığından, öncelikle yüzü, ardından dudak bölümleri kesilerek modele dahil edilir. Daha sonra, OpenCV [21] ve dlib [22] kütüphaneleri kullanılarak görseldeki yüz ve ağız kesimi için yüz işaret noktaları tespit edilir (Bkz. Şekil 3). Son olarak her bir dudak görüntüsü sabit bir boyut olarak yeniden boyutlandırılır. Bu boyut, hesaplama maliyetini azaltmak amacıyla mümkün olduğunca küçük tutulur.

Her görüntü için yapılan ön işlemeye ek olarak, her örnek için kare sayısı sabit bir değere oturtulmuştur. Bazı kelimelerin video sekanslarındaki kare sayıları, her konuşmacının konuşma hızına ve kelimenin uzunluğuna göre değişiklik gösterdiğinden, tutarlılığı sağlamak amacıyla sabit bir boyuta getirilmiştir. Yapılan deneylerde, kelime uzunluklarına göre en iyi sonuçlar 15 değeri için elde edilmiştir. Eğer örneklerin

çerçeve sayısı 15'ten büyükse devam eden çerçeveler göz ardı edilir; 15'ten küçükse, boş çerçeveler ile doldurulmaktadır.

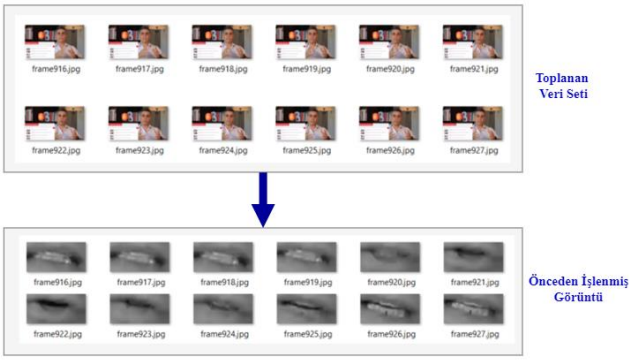
C. *Modeller*

Dudak okuma sınıflandırma problemi üzerinde CNN ve LSTM kullanarak çalışmalar gerçekleştirilmiştir. CNN, görüntü işleme problemlerinde en sık kullanılan yaklaşımlardan biridir. CNN mimarisi, ReLU aktivasyon fonksiyonunu kullanan iki evrişim katmanı ve ardından gelen evrişimlerde iki maksimum havuzlama (max pooling) katmanı içermektedir. Evrişim katmanında pencere boyutu 3, filtrelerin adım sayısı ise 2 olarak belirlenmiştir. Sınıflandırma kısmı, tam bağlantı (fully connected) katmanları ve aşırı uyumu (over-fitting) önlemek için kullanılan bırakma (dropout) katmanını içermektedir (Bkz. Şekil 4). Son olarak softmax katmanı, Türkçe'deki üç kelime ve üç kelime öbeği için bir skor döndürür.

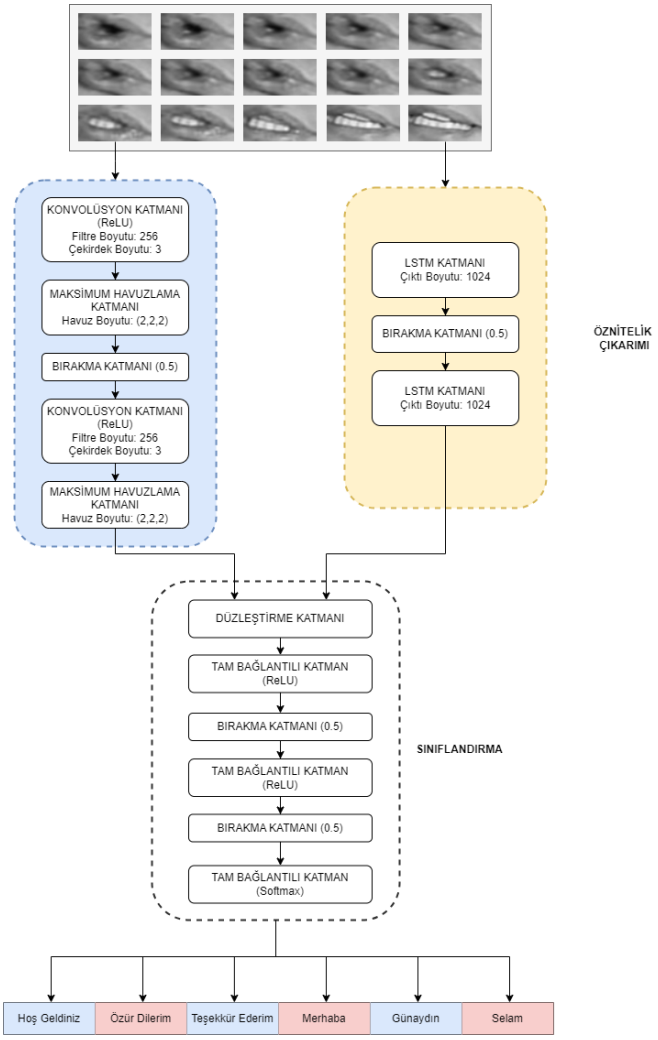


Şekil 2. Tüm Sınıflara Ait Çerçeve Sayıları

Önerilen derin öğrenme modelleri, sınıflandırma katmanında aynı mimariyi kullanırken, özellik çıkarım katmanında CNN ve LSTM kullanılarak oluşturulmuştur (Bkz. Şekil 4). Kullandığımız bir diğer derin öğrenme mimarisi de LSTM'dir. LSTM, zaman serisi yaklaşımlarında genel olarak çok iyi sonuçlar üretmektedir. Yapılan çalışmada dudak görselleri bir sekans oluşturduğundan LSTM'deki girdi, unutma ve çıktı kapıları sayesinde probleme uygun bir algoritmadır. Dudak okuma problemlerine zaman serisi yaklaşımı uygulandığından, teorik olarak LSTM kullanmak mantıklıdır. Temel LSTM modelimiz, 1024 çıktı boyutuna sahip iki LSTM katmanından oluşmaktadır.



Şekil 3. Uçtan Uca Veri Ön İşleme Örneği



Şekil 4. Model Mimarisi: Diyagramın mavi kısmı CNN modelinin özellik çıkarma katmanını; sarı kısım LSTM modelinin özellik çıkarma katmanını göstermektedir.

D. *Eğitim*

Eğitim süreci boyunca birçok değer için hiperparametre ayarlaması (hyperparameter tuning) deneysel olarak gerçekleştirilmiştir (bkz. Tablo 1). Deneysel çalışmalar, her bir örnekteki dudak görüntüleri üzerinde CNN katmanlarının filtre boyutu, LSTM çıktı katmanları ve boyutları, öğrenme oranı, giriş boyutu ve eğitime dahil edilecek çerçeve sayısı için farklı değerleri tekrar etmiştir. Ayrıca doğrulama kaybı değerinin iyileştirilmemesi durumunda erken durdurma stratejisinin eklendiği eğitimi de deneye dahil ettik. 1390 örnek, eğitim ve model testi için %70 eğitim, %15 test ve %15 doğrulama olarak bölünmüştür.

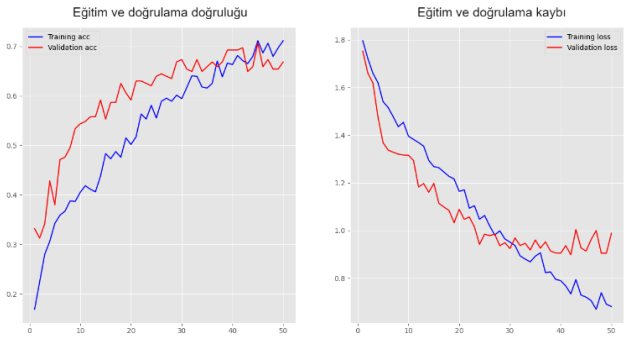
Tablo 1 CNN ve LSTM modelleri hiperparametreleri

	CNN	**LSTM**
Epok	50	65
Yığın Boyutu (Batch Size)	16	4
Öğrenme Oranı (Learning Rate)	2e-4	2e-4
Katman Sayısı	2	2
Filtre Boyutu	3x3	-
Bırakma katmanı	0.5	0.5
Aktivasyon Fonksiyonu	ReLU	ReLU
Optimizasyon Fonksiyonu	Adam	Adam
Kayıp Fonksiyonu (Loss Function)	Kategorik Çapraz Entropi (Categorical Cross Entropy)	Kategorik Çapraz Entropi (Categorical Cross Entropy)
Kelime Boyutu	15	15
Girdi Boyutu	50	50

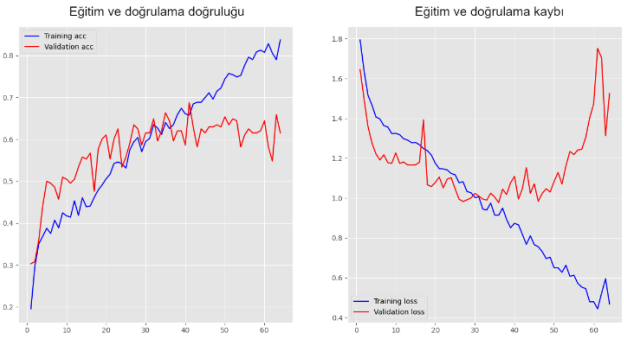
Çalışmalar Python dilinde CUDA 11.2 kullanılarak gerçekleştirilmiştir. Ekran kartı olarak NVIDIA GeForce GTX 1650 Ti 4GB kullanılmıştır.

IV. SONUÇLAR

Erken durdurma (early stopping) stratejisine sahip eğitilmiş CNN ve LSTM modelleri, CNN için 50 epok ve LSTM için 65 epok sonunda elde edilmiştir. Doğruluk ve kayıp grafiği eğitim sürecinden elde edilir. Şekil 5 ve 6'da görülebileceği gibi eğitimin daha fazla devam etmesi durumunda model verilerden daha fazla şey öğrenecektir.



Şekil 5. CNN Modelinin Eğitim ve Doğrulama Kaybı ve Doğruluğu



Şekil 6. LSTM Modelinin Eğitim ve Doğrulama Kaybı ve Doğruluğu

CNN ve LSTM modellerini, veri çeşitliliğini göz ardı etmemek adına, duyarlılık (recall), kesinlik (precision), f1-skor (f1-score) ve doğruluk (accuracy) performans metriklerini kullanarak değerlendirdik ve karşılaştırdık. Mikro doğruluk, modelin toplamda ne kadar doğru tahminde bulunduğunu ölçen bir metriktir. Mikro kesinlik, modelin pozitif olarak tahminlediği örneklerin ne kadarının gerçekten pozitif olduğunu ölçerken, duyarlılık modelin gerçek pozitifleri ne kadar iyi yakaladığını ölçer. Yani, gerçek pozitiflerin doğru tahminlere oranıdır. F1 skor ise, kesinlik ve duyarlılığın harmonik ortalamasıdır, yani iki metrik arasındaki dengeyi sağlar.

$$Mikro\ Doğruluk = \frac{Toplam\ doğru\ tahimler}{Toplam\ örnekler} \quad (1)$$

$$Mikro\ Kesinlik = \frac{Toplam\ gerçek\ pozitifler}{Toplam\ gerçek\ pozitifler + Toplam\ yanlış\ pozitifler} \quad (2)$$

$$Mikro\ Duyarlılık = \frac{Toplam\ gerçek\ pozitifler}{Toplam\ gerçek\ pozitifler + Toplam\ yanlış\ negatifler} \quad (3)$$

$$\text{Mikro F1} - \text{skor} = 2 \text{ x } \frac{\text{Mikro kesinlik x Mikro duyarlılık}}{\text{Mikro kesinlik+Mikro duyarlılık}} \quad (4)$$

Dudak okuma modellerinin 6 sınıfından elde ettiğimiz test doğrulukları (mikro), CNN ve LSTM için sırasıyla %60 ve %56 (Bkz. Tablo 2 ve Tablo 3). Ancak bazı kelimeler için tespit performansı genel doğruluktan daha iyi olurken bazı kelimeler için bu skor daha düşüktür. Örneğin, "özür dilerim" sınıfı için kesinlik skoru, CNN modeli için %74 ve LSTM modeli için %82 olup, bu değerler genel doğruluktan daha yüksektir. Buna karşın, "özür dilerim" sınıfı için duyarlılık değeri, CNN modeli için %59 ve LSTM modeli için %28 olup, bu değerler genel doğruluktan daha düşüktür. Ayrıca, bazı sınıflar için CNN ve LSTM modeli sonuçlarında skorun daha düşük ve daha yüksek olduğu durumlar mevcuttur. Örneğin, "özür dilerim" sınıfı için doğruluk, LSTM modeli için genel doğruluğa düşük iken, CNN modeli için genel doğruluğa denktir. Bu durum, farklı YouTube videolarından elde edilen veri kümesinin çeşitliliğinden ve sınıf örnek setinin farklı olmasından kaynaklanmaktadır. CNN ve LSTM modelleri toplamda 209 örnek ile test edilmiştir.

Tablo 1. CNN Model Sonuçları

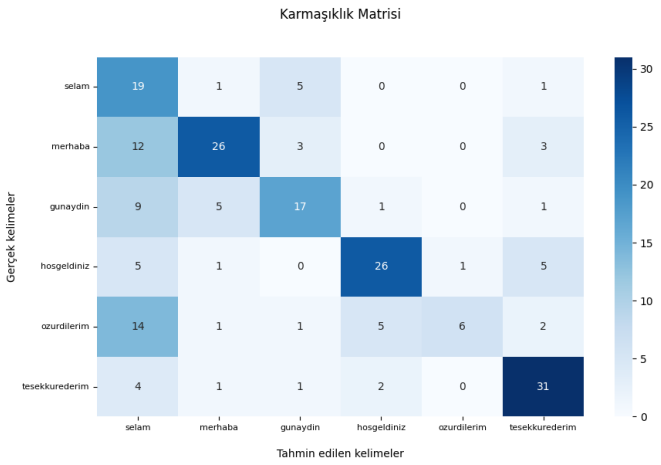
Kelime	Doğruluk	Kesinlik	Duyarlılık	F1-Skor	Boyut
Hoş geldiniz	0,73	0,30	0,73	0,43	26
Özür dilerim	0,59	0,74	0,59	0,66	44
Teşekkür ederim	0,51	0,63	0,52	0,57	33
Merhaba	0,68	0,76	0,68	0,72	38
Selam	0,20	0,86	0,21	0,33	29
Günaydın	0,79	0,72	0,79	0,76	39
Tüm Kelimeler	0,60	0,69	0,60	0,60	209

Tablo 2. LSTM Model Sonuçları

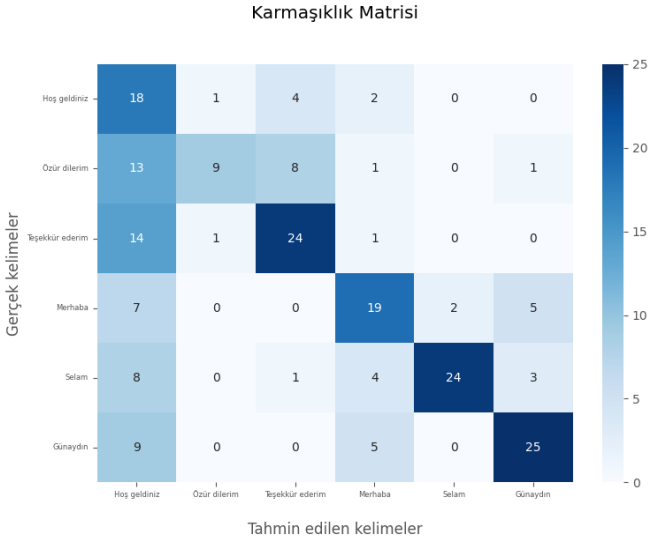
Kelime	Doğruluk	Kesinlik	Duyarlılık	F1-Skor	Boyut
Hoş geldiniz	0,72	0,26	0,72	0,38	25
Özür dilerim	0,28	0,82	0,28	0,42	32
Teşekkür ederim	0,60	0,65	0,60	0,62	40
Merhaba	0,57	0,59	0,58	0,58	33
Selam	0,60	0,92	0,60	0,73	40
Günaydın	0,64	0,74	0,64	0,68	39
Tüm Kelimeler	0,56	0,69	0,57	0,59	209

Şekil 7 ve 8'de her kelime için yapılan tahminler ve yanlış tahminle sonuçlanan kelimeler gösterilmektedir. Köşegendeki yoğunluğa bakıldığında çoğunlukla her kelime için iyi bir performansın olduğu görülmektedir. Yanlış tahmin edilen "günaydın" kelimesi için "merhaba" ve "selam" kelimelerine odaklanılmaktadır. "özür dilerim" ve "teşekkür ederim" yanlış tahmin edildiğinde, doğru tahminin "hoş geldiniz" olması gerektiği belirtilmiştir. Bu iki duruma göre yorum yapılırsa, kelimelerin ve ifadelerin kendi içlerinde tahmin karışıklığına neden olduğu söylenebilir.

Şekil 7. CNN Modelinin Karmaşıklık Matrisi



Şekil 8. LSTM Modelinin Karmaşıklık Matrisi



V. TARTIŞMA

Bu çalışmada, yeni bir Türkçe dudak okuma veri kümesinin kazandırılmasının yanı sıra bu veri ile yapılan CNN ve LSTM modellerinin değerlendirilmesi yapılmıştır. Gerçek dünyaya uyan bir model geliştirmeye çalışıldığından, bu veri kümesinin oluşturulmasındaki en önemli kısım, tamamen gerçek dünya verilerinden elde edilmesi oldu. Veri kümesi hem ön işleme hem de eğitim açısından zorlu olsa da, CNN modelinin LSTM'den daha iyi olduğu çok sınıflı sınıflandırma problemlerinde iyi bir sonuç elde edilmiştir. Ayrıca kelime ve cümleleri kendi aralarında sınıflandırmada yanlış pozitif ve yanlış negatif değerlerin daha yaygın olduğunu gördük ki bunun da en doğal sonuçlardan biri olduğunu düşünüyoruz.

Çalışma kapsamında yapılan değerlendirmeler için her bir kelime sınıfı test edilirken daha fazla veriye sahip olunması durumunda daha iyi sonuçlara ulaşması beklenmektedir.

VI. GELECEK ÇALIŞMALAR

Gelecek çalışmalarda, farklı ön işleme stratejileri uygulamayı ve yüz ve ağız bölgelerinin kesilmesi için yeni algoritmalar geliştirmeyi planlıyoruz. Böylece, yeni hibrit modellerle sınıflandırma skorunu artırmayı hedefliyoruz. Ayrıca, 10 kelime ve kelime öbeği sınıfı oluşturduk. Ancak, model yalnızca bunların 6'sı ile test edilmiştir. Gelecekteki diğer bir çalışma ise genişletilmiş veri kümesi ile yeni modeller denemek olacaktır.

KAYNAKLAR

[1] C. G. Fisher. “Confusions among visually perceived consonants.” Journal of Speech, Language, and Hearing Research, 11(4) pp. 796–804, Dec. 1968.

[2] R. D. Easton and M. Basala. “Perceptual dominance during lipreading”. Perception and Psychophysics, 32(6) pp.562–570, Nov. 1982.

[3] Cecilia Tejedor, A. Leer en los labios. Manual práctico para entrenamiento de la comprensión labiolectora. Madrid: CEPE, 2000.

[4] Shrestha, K. (n.d.). “Lip Reading using Neural Network and Deep learning.” 1802.

[5] T. Ozcan, and A. Basturk, “Lip Reading Using Convolutional Neural Networks with and without Pre-Trained Models.” Balkan Journal of Electrical and Computer Engineering, vol. 7(2) pp. 195-201, Apr. 2019.

[6] J. S. Chung, A. Senior, O. Vinyals, and A. Zisserman, “Lip reading sentences in the wild.” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6447-6456.

[7] Chitu, A., Rothkrantz, L. "Visual Speech Recognition Automatic System for Lip Reading of Dutch". Journal on Information Technologies and Control, vol. 7, no. 3, pp. 2-9, 2009.

[8] K. Saenko, K. Livescu, M. Siracusa, K. Wilson, J. Glass, T. Darrell "Visual Speech Recognition with Loosely Synchronized Feature Streams," in Proceedings of the 10th International Conference on Computer Vision, 2005, pp.1424–1431.

[9] K. Iwano, T. Yoshinaga, S. Tamura, S. Furui. "Audio-Visual Speech Recognition Using Lip Information Extracted from Side-Face Images", Hindawi Publishing Corporation EURASIP Journal on Audio, Speech, and Music Processing vol. 2007, pp.1-9, 2007

[10] S. Fenghour, D. Chen, K. Guo, B. Li, and P. Xiao, "Deep learning-based automated lip-reading: A survey," IEEE Access, vol. 9, pp. 121184–121205, 2021.

[11] M. Faisal, and S. Manzoor, "Deep Learning for Lip Reading using Audio-Visual Information for Urdu Language". CoRR, 2018. DOI: https://doi.org/10.48550/arXiv.1802.05521

[12] L. Pandey and A. S. Arif, "LipType: A Silent Speech Recognizer Augmented with an Independent Repair Model." in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, 2021 Article 1, pp. 1–19. DOI: https://doi.org/10.1145/3411764.3445565

[13] Y. Lu and H. Li, "Automatic Lip-Reading System Based on Deep Convolutional Neural Network and Attention-Based Long Short-Term Memory". Applied Sciences. 9(8) 1599. 2019. DOI: https://doi.org/10.3390/app9081599

[14] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," arXiv:1803.01271, 2018.

[15] B. Martinez, P. Ma, S. Petridis, and M. Pantic, "Lipreading using temporal convolutional networks." in ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2020. pp. 6319-6323 https://doi.org/10.1109/icassp40776.2020.9053841

[16] G. Amit, J. Noyola, and S. Bagadia. "Lip reading using CNN and LSTM". Stanford University, CS231n project report, 2016.

[17] Y. M. Assael, B. Shillingford, S. Whiteson, and N. de Freitas. "Lipnet: End-to-End Sentence-Level Lipreading." Dec. 2016. http://arxiv.org/abs/1611.01599

[18] A. Graves, S. Fernandez, F. Gomez, and J. Schmidhuber. "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks." in ICML, 2006 pp. 369–376.

[19] M. Cooke, J. Barker, S. Cunningham, and X. Shao. "An audio-visual corpus for speech perception and automatic speech recognition." The Journal of the Acoustical Society of America, vol. 120(5) pp. 2421–2424, 2006. DOI: https://doi.org/10.1121/1.22290.

[20] https://doi.org/10.17632/4t8vs4dr4v.1.

[21] OpenCV Team. (2024). *OpenCV: Open Source Computer Vision Library*. Version 4.6.0. https://opencv.org/

[22] King, D. E. (2009). *dlib: A C++ Library for Machine Learning*. Version 19.24. http://dlib.net/

[23] Jittakoti, A., & Phumeechanya, S. (2024, March). Temporal Keyframe Technique based on CNN and LSTM for Enhancing Lip Reading Performance. In 2024 12th International Electrical Engineering Congress (iEECON) (pp. 1-5). IEEE.

[24] Shashidhar, R., Shashank, M. P., & Sahana, B. (2023). Enhancing visual speech recognition for deaf individuals: a hybrid LSTM and CNN 3D model for improved accuracy. Arabian Journal for Science and Engineering, 1-17.

[25] Pourmousa, H., & Özen, Ü. (2022). LIP READING USING CNN FOR TURKISH NUMBERS. Journal of Business in The Digital Age, 5(2), 155-160.

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 76-80
DOI: 10.36287/ijmsit.8.2.5
Received: November 11, 2024; Accepted: December 9, 2024

A Prototype Study on YOLOv10-Based Bird Gesture Recognition

Rıdvan Yayla*

[1*]*Computer Engineering Department, Engineering Faculty Bilecik Şeyh Edebali University, Bilecik, Türkiye (ridvan.yayla@bilecik.edu.tr) (ORCID: 0000-0002-1105-9169)*

Abstract – Birds are one of the most abundant types of creatures on Earth. However, it is also known that many taxonomically diverse bird species exist in nature. The bird network has standard behavioural patterns such as flying, perching, feeding and walking. In this study, 2372 bird images are used for five standard bird gestures detection: flying, perching, swimming, eating, and walking with the Yolov10 algorithm from Caltech-UCSD Birds-200-2011 dataset. Firstly, the dataset is prepared for detection by classifying these gestures. Secondly, the bird gesture images are trained with Yolov10, thirdly the trained model is tested with bird motion short videos and finally, the evaluation results are shown with evaluation metrics. In this prototype study, it was observed that the obtained model had results with an accuracy higher than 70%. The study can be used to make sense of bird communication for future studies.

Keywords *– Bird gesture, target detection, classification, deep learning, Yolov10.*

Citation: Yayla, R., (2024). A Prototype Study on YOLOv10-Based Bird Gesture Recognition. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 76-80

I. INTRODUCTION

Object detection is one of the methods used in all intelligent systems in recent years. It is used for all intelligent systems from home management systems to target detection in military defense systems. In this way, the workload requiring manpower can be reduced. In recent years, a few effective object detection algorithms such as R-CNN, Fast R-CNN, Mask R-CNN and Yolo (You Only Look Once) have been developed for object detection. These methods are widely used for different purposes such as military defence, unmanned vehicles, agriculture, health and commerce.

On the other hand, a bird network is a widely complex structure in nature. According to ornithologists, 9500 to 11000 species of birds live in the world as a taxonomy [1]. In contrast, these animals have standard behaviour types such as flying, feeding and perching. In this scope, Yolov10 that is last version of the Yolo algorithm is used for bird gesture recognition in this study [2].

II. MATERIALS AND METHOD

A. Level Literature Review

In literature, there are various object detection studies with Yolo such as from traffic sign recognition to military target detection. Yolo algorithm is a flexible algorithm that can be used for a wide range of different objects. For bird studies, the Yolo algorithm has generally been used for purposes such as bird species recognition and bird flock detection for safe flight. Liang et al. developed the SMB-YOLOv5 model to detect birds near airports for safe flight in their study [3]. Datar et al. conducted a comparative study for bird detection using YOLOv2, YOLOv3 and Mask R-CNN algorithms [4]. Xie et al. used the yolov5 algorithm to detect the presence of bird nests that threaten transmission lines and transformer substations [5]. Ou et al. developed a system that detects birds and identifies bird species using Yolo and Custom Vision algorithms [6]. Zhao used the yolov4 algorithm to detect bird movements in 3 classes: staying, flying and swimming in a study she conducted in 2022 [7]. Bird gestures are improved by using Yolov10 for 5 movements: perching, swimming, opening wings-flying (ow-flying), eating, and walking in this study.

B. Yolo Algorithm

Yolo is the most common algorithm that uses CNN for real-time object tracking. Applications such as R-CNN, Fast R-CNN, and Faster R-CNN were popular applications used for real-time object tracking until Yolo was released in 2015 [8] [9] [10] [11].

Yolo predicts the coordinates and class of objects in the image by passing the image through a neural network one at a time. While making this definition, it divides the image into grids. There is no specific condition to determine the number of grids. It is sufficient to have it in N*N format. Each grid determines whether there is an object in it and if it detects an object, it estimates whether the center point is in its area. After the image passes through the neural network, a vector is produced as output. The working principle of YOLO is as follows:

- Yolo first divides the image into grids.
- It draws a frame (bounding box) around the objects
- in each region.
- It calculates the probability of finding an object in each region.
- It calculates a confidence score for each frame (bounding box). Confidence score predicts the

probability that an object is that object and is calculated as (1) [12].

$$Confidence = \Pr(obj) \times IoU \quad (1)$$

- In (1), Pr(obj) represents the probability of finding the object in the grid, and Intersection over Union (IoU) represents the intersection of the predicted box with the box where the object is located.

In YOLO, error functions are expressed with three basic error rates:

Confidence loss: It expresses how wrong it is to determine whether there is an object inside the grid. If there is an object in the image, it is calculated as (2) and if not, it is calculated as (3) [14].

$$\sum_{i=0}^{s^2}\sum_{j=0}^{B} 1_{ij}^{obj}\left(C_i - \hat{C}_i\right)^2 \quad (2)$$

$$\lambda_{noobj}\sum_{i=0}^{s^2}\sum_{j=0}^{B} 1_{ij}^{noobj}\left(C_i - \hat{C}_i\right)^2 \quad (3)$$

S is the grid cell number, and B is the number of estimated bounding boxes in each grid cell. 1_{ij}^{obj} indicates whether the j bounding box of the i grid cell contains an object. λ_{noobj} is a weighting parameter used to compensate for the loss of confidence in places where objects are not present. C_i real confidence score i.e. the confidence score in the grid cell of the real object box and $\hat{C}_i$ is the confidence score estimated by the model.

Location loss: It expresses how wrong the predicted box is and it is calculated as (5) [13].

$$\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B} 1_{ij}^{obj}[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B} 1_{ij}^{obj}\left[\left(\sqrt{w_i} - \sqrt{\widehat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2\right] \quad (5)$$

λ_{coord} is a weighting parameter used to compensate for position loss. x_i and y_i indicate the x and y coordinates of the centre of the bounding box. $\hat{x}_i$ and $\hat{y}_i$ are the x and y coordinates of the centre of the bounding box predicted by the model. w_i , h_i represent the width and height of the bounding box, and $\widehat{w}_i$, $\hat{h}_i$ represent the width and height of the bounding box predicted by the model.

Classification loss: It expresses how wrong the predicted object is and is calculated as (4) [13].

$$\sum_{i=0}^{S^2} 1_i^{obj}\sum_{c\,\in\,classes}(p_i(c) - \hat{p}_i(c))^2 \quad (4)$$

$p_i(c)$ is a probability vector of real class labels, an array of 1 for the real object class and 0 for other classes. $\hat{p}_i(c)$ is the vector containing the class probabilities predicted by the model, that is, it represents the probability of the predicted class.

Finally, the total loss function is computed by summing three loss functions and it is shown as (6) [14]. The smaller the total loss function result, the higher the success of the algorithm.

Additionally, Mean Average Precision (mAP) is used to determine the accuracy of object identification by the Yolo algorithm [15]. mAP is a metric used to evaluate the accuracy of an object detection model. It measures the model's performance over a certain threshold value, averaged over all classes Average Precision (AP) is defined as the area under the precision and recall curve for a given class.

$$\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B} 1_{ij}^{obj}[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B} 1_{ij}^{obj}\left[\left(\sqrt{w_i} - \sqrt{\widehat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2\right] + \sum_{i=0}^{s^2}\sum_{j=0}^{B} 1_{ij}^{obj}\left(C_i - \hat{C}_i\right)^2 + \lambda_{noobj}\sum_{i=0}^{s^2}\sum_{j=0}^{B} 1_{ij}^{noobj}\left(C_i - \hat{C}_i\right)^2 + \sum_{i=0}^{S^2} 1_i^{obj}\sum_{c\,\in\,classes}(p_i(c) - \hat{p}_i(c))^2 \quad (6)$$

IoU is the ratio of the intersection of the predicted boundary box and the true boundary box to the union area [12]. IoU is used to measure how accurate the model's detections are. mAP50 is a mAP calculation method where the IoU threshold is set to 0.5 (i.e. 50%). This means that if the intersection ratio of the predicted boundary box and the true boundary box is at least 50%, the prediction is considered accurate. The mAP50 value is the average of the AP values calculated for all classes and the IoU threshold of 0.

Moreover, mAP50-95 is another average AP value obtained by calculating the IoU threshold values from 0.5 to 0.95 with every 0.05 increments. The mAP50-95 metric measures how well the model performs not only at a 50% IoU threshold but also at higher IoU values. This provides a more challenging and detailed assessment. In the Yolo algorithm training process, these loss functions are computed by using Python programming.

III. System Design and Components

C. Dataset

2372 bird images has been used for bird gesture classification from the Caltech-UCSD Birds-200-2011 (CUB-200-2011) dataset [16]. 1161 train set (80%), 207 validation set (%10) and 207 test set (%10) images are used for the Yolo training process. Additionally, the dataset image number has been increased by using data augmentation with different views such as rotation, and mirroring. A few sample bird images with different gestures are shown in Figure 1.

D. Pre-process

Caltech dataset originally contains 11788 bird images with different species, views and poses. In this study, 2372 bird images have been used for five classes. Image data should be optimized for the training process. The bird images have been classified into 5 classes that are perching, opening wings-flying (ow-flying), swimming, eating, and walking in this study.

Each dataset image has been prepared for the training process. The images should be segmented for each defined class before the training process. The data set must be balanced and the epoch number must not be excessive to prevent over-learning during the training process.

250 images are selected for each movement class and each class has been enhanced by data augmentation from the Caltech dataset. These selected images are increased by using the Roboflow framework as balanced up to 2372 images.

Additionally, the training process is limited by 100 epochs to prevent over-fitting.

Fig. 1. Sample bird images in Caltech dataset [17][18]

Roboflow is a computer vision developer framework that focusses on improving data gathering, preprocessing, and model training procedures [18]. Data images are segmented for each class by using the Roboflow framework. Two sample segmented bird images from the Caltech dataset are shown in Figure 2.

Table 1. The dataset image numbers for each classes

Class	The number of selected images	The number of data augmentation images	The final numbers of images
eating	250	224	474
ow- flying	250	223	473
perching	250	226	476
swimming	250	220	470
walking	250	229	479
Total	1250	1122	2372

Fig. 2. Segmented sample bird images with RoboFlow framework[18][19]

Additionally, Roboflow provides to increase images for each class by using different techniques such as reflector, reversing, rotating right, left, above or below the image for data augmentation. In this way, the selected images from the Caltech dataset have been increased based on these five classes in this study. The numbers of images in the final dataset for each class are shown in Table 1.

E. Bird Gesture Detection

The study aims to determine bird gesture detection with the highest accuracy. In this scope, the dataset is prepared by using pre-processing techniques. The five-movement classes have been drawn by Roboflow and the final dataset has been prepared for the training process. The training has been made by NVIDIA L4 GPU and 52GB RAM, 22,5 GB GPU RAM and 78GB disk space on the Google Colab platform with 100 epochs and 8 batch sizes. At the end of the training process, a bird gesture detection trained model file has been obtained for 5 classes (perching, swimming, opening wings-flying (ow-flying), eating, and walking). In the test process, the detected bird gestures are shown in Figure 3 by using various bird video samples.

IV. RESULTS

The study is built based on the Yolov10 algorithm for detection bird gesture detection. mAP50 is one of the most commonly used metrics to evaluate the performance of an object detection model. It measures the average precision at IoU=0.5, taking into account both the accuracy of the objects detected by the model and the proportion of objects it misses. The accuracy metrics which are precision, and recall based on mAP scores are shown in Table 2.

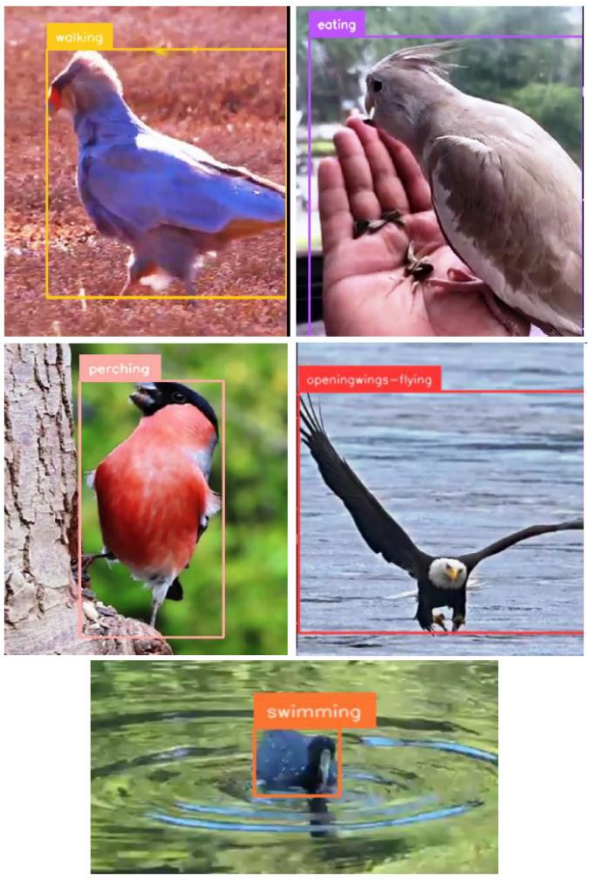


Fig. 3. Bird gesture detection samples for each class after training with the YOLOv10 model [20][21][22][23][24]

Table 2. Yolo Performance Metrics for each class

Class	Images	Instances	Performance Metrics			
			Precision	Recall	mAP50	mAP50-95
All	207	224	0.814	0.732	0.768	0.696
eating	207	39	0.833	0.640	0.756	0.688
ow- flying	207	43	0.943	0.860	0.913	0.882
perching	207	62	0.588	0.387	0.433	0.304
swimming	207	35	0.905	0.971	0.939	0.863
walking	207	45	0.800	0.800	0.799	0.743

Confusion Matrix is a table used to visualize and evaluate the prediction performance of a model. Confusion Matrix is used to better understand the classification or detection accuracy of a model [25]. Normalized Confusion Matrix is obtained by normalizing the value in each cell per class. Normalization is done by dividing the value in each cell by the total real count of that class. In this study, the normalized confusion matrix obtained as a result of training for five-bird gesture detection is shown in Figure 4.

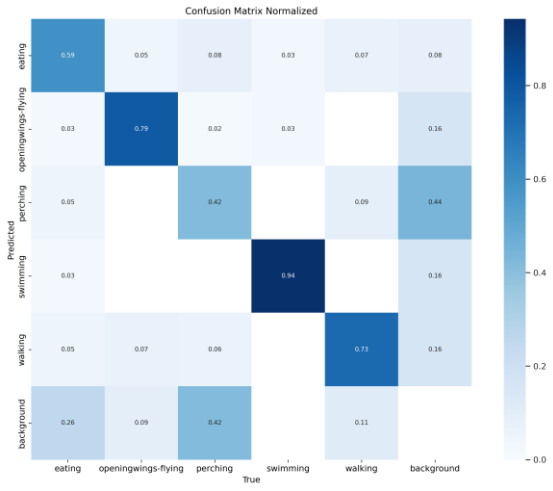


Fig. 4. Normalized confusion matrix for five classes

According to Table 2, the most challenging classes are perching and eating due to the similar pose of the bird. Because of the gesture similarity, the accuracy of these classes more less than the other classes. However, the perching position has been segmented with tree branches together to prevent gesture similarity. The best result was obtained from the flying position as it is a unique movement in birds. For the eating class, segmentation was made by taking into account whether the bird's beak is open or there is food in its beak. Because this segmentation is the most important feature that distinguishes the bird from the walking or perching class. In addition, since swimming birds such as ducks do not have visible feet on the water surface, better results can be obtained. Since the foot segmentation of the bird is particularly important in the walking position, the foot segmentation is highlighted together with the total segmentation of the bird in such images.

When mAP50 values are examined, it is seen that all classes except the perching position have an accuracy rate above 70%. In this study, a prototype study was conducted for bird gesture segmentation. The accuracy values can be increased with more balanced data and training to increase the accuracy rate of mAP50-95 values,

V. DISCUSSION

In this study, bird movements were classified into five different categories using the latest YOLO algorithm, YOLOv10, demonstrating its effectiveness in detecting and categorizing bird behaviours. Compared to previous studies, an increase in the number of movement classes was achieved, highlighting the potential of the proposed approach for capturing more granular behavioural distinctions. This advancement underscores the importance of leveraging state-of-the-art algorithms in ecological and behavioural research, as it allows for a more nuanced understanding and classification of animal behaviours.

However, despite the promising results, certain limitations were observed, primarily related to dataset size and balance. A balanced dataset is critical to avoid biases in classification performance across different movement categories. Future studies can enhance the dataset by including diverse and representative samples from various bird species and environments. This would not only improve the accuracy of movement detection but also enhance the generalizability of the model.

Additionally, integrating multimodal data such as bird sounds alongside visual data presents an exciting avenue for future research. Combining audio and visual modalities can enable a more comprehensive interpretation of bird communication, potentially leading to ground breaking insights into the interplay between movement and vocalization.

VI. CONCLUSION

This study successfully classified bird movements into five distinct categories using the advanced YOLOv10 algorithm, demonstrating the potential of deep learning models in ecological and behavioral research. By increasing the number of classes compared to previous studies, the model provides a more detailed understanding of bird behaviors. However, the findings also emphasize the importance of balanced datasets and adequate training to achieve optimal results. Future work integrating bird sounds and visual data could open new horizons in understanding bird communication, paving the way for more holistic and multimodal approaches in this field.

Authors' Contributions

The authors' contributions to the paper are equal.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics

REFERENCES

[1] T. Puiu, "How many birds are there in the world?" ZME Science: https://www.zmescience.com/feature-post/natural-sciences/animals/birds/how-many-birds-are-there-in-the-world/, 2023.

[2] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, "Yolov10: Real-time end-to-end object detection," arXiv preprint arXiv:2405.14458, 2024.
[3] H. Liang, X. Zhang, J. Kong, Z. Zhao, and K. Ma, "Smb-yolov5: A lightweight airport flying bird detection algorithm based on deep neural networks," IEEE Access, vol. 12, pp. 84 878–84 892, 2024.
[4] P. Datar, K. Jain, and B. Dhedhi, "Detection of birds in the wild using deep learning methods," in 2018 4th International Conference for Convergence in Technology (I2CT), 2018, pp. 1–4.
[5] M. Xie, X. Li, C. Zhao, and C. Xu, "Identification of bird nest based on yolov5 algorithm," in 2023 5th International Academic Exchange Conference on Science and Technology Innovation (IAECST), 2023, pp. 811–814.
[6] Y.-Q. Ou, C.-H. Lin, T.-C. Huang, and M.-F. Tsai, "Machine learning-based object recognition technology for bird identification system," in 2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan). IEEE, 2020, pp. 1–2
[7] S. Zhao, "Bird movement recognition research based on yolov4 model," in 2022 4th International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), 2022, pp. 441–444.*FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
[8] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, "Oriented r-cnn for object detection," in Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 3520–3529.
[9] R. Girshick, "Fast r-cnn," in 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448.
[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 6, pp. 1137– 1149, 2016.
[11] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.
[12] P. Shivaprasad, "A comprehensive guide to object detection using yolo framework," https://towardsdatascience.com/a-comprehensive-guide-to-object-detection-using-yolo-framework-24f8e2e5c6ab, jan 2019.
[13] E. Yildirim, U. G. Sefercik, and T. Kavzoglu, "Automated identification of vehicles in very high-resolution uav orthomosaics using yolov7 deep learning model." Turkish J. Electr. Eng. Comput. Sci., vol. 32, no. 1, pp. 144–165, 2024.
[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA, USA: IEEE Computer Society, jun 2016, pp. 779–788. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.91
[15] Y. Indulkar, "Alleviation of covid by means of social distancing face mask detection using yolo v4," in 2021 International Conference on Communication information and Computing Technology (ICCICT), 2021, pp. 1–8.
[16] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "Caltech-ucsd birds-200-2011 (cub-200-2011)," California Institute of Technology, Tech. Rep. CNS-TR-2011-001, 2011.
[17] Wang, F., Zhou, H., Li, S., Lei, J., & Zhang, J. (2020). Convolutional Attention Network with Maximizing Mutual Information for Fine-Grained Image Classification. Symmetry, 12(9), 1511.
[18] B. Dwyer, J. Nelson, T. Hansen, and et al., "Roboflow (version 1.0)," https://roboflow.com, 2024, computer vision software.
[19] Shandilya, S. K., Srivastav, A., Yemets, K., Datta, A., & Nagar, A. K. (2023). YOLO-based segmented dataset for drone vs. bird detection for deep and machine learning algorithms. Data in Brief, 50, 109355.
[20] HAWIStudios. (2024, Jan 17). Caracara bird walking in field. [Short video]. https://youtube.com/shorts/fGuZgvVb3uc?si=AYq2f8v93TthqseH
[21] DiscoverAnimalAll. (2024, Jul 21). Bird Perched On A Tree While Eating. [Short video]. https://youtube.com/shorts/Qk53nb8kEQM?si=f0EUK-u_73BTRf8s
[22] devang_jani. (2023, Nov 27). Bird eating. [Reel]. Instagram. https://www.instagram.com/devang_jani/reel/C0I5L3wpkfH/
[23] MarkSmithphotography. (2023, June 30).Must SEE!! Extreme close up off a Bald Eagle snatching a fish from a whirlpool.[Short video]. https://youtube.com/shorts/saoQHEJCkBM?si=OHzeoK6BLCsiRTKX
[24] @recep.kaplan. (2023, Nov 8). Ankara Gölbaşı Mogan Gölü'nde Ördekler. [Short video]. https://youtube.com/shorts/aGSCvYjjt9I?si=mWdb6y3j9E1tpVUC
[25] N. Herbaz, H. El Idrissi, and A. Badri, "Deep learning empowered hand gesture recognition: using yolo techniques," in 2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA), 2023, pp. 1–7.

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 81-85
DOI: 10.36287/ijmsit.8.2.6
Received: November 11, 2024; Accepted: December 9, 2024

Autonomous Flight Systems and Generative AI

Ali Berkol[1*], İdil Gökçe Demirtaş [1]

[1*]*Aselsan-Bites – Defence and Information Systems, Ankara, Turkey (ali.berkol@yahoo.com) (ORCID: 0000-0002-3056-1226)*
[1*]*Aselsan-Bites – Defence and Information Systems, Ankara, Turkey (idil.demirtas@bites.com.tr) (ORCID: 0000-0003-1704-1581)*

Abstract – Autonomous flight systems have emerged as a significant area of research and development within the aviation industry. With the advancements in artificial intelligence (AI), particularly generative AI, these systems have witnessed substantial improvements in their capabilities and efficiency. This abstract explores the integration of generative AI techniques in autonomous flight systems and its implications on the aviation sector. Generative AI algorithms play a crucial role in various aspects of autonomous flight, including flight path planning, obstacle detection and avoidance, decision-making processes, and even aircraft design optimization. By leveraging generative AI, autonomous flight systems can adapt and respond to dynamic environments in real-time, enhancing safety, efficiency, and reliability. Furthermore, generative AI enables the generation of innovative solutions and designs that may not be apparent through traditional methods, leading to more optimized and efficient aircraft configurations. This abstract also discusses the challenges and future directions in the utilization of generative AI in autonomous flight systems, including regulatory considerations, ethical concerns, and the need for continued research and development. Overall, the integration of generative AI in autonomous flight systems represents a promising avenue for advancing the capabilities and effectiveness of aviation technology in the 21st century.

Keywords *– Autonomous Flight Systems, Generative AI, Aviation Technology, Aircraft Automation, Artificial Intelligence Integration*

Citation: Berkol, A., Demirtaş, I. (2024). Autonomous Flight Systems and Generative AI. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 81-85.

I. Introduction

Integrating autonomous flight systems with generative artificial intelligence (AI) represents a pivotal advancement in aviation technology, offering unprecedented opportunities for safety, efficiency, and innovation. In recent years, the aviation industry has witnessed a significant shift towards autonomous capabilities, driven by advancements in AI algorithms, sensor technologies, and computational power. Generative AI, in particular, has emerged as a transformative tool in enhancing the capabilities of autonomous flight systems by enabling adaptive decision-making, real-time response to dynamic environments, and the generation of novel solutions. This introduction sets the stage for exploring the intersection of autonomous flight systems and generative AI, highlighting the potential benefits, challenges, and future directions in this rapidly evolving field. By examining the role of generative AI in various aspects of autonomous flight, from flight path planning to aircraft design optimization, this paper aims to provide insights into the transformative impact of AI-driven technologies on the future of aviation. Moreover, it addresses key considerations such as regulatory frameworks, ethical implications, and the need for continued research and development to realize the full potential of autonomous flight systems integrated with generative AI.

II. Materials and Method

The study employed a combination of experimental testing and computational modeling techniques to evaluate the performance of autonomous flight systems integrated with generative AI algorithms. Various UAV platforms equipped with state-of-the-art sensors were utilized for data collection, while advanced machine learning algorithms were implemented for real-time decision-making and trajectory optimization.

A. Foundations of Autonomous Flight Systems: Evolution, Components, and Challenges

The evolution of autonomous flight systems spans over a century of technological advancements, innovation, and paradigm shifts in the aviation industry. At the dawn of aviation, rudimentary autopilot systems were developed to assist pilots in maintaining straight and level flight. However, these early systems were limited in functionality and primarily used as aids rather than fully autonomous solutions.

The true evolution of autonomous flight systems began with the emergence of unmanned aerial vehicles (UAVs) during the mid-20th century. Initially employed for reconnaissance and surveillance purposes, UAVs gradually evolved to encompass a wide range of applications, including military missions, scientific research, and commercial operations. Early UAVs relied on pre-programmed flight paths and basic navigation systems, but advancements in sensor technologies and

computing power paved the way for more sophisticated autonomous capabilities.

In recent decades, the convergence of artificial intelligence, machine learning, and sensor fusion technologies has propelled autonomous flight systems to unprecedented levels of sophistication. Modern UAVs are equipped with a myriad of sensors, including GPS, cameras, LiDAR, radar, and inertial measurement units (IMUs), enabling precise navigation, obstacle detection, and environmental awareness. Furthermore, advancements in machine learning algorithms have empowered UAVs to adapt and respond to dynamic operating conditions, autonomously optimizing flight paths, avoiding obstacles, and making real-time decisions.

The evolution of autonomous flight systems has not only revolutionized military operations but also transformed various civilian industries. In agriculture, UAVs equipped with multispectral cameras and infrared sensors are used for crop monitoring and precision agriculture. In infrastructure inspection, UAVs equipped with LiDAR scanners and high-resolution cameras can efficiently survey bridges, pipelines, and power lines, reducing inspection costs and enhancing safety.

Looking ahead, the evolution of autonomous flight systems is expected to continue rapidly, driven by ongoing advancements in sensor technologies, artificial intelligence, and robotics. As regulatory frameworks evolve to accommodate autonomous aerial vehicles, the integration of UAVs into everyday life is poised to revolutionize transportation, logistics, and emergency response, ushering in a new era of autonomous aviation.

Autonomous flight systems comprise several interconnected components that work together to enable unmanned aerial vehicles (UAVs) to operate autonomously and safely. These components can be broadly categorized into three main groups: sensing and perception systems, processing and decision-making units, and control and communication interfaces.

Sensing and Perception Systems:

Sensing and perception systems are critical for enabling UAVs to perceive their environment, detect obstacles, and navigate autonomously. These systems typically include a variety of sensors such as:

GPS (Global Positioning System): Provides accurate positioning and navigation information.

Cameras: Capture visual data for navigation, obstacle detection, and situational awareness.

LiDAR (Light Detection and Ranging): Measures distances by illuminating targets with laser light and analyzing the reflected signals, used for terrain mapping and obstacle avoidance.

Radar: Detects objects and measures their distance and velocity, particularly useful in adverse weather conditions or low visibility situations.

IMU (Inertial Measurement Unit): Measures and reports a vehicle's specific force, angular rate, and sometimes the magnetic field surrounding the vehicle, aiding in navigation and stabilization.

Processing and Decision-Making Units:

Processing and decision-making units are responsible for analyzing sensor data, making real-time decisions, and planning optimal trajectories for the UAV. These units typically consist of:

Onboard computers: Process sensor data and run algorithms for perception, path planning, and decision-making.

Machine learning algorithms: Analyze sensor data to recognize objects, predict their behavior, and make decisions based on learned patterns and models.

Path planning algorithms: Generate optimal flight paths considering factors such as mission objectives, airspace regulations, and obstacle avoidance.

Decision-making algorithms: Evaluate sensor data, assess potential risks, and make decisions to ensure safe and efficient flight operations.

Control and Communication Interfaces:

Control and communication interfaces enable operators to interact with the UAV and monitor its status remotely. These interfaces include:

Flight control systems: Translate trajectory commands into control inputs for the UAV's actuators (such as motors and servos) to execute desired maneuvers.

Telemetry systems: Transmit real-time data from the UAV to the ground control station, including telemetry data (e.g., altitude, speed, battery status) and video feeds.

Ground control stations (GCS): Provide operators with a user interface for mission planning, monitoring, and controlling UAV operations.

Autonomous flight systems, while offering numerous advantages, also face several significant challenges that must be addressed for widespread adoption and successful integration into various industries. These challenges span technical, regulatory, safety, and societal domains, and include:

Adverse Weather Conditions:

Autonomous flight systems must operate reliably in a wide range of weather conditions, including high winds, fog, rain, and snow. Adverse weather can impair sensor performance, reduce flight stability, and increase the risk of accidents, necessitating robust weather-proofing solutions and sophisticated control algorithms.

Obstacle Detection and Avoidance:

Accurate and reliable obstacle detection and avoidance are critical for ensuring the safety of autonomous flight operations. Challenges include the detection of small or low-contrast obstacles, dynamic obstacle tracking, and effective collision avoidance strategies in complex environments such as urban areas and densely populated airspace.

Regulatory Frameworks:

The integration of autonomous flight systems into airspace regulations presents significant regulatory challenges. Existing regulations often prioritize manned aviation and may not fully accommodate the unique capabilities and operational characteristics of unmanned aerial vehicles (UAVs). Addressing regulatory barriers requires collaboration between industry stakeholders, regulatory agencies, and policymakers to develop clear guidelines and standards for safe and lawful autonomous flight operations.

4. Safety and Security Concerns:

Safety and security are paramount considerations in autonomous flight systems, particularly in light of potential cyber threats and malicious attacks. Ensuring the integrity and resilience of UAV systems against hacking, spoofing, and

jamming attacks is essential to safeguarding airspace and protecting public safety.

5. Human Factors and Public Acceptance:

Integrating autonomous flight systems into society requires addressing human factors and fostering public acceptance. Concerns regarding privacy, noise pollution, and the impact on traditional aviation jobs must be carefully addressed through education, outreach, and transparent communication to build trust and confidence in autonomous flight technologies.

Addressing these challenges will require collaborative efforts from industry, academia, government agencies, and regulatory bodies to develop innovative solutions, establish clear guidelines, and ensure the safe and responsible integration of autonomous flight systems into everyday life.

B. Challenges and Considerations

The integration of generative AI into autonomous flight systems presents several challenges and considerations that must be carefully addressed to ensure the safe and effective deployment of these technologies. Key challenges and considerations include:

Data Quality and Reliability:

Generative AI algorithms rely heavily on large datasets to generate accurate and reliable outputs. Ensuring the quality, diversity, and integrity of training data is crucial for the effectiveness and robustness of generative AI models. Challenges include data scarcity, bias, and inaccuracies, as well as the need for data annotation and labeling to facilitate supervised learning.

Algorithm Robustness and Generalization:

Generative AI algorithms must demonstrate robustness and generalization across diverse operating conditions, environments, and scenarios. Challenges include adversarial attacks, domain shifts, and the ability to adapt to novel or unforeseen situations. Ensuring the reliability and scalability of generative AI algorithms requires rigorous testing, validation, and benchmarking against real-world data and performance metrics.

Ethical and Legal Considerations:

The use of generative AI in autonomous flight systems raises ethical and legal considerations regarding privacy, accountability, and transparency. Challenges include the ethical implications of AI-generated content, such as deepfakes and synthetic media, and concerns about algorithmic bias and fairness. Addressing these considerations requires the development of ethical guidelines, regulatory frameworks, and governance mechanisms to promote responsible AI use and mitigate potential risks.

Safety and Security:

Ensuring the safety and security of autonomous flight systems augmented with generative AI is paramount. Challenges include the potential for AI-induced failures or errors, cybersecurity threats, and adversarial attacks targeting AI models or data pipelines. Robust safety and security measures, including redundancy, fail-safes, and cybersecurity protocols, are essential to mitigate risks and safeguard against potential hazards.

Human-AI Collaboration:

Effective human-AI collaboration is essential for maximizing the benefits of generative AI in autonomous flight systems. Challenges include user acceptance, trust, and understanding of AI-generated outputs, as well as the integration of AI recommendations into human decision-making processes. Human-centered design principles and user interface enhancements can facilitate seamless interaction and collaboration between humans and AI systems, promoting user trust, confidence, and acceptance.

Addressing these challenges and considerations will require interdisciplinary collaboration, stakeholder engagement, and ongoing research and development efforts to ensure the responsible and beneficial integration of generative AI into autonomous flight systems.

C. Future Directions and Emerging Trends:

The future of autonomous flight systems augmented with generative AI holds immense potential for innovation, transformation, and advancement in various domains. Emerging trends and future directions in this rapidly evolving field include:

Advancements in Generative AI Algorithms:

Continued advancements in generative AI algorithms, including deep learning models, reinforcement learning techniques, and evolutionary algorithms, are expected to drive improvements in autonomous flight systems' capabilities and performance. Future research efforts will focus on developing more robust, efficient, and adaptive AI models capable of addressing complex real-world challenges and scenarios.

Integration of Multi-Sensory Data Fusion:

Integrating multi-sensory data fusion techniques, combining inputs from diverse sensor modalities such as cameras, LiDAR, radar, and GPS, will enhance autonomous flight systems' perception, situational awareness, and decision-making capabilities. Future trends will emphasize the development of sensor fusion architectures and algorithms optimized for real-time, multi-modal data processing and analysis.

Autonomous Collaborative Swarms:

The emergence of autonomous collaborative swarms, comprising fleets of interconnected UAVs operating collaboratively to achieve shared objectives, represents a promising future direction for autonomous flight systems. Collaborative swarm technologies enable distributed sensing, communication, and coordination among multiple UAVs, facilitating scalable, resilient, and adaptive mission execution in dynamic environments.

Urban Air Mobility (UAM) Solutions:

The rise of urban air mobility (UAM) solutions, including passenger drones, air taxis, and autonomous aerial delivery services, is poised to revolutionize urban transportation and logistics. Future trends will focus on developing UAM infrastructure, airspace management systems, and regulatory frameworks to support the safe, efficient, and sustainable integration of autonomous aerial vehicles into urban environments.

Human-Centered Design and Human-AI Interaction:

Human-centered design principles and human-AI interaction paradigms will play a crucial role in shaping the future of autonomous flight systems. Future trends will emphasize the development of intuitive user interfaces, transparent AI decision-making processes, and collaborative human-AI interaction frameworks to enhance user trust, acceptance, and engagement with autonomous flight technologies.

Sustainability and Environmental Impact:

Addressing sustainability and environmental impact considerations will be increasingly important in the future development and deployment of autonomous flight systems. Future trends will focus on optimizing energy efficiency, reducing carbon emissions, and mitigating environmental impact through innovations in electric propulsion, lightweight materials, and eco-friendly operational practices.

As autonomous flight systems continue to evolve and mature, interdisciplinary collaboration, stakeholder engagement, and responsible innovation will be essential to realize their full potential in shaping the future of aviation and society.

III. DISCUSSION

Integrating generative AI into autonomous flight systems represents a significant advancement in aviation technology, offering transformative opportunities for safety, efficiency, and innovation. In this discussion, we reflect on the key findings, implications, and future directions arising from exploring autonomous flight systems augmented with generative AI.

Enhancements in Autonomous Flight Capabilities:

Incorporating generative AI algorithms has led to notable enhancements in the capabilities of autonomous flight systems. These enhancements include improved situational awareness, adaptive decision-making, and real-time response to dynamic environments. By leveraging generative AI, autonomous flight systems can navigate complex scenarios with greater precision, efficiency, and autonomy, thereby enhancing overall operational effectiveness and reliability.

Challenges and Considerations:

However, integrating generative AI into autonomous flight systems also presents several challenges and considerations that must be carefully addressed. These include concerns related to data quality and reliability, algorithm robustness and generalization, ethical and legal implications, safety and security, and human-AI collaboration. Addressing these challenges will require interdisciplinary collaboration, stakeholder engagement, and ongoing research and development efforts to ensure the responsible and beneficial integration of generative AI technologies into aviation.

Future Directions and Emerging Trends:

Looking ahead, the future of autonomous flight systems augmented with generative AI is characterized by several emerging trends and future directions. These include advancements in generative AI algorithms, integration of multi-sensory data fusion techniques, development of autonomous collaborative swarms, emergence of urban air mobility solutions, emphasis on human-centered design and human-AI interaction, and considerations for sustainability and environmental impact. By embracing these trends and leveraging emerging technologies, autonomous flight systems can continue to evolve and mature, shaping the future of aviation and society.

Implications for Research and Practice:

The insights gained from this discussion have several implications for both research and practice in the field of autonomous flight systems. Researchers are encouraged to explore innovative solutions to address the challenges and considerations identified, while practitioners should prioritize safety, ethics, and sustainability in the development and deployment of autonomous flight technologies. Furthermore, collaboration between academia, industry, government agencies, and regulatory bodies is essential to foster innovation, ensure regulatory compliance, and promote responsible AI use in aviation.

IV. CONCLUSION

In conclusion, the integration of generative AI into autonomous flight systems holds immense promise for revolutionizing aviation technology. By addressing the challenges, embracing emerging trends, and fostering interdisciplinary collaboration, we can unlock the full potential of autonomous flight systems augmented with generative AI, paving the way for safer, more efficient, and sustainable aviation solutions in the future.

Throughout this study, we have examined the evolution of autonomous flight systems, explored the components and challenges of integrating generative AI, and discussed emerging trends and future directions in the field. Key findings include the transformative potential of generative AI algorithms in enhancing the capabilities of autonomous flight systems, as well as the challenges and considerations that must be addressed to ensure their safe and effective integration.

The insights gained from our discussion have several implications for both research and practice in the aviation industry. Practitioners are encouraged to prioritize safety, ethics, and regulatory compliance in the development and deployment of autonomous flight technologies, while researchers should continue to explore innovative solutions to address the challenges identified. Furthermore, collaboration between academia, industry, and regulatory bodies is essential to foster responsible innovation and ensure the ethical and beneficial use of generative AI in aviation.

Looking ahead, the future of autonomous flight systems augmented with generative AI is characterized by several emerging trends and opportunities. Advancements in generative AI algorithms, integration of multi-sensory data fusion techniques, development of autonomous collaborative swarms, emergence of urban air mobility solutions, and emphasis on human-centered design and human-AI interaction are among the key trends shaping the future of aviation. By embracing these trends and leveraging emerging technologies, we can unlock the full potential of autonomous flight systems, paving the way for safer, more efficient, and sustainable aviation solutions in the future.

Authors' Contributions

The authors' contributions to the paper are equal.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics

REFERENCES

[1] Shao, Y., & Yang, S. (2020). Applications of Generative Adversarial Networks (GANs) in Aircraft Design: A Survey. Aerospace, 7(8), 108. https://doi.org/10.3390/aerospace7080108

[2] Zhou, D., Hu, H., Wu, J., & Peng, Z. (2020). Generative adversarial networks in UAV-aided communications: A comprehensive survey. IEEE Access, 8, 135857-135876. https://doi.org/10.1109/ACCESS.2020.3015019

[3] Karaman, S., & Frazzoli, E. (2020). Autonomous aerial vehicles: A survey of the state of the art. Proceedings of the IEEE, 108(2), 252-294. https://doi.org/10.1109/JPROC.2019.2955685

[4] Brown, C., De Wet, P., Bruckstein, A. M., & Sukkarieh, S. (2020). A survey of visual SLAM in unmanned aerial vehicles. Journal of Field Robotics, 37(3), 405-448. https://doi.org/10.1002/rob.21892

[5] Khalifa, S., Mustafa, M., & Guizani, M. (2021). Machine Learning-Driven Autonomous UAV Navigation: Challenges and Opportunities. IEEE Access, 9, 73138-73156. https://doi.org/10.1109/ACCESS.2021.3073353

[6] Yang, L., Cui, W., & Meng, X. (2021). A Survey of Key Technologies for Autonomous Flight of UAVs. IEEE Access, 9, 68745-68766. https://doi.org/10.1109/ACCESS.2021.3070011

[7] Zhang, Y., Chen, L., & Zhang, H. (2020). A survey on deep learning for UAV-based communication networks. IEEE Access, 8, 22147-22159. https://doi.org/10.1109/ACCESS.2020.2960668

[8] Sun, Y., Yu, L., & Zhang, Y. (2021). A Survey of Machine Learning for Unmanned Aerial Vehicles: Recent Advances, Taxonomy, and Challenges. IEEE Transactions on Intelligent Transportation Systems, 22(1), 543-562. https://doi.org/10.1109/TITS.2020.2979192

[9] Roldán-Álvarez, D., Del-Blanco, C. R., & Cuadra-Troncoso, A. (2020). UAV-based data collection, processing, and analysis for smart agriculture: A comprehensive review of advances and applications. Computers and Electronics in Agriculture, 178, 105766. https://doi.org/10.1016/j.compag.2020.105766

[10] Jia, F., Dong, Y., Shi, H., Wang, J., & Chen, L. (2021). Recent advances in machine learning for unmanned aerial vehicles: Algorithms, tools, and applications. Journal of Aerospace Information Systems, 18(1), 32-49. https://doi.org/10.2514/1.I010984

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Review Article

2024, 8 (2), 86-91
DOI: 10.36287/ijmsit.8.2.7
Received: November 8, 2024; Accepted: December 5, 2024

Deep Learning Based Color and Style Transfer: A Review and Challenges

Melike Bektaş Kösesoy[1*] and Seçkin Yılmaz [2]

[1*] *Department of Computer Engineering, Graduate School, Bursa Technical University, Bursa, Turkey (melike.bektas@btu.edu.tr) (ORCID: 0000-0002-1944-1928)*

[2] *Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Bursa Technical University, Bursa, Turkey (seckin.yilmaz@btu.edu.tr) (ORCID: 0000-0001-6791-1536)*

Abstract – Deep learning methods have been applied in many fields in recent years, and successful results have been obtained. Image processing is one of these areas. One of the image processing applications using deep learning is color and style transfer. Color and style transfer is aimed at transferring the color and texture from the source image to another image (the target image). In color transfer, the colors in the source image are transferred, while in style transfer, texture is transferred as well as color. In the literature, color transfer has been studied for many years, and traditional methods such as PCA have been used in addition to deep learning. On the other hand, studies about the style transfer are relatively new and mostly realized by using deep learning methods. In this study, color and style transfer studies in the literature are examined. The methods used in these studies are mentioned, and the current problems in this field are shared.

Keywords *– Color Transfer, Neural Style Transfer, Image Colorization, Image Recoloring, CNN.*

Citation: Bektaş Kösesoy, M., Yılmaz, S. (2024). Deep Learning Based Color and Style Transfer: A Review and Challenges. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 86-91.

I. INTRODUCTION

The development of technologies such as artificial intelligence and image processing has made it possible to modify images [1], [2]. As a result of these advancements, aspects like the colorization of black and white images, the recoloring of colored images, and the transfer of color and style between images have gained popularity. As digital images can be seen as a well-known artwork using style transfer, color transfer can make a recolored image. In this paper, literature review about color and neural style transfer have been addressed. In the context of colorization, limitations and challenges of these studies have been discussed.

Color and style transfer problems are very comprehensive and hot topics and there are many studies in the literature on this subject. Graphics in Figure 1 present the number of manuscript's numbers of in the Scopus journals lists between 2014 and 2024 which include “neural style transfer” and “image color transfer” keywords, respectively.

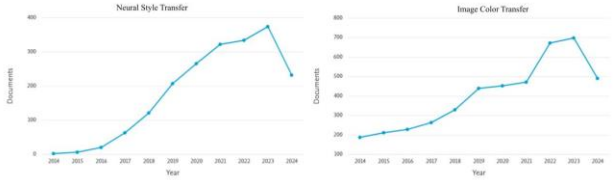


Fig. 1 Number of publications between 2014 and 2024 [3]

Figure 1 shows the number of publications for the last decade. As seen in Figure 1, the number of studies on this topic has been steadily increasing (except for 2024). Considering that 2024 is not yet complete, we expect that there will be more publications on color and style transfer at the end of this year than in 2023.

The rest of the paper is organized as follows: in Section 2, the neural style transfer and literature review about the neural style transfer are explained; in Section 3, the image color transfer and literature review about the image color transfer are explained, in Section 4, encountered difficulties about color and neural style transfers are presented, in Section 5, the conclusions are given.

II. NEURAL STYLE TRANSFER

Neural Style Transfer (NST) is the process of transferring the style of one image to another by the neural network methods [4]. This process requires a style and a context image. The artistic style, colors and texture features of the style image are considered without changing the content and structural features of the context image. Usually, deep learning algorithms are used to extract the features of the context and style images. The features of the style image are transferred to the context image and a target image is generated by doing this [5]. Today, it is possible to use the style of a painting by a famous artist who is no longer alive to obtain multiple target images that look as if they were drawn by the artist [6][7]. Figure 2 shows an example of NST [8].

The first and second column in Figure 2 shows the content and style image respectively, and the last column shows the target image obtained after the NST. The images in Figure 2 are generated using the method proposed by Gayts et al.

Fig. 2 An example of NST

There are many deep learning-based methods such as convolutional neural network (CNN), Generative Adversarial Network (GAN), Variational Autoencoder, Transformer, and Attention mechanism for the implementation of style transfer [4]. These methods are typically used in the RGB color space. Gatys et al. conducted the first study on NST in 2015 [9]. CNN-based VGG16 model was used in this study. An image depicting the Neckarfront in Tübingen was used as a content image. JMV's Shipwreck of the Minotaur, Vincent van Gogh's Starry Night, Edvard Munch's Scream, Pablo Picasso's Seated Nude Woman, and Wassily Kandinsky's Composition VII were selected as style images. The artistic styles in the style images were transferred to the content image.

Zhang et al. colorized anime images using Residual U-net and Auxiliary Classifier GAN (AC-GAN) algorithms [10]. Because anime images are randomly colored using the style transfer method, in this study, areas in the anime image were classified and colored by hair, eyes, clothing, and skin color. VGG model was used for the classification process. However, it was observed that the VGG model is feasible for classification of photographs, but not for drawings. In addition, it was observed that because the obtained colorized images were blurred, quality of these images were low.

Karadağ et al. compared the NST performance of VGG16, VGG19 and ResNet50 models using different optimization algorithms. Leonardo da Vinci's Mona Lisa, Pablo Picasso's Weeping Woman, Vincent van Gogh's Cypresses and Edvard Munch's Scream were used as style images. An image of an asphalt road was chosen as the content image. The best visual performance was achieved with the VGG19 model using the Stochastic Gradient Descent (SGD) optimization algorithm. The fastest results in terms of time were obtained using the ResNet50 model and the SGD optimization algorithm [11].

Lian and Ciu transferred the colors of hair, clothing, skin, and other features of characters in anime images to grayscale anime images. In this study, the Spatially Adaptive (DE) Normalization (SPADE) method was proposed for style transfer. It was reported that the colorized images obtained were consistent with the style image and exhibited good visual quality. However, upon examining the resulting images, it was observed that the shades of the colors in the style image and the target image did not match precisely, leading to tonal differences between the colors [12].

JinKua et al. performed style transfer and colorization in their study. The VGG19 model was used for style transfer. The style transferred black and white image was colorized by estimating the a* and b* channels in the La*b* color space. Then, the resolution of the resulting images was increased using the CNN algorithm [13].

Ke et al. proposed the Neural Preset method to solve the problems of style transfer methods such as high memory requirements and time consuming. The proposed method consists of two components. One of these components is deterministic neural color mapping. This component reduces error, blur and distortion by providing a consistent color mapping in each pixel. The other component is a two-stage pipeline for color normalization and stylization. It has been reported that the proposed method has advantages over existing techniques, such as preserving image textures, providing color properties more consistent with style images. However, it has limitations in style transfer between different colors and local color matching [14].

Virtusio et al. proposed the Neural Style Palette (NSP) method. Existing NST methods produce a limited variety of outputs. The proposed method aimed to generate various stylized images from a single style image input. A sample user interface was also developed in the study. Consequently, a user-interactive style transfer method was proposed, enabling users to maximize, minimize, or remove certain style features. [15].

Deng et al. proposed a transformer-based approach StyTr2 for NST. This approach utilized two distinct transformer encoders: content and style transformer encoders. These encoders process content and style sequences separately, while the transformer decoder stylizes content sequences according to style sequences. It has been reported that the proposed method was more successful than traditional CNN-based models but slower in terms of speed. [16].

Fu focused on the blurring problem encountered in traditional methods in style transfer studies. He proposed a CycleGAN-based method to solve the blurring problem [17]. Huang et al. proposed a method called QuantArt to improve the visual fidelity of NST. They reported that they adopted a vector quantization-based approach to ensure that the latent features of the target image generated in the proposed method are closer to the real distribution. The method was tested for digital image-to-artwork, artwork-to-artwork, and digital image-to-digital image style transfer operations, in [18].

Fang et al. colorized anime images using the style image and the coloring style given in the text. In this study, a GAN-based method with one generator and two discriminator networks was proposed. The generator network was designed using U-Net architecture. As input, the drawing was stylized and colored by taking the anime image and the coloring style in the text. One of the discriminative networks was used for color and the other for style. When the results of the study were examined, it was observed that some images had problems due to blur and color tones [19].

Zheng and Zhang colorized flower images. They proposed a two-stage method for drawing extraction and colorization. In the first stage, drawing was considered a type of style rather than edge detection algorithms. CNN-based style transfer method was used to obtain the drawing. In the second stage, image colorization was performed with GAN-based style transfer. It has been reported that the two-stage method captures the color tone and outlines better than the one-stage method. However, in some cases, the two-stage method was not able to preserve the color features of the content images [20].

Luan et al. performed style transfer between images using a deep learning approach. To improve the success of the style transfer, the colors or pixels in local regions of the image were changed by linear transformations. Furthermore, the approach is supported by semantic segmentation. The proposed method was tested in many scenarios, and realistic results were

obtained [21]. Wang et al. presented a multimodal transfer method for fast and efficient transfer of artistic styles to daily photographs. In style transfer studies, a problem in high-resolution images is that local regions appear less like the desired artistic style. In this study, an approach that learns style cues (such as color, texture structure) at various scales was proposed to solve this problem [22].

Ciu presented a deep learning-based style transfer method based on HSV color space. The proposed method aimed to exploit the strengths of the HSV color space model in representing color types. In [23], the L2 distance between the H factor in the HSV color model of the style-transferred image and the content image is added to the loss function. As a result of the tests, it was reported that proposed algorithm preserves the color tone of the original content image.

Liao and Huang focus on the CycleGAN algorithm to perform style transfer. Also, they included comparisons of various loss functions [24]. NST studies usually focus on color and texture transfer and ignore other components of style. Liu et al. proposed a new network architecture that enables the transfer of both texture and geometric style. As a result of the tests, it was reported that the proposed method improves the qualitative expressive power of stylized images and shows more similarity to the target styles than other algorithms. However, some styles such as Cubism were out of the scope of this method [25].

Han et al. proposed depth extraction generative adversarial network (DE-GAN) model. In the proposed model, they applied multiple feature extractors such as U-Net, multi-item extractor, Fast Fourier Transform and MiDas depth estimation network to extract color, texture, depth features and shape masks from style images. After experimental comparisons with StyleGAN and CycleGAN, it was reported that the images produced using DE-GAN have higher image quality. However, it was reported that the DE-GAN model is a general artistic style transition network and has a worse transition effect compared to some specialized style transition methods such as CartoonGAN [26].

III. Image Color Transfer

Image color transfer is the process of changing the color content of the target image by transferring the colors of the source image to the target image [27]. In the image color transfer, only color is transferred, while in the style transfer process, features of the style image such as color and texture are transferred. Figure 3 shows an example of color transfer [28].

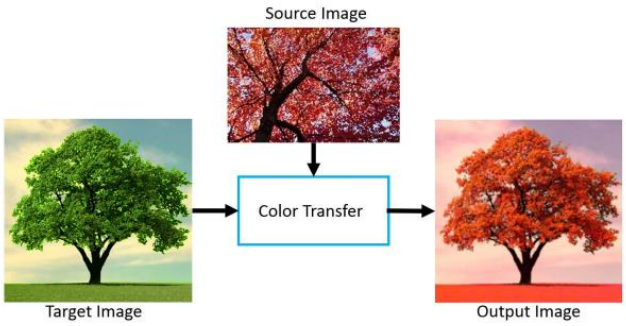


Fig. 3 An example of image color transfer

Today, color transfer is usually performed using deep learning algorithms, but it can also be performed based on statistical information, user interaction or hybrid methods [29]. One of the important studies on color transfer was conducted by Reinhard et al. in 2001. In this study, color transfer was performed using statistical information (color mean and standard deviation) in the image. It was reported that the proposed method was applied in the lαβ color space [30]. Abadpour and Kasaei also performed color transfer between images using Principal Component Analysis (PCA) dimensionality reduction method. In the study, both grayscale images were colored, and color images were recolored [31].

An and Pellacini proposed a framework in which the user marks the colors with a brush on the source and target image and performs color transfer from the target image to the source image according to the marked areas. It was reported that the proposed framework matches color distributions using a transfer function for each pair of brushes and minimizes visible distortions [32]. Arbelot et al. proposed a method for color transfer and colorization in their work. The proposed method performs local color transformations between input and reference images using an edge-aware texture descriptor. However, it was reported that the lack of sufficient similarity between the input and reference images in the proposed method negatively affects the success of color transfer [33].

Xu et al. proposed a Color Network Model to transfer colors from an image to another. This model consists of two sub-networks: source network, which describes the color information of the reference image, and target network, which indicates the target object to be colored. Color extraction from the reference image was performed with the k-means algorithm. In the study, the k-means algorithm was applied twice to perform color extraction. The limitations of the study are the evaluation of the colorized image by the designer and the many parameters that are manually determined in the process. For example, parameters such as the number of groups determined by the designer in the colorization process are manually entered. In addition, the k parameters of the k-means clustering algorithm used to extract the colors from the reference image were also determined by the designer [34].

Gu et al. reported using the Gaussian Mixture Model (GMM) algorithm for pixel-wise color transfer. They found that the proposed method produced successful results and that multiple transfer results could be obtained. However, they noted that the method struggles to produce the desired results when there are very similar color tones between the two input images [35]. Xiao and Ma focused on the fidelity problem in the color transfer. A gradient preservation algorithm was proposed to solve this problem. The algorithm was formulated as a two-stage optimization problem to achieve these goals. Furthermore, a new metric was introduced in the study to objectively evaluate the fidelity of global color transfer algorithms [36].

Lee et al. proposed a deep neural network method that uses histogram similarity for color transfer. The proposed method was tested for different scenarios where the relationship between the source and reference image was strongly correlated, weakly correlated and uncorrelated. As a result of the tests, it was reported that the method showed moderate performance for all cases and was comparable to specialized state-of-the-art color transfer methods [37]. Yin et al. proposed a color transfer method based on CNN algorithm to remove blur in images. They reported that the brightness and clarity of a blur image can be effectively recovered with this process [38].

Liu et al. performed deep learning based emotional color transfer. The deep learning model reportedly consists of four main networks: a low-level feature network, an emotion classification network, a fusion network, and a colorization network. Since the training set of the emotional colorization network was manually created, it was noted that the training and test sets are limited. As a result of, they indicated that they aim to achieve more accurate results by expanding the training set of the emotional colorization network [39]. Zhang et al. proposed a deep learning-based color transfer method for recoloring 3D models. The proposed model consists of two modules: Color Transfer Network and 3D Texture Optimization Module [40].

IV. Challenges in Color and Neural Style Transfer

NST and image colorization are important research topics in image processing. In the literature, it is observed that various studies have been carried out by considering these two separate topics together. When these studies are examined, it is observed that style transfer and colorization have various difficulties, limitations and disadvantages. NST aims to apply the style of another image while preserving the content of an image, while colorization aims to color an image in black and white or drawing form with appropriate colors.

Color transfer is the process of changing the colors of the target image with respect to the reference image [27]. This image editing process is currently used in many different fields such as colorization of grayscale images, recoloring of color images [37], image de-blurring [38], image stitching [41][42]. However, colorization by color transfer has some limitations and challenges. Some of these limitations that we have encountered in the literature are as follows:

- *Computational cost:* For color and neural style transfer generally deep learning-based methods are used. These methods require a large amount of computing power and memory, especially when they are trained on large datasets. This can make the process time-consuming and more costly. For example, style transfer of a high-resolution image may require a high computational cost[14], [21], [22], [37].
- *Re-editing the output images:* One of the challenges is that the images obtained after color and neural style transfer operations can't be edited in real time or afterwards according to user requirements [33].
- *Lack of control:* In color and neural style transfer, images are perceived as a whole, and the transfer process is usually applied to the entire target image. Thus, it may be difficult to select local regions of the image to be colored and to assign colors to the selected regions in terms of control. However, it is possible to perform style and color transfer to certain regions of the image. For example, Ding et al. applied segmentation separated foreground objects. Then, they applied style transfer to the background and color transfer to the foreground objects [43].
- *Unnatural image outputs:* As a result of the neural style and color transfer, problems such as unnatural appearance of the output image, low image quality and resolution, or blurred image may be encountered [10], [12], [19], [21]. Such problems cause the images to look like unnatural. Figure 4 shows the unnatural images obtained after NST.
- *Color tone differences*: When transferring color and style from the source image to the target image, problems such as mismatched color tones can occur [10], [12]. Color and style transfer algorithms usually try to match the color palette of the target image with the color palette of the source image. However, in this process, some color tones of the style image may be lost, or color deviations may occur. Figure 5 shows the color tone differences in the target image after NST.

Fig. 4. Unnatural images obtained after NST [21]

Figure 4 shows green color tones in the cloud parts of the output image, so an artificial image was obtained. The output image was generated using the CNN algorithm.

Fig. 5 Image obtained after NST [10]

Analyzing the images in Figure 5, it is observed that the colors in the style image and the colors in the output image do not match each other in terms of tone, and some colors in the reference image are not included in the output image.

- *Visual fidelity problem:* The concept of fidelity can be defined as the accuracy with which the output image reflects the scene in the content image and the color distribution in the style image [18]. The more the output image obtained after the style transfer process resembles the style and content images, the higher the visual fidelity. This issue has been addressed in the literature, particularly in [18], [44], [45].

V. CONCLUSION

In this study, the studies in the literature on color and neural style transfer are discussed within the scope of image colorization. When the studies in the literature are examined, it has been seen that style and color transfer processes have various limitations and difficulties. The literature has encountered difficulties such as generating unnatural images, loss of fine details, difficult control, high computational cost, and color tone differences. To overcome these challenges, real-time, efficient methods can be developed that optimize computational costs and take user feedback into account. New research can be done on these issues. This literature review is

intended to guide researchers who want to work in this field. Future works will aim to overcome these challenges by proposing innovative algorithms and techniques that balance efficiency and quality.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics

REFERENCES

[1] Q. Cai, M. Ma, C. Wang, and H. Li, "Image Neural Style Transfer: A Review," *Computers and Electrical Engineering*, vol. 108, p. 108723, 2023.

[2] L. Jiao and J. Zhao, "A survey on the new generation of deep learning in image processing," *IEEE Access*, vol. 7, pp. 172231–172263, 2019.

[3] "Neural Style Transfer, Image Color Transfer." Accessed: Sep. 15, 2024. [Online]. Available: https://www.scopus.com/search/

[4] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2414–2423.

[5] Y. Jing, Y. Yang, Z. Feng, J. Feng, Y. Yu, and M. Song, "Neural Style Transfer: A Review," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 11, pp. 3365–3385, 2019.

[6] A. Singh, V. Jaiswal, G. Joshi, A. Sanjeeve, S. Gite, and K. K., "Neural Style Transfer: A Critical Review," *IEEE Access*, vol. 9, pp. 131583–131613, 2021.

[7] G. Sohaliya and K. Sharma, "An Evolution of Style Transfer from Artistic to Photorealistic: A Review," in *2021 Asian Conference on Innovation in Technology, ASIANCON 2021*, Institute of Electrical and Electronics Engineers Inc., Aug. 2021. doi: 10.1109/ASIANCON51346.2021.9544924.

[8] J. W. Johnson, "Towards the Algorithmic Detection of Artistic Style," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 1, pp. 76–81, 2019.

[9] L. A. Gatys, A. S. Ecker, and M. Bethge, "A Neural Algorithm of Artistic Style," *arXiv preprint arXiv:1508.06576*, 2015.

[10] L. Zhang, Y. Ji, X. Lin, and C. Liu, "Style Transfer for Anime Sketches with Enhanced Residual U-Net and Auxiliary Classifier GAN," in *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, 2017, pp. 506–511.

[11] B. Karadağ, A. Arı, and M. Karadağ, "Derin Öğrenme Modellerinin Sinirsel Stil Aktarımı Performanslarının Karşılaştırılması," *Journal of Polytechnic*, vol. 24, no. 4, pp. 1611–1622, 2021, doi: 10.2339/politeknik.885838.

[12] J. Lian and J. Cui, "Anime Style Transfer with Spatially-Adaptive Normalization," in *2021 IEEE International Conference on Multimedia and Expo (ICME)*, 2021, pp. 1–6.

[13] L. JinKua, C. Yang, and H. B. Abdalla, "Enhanced Style Transfer with Colorization and Super-Resolution," in *2022 7th International Conference on Communication, Image and Signal Processing (CCISP)*, 2022, pp. 166–172.

[14] Z. Ke, Y. Liu, L. Zhu, N. Zhao, and R. W. Lau, "Neural Preset for Color Style Transfer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 14173–14182.

[15] J. J. Virtusio, J. J. Ople, D. S. Tan, T. M., N. Kumar, and K. L. Hua, "Neural Style Palette: A Multimodal and Interactive Style Transfer from a Single Style Image," *IEEE Transactions on Multimedia*, vol. 23, pp. 2245–2258, 2021.

[16] Y. Deng *et al.*, "Stytr2: Image Style Transfer with Transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11326–11336.

[17] X. Fu, "Digital Image Art Style Transfer Algorithm Based on CycleGAN," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, p. 6075398, 2022.

[18] S.-H. Huang, J.-A. An, D. Wei, J. Luo, and H. Pfister, "QuantArt: Quantizing Image Style Transfer Towards High Visual Fidelity," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5947–5956.

[19] T. T. Fang, D. M. Vo, A. Sugimoto, and S.-H. Lai, "Stylized-Colorization for Line Arts," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 2033–2040.

[20] C. Zheng and Y. Zhang, "Two-Stage Color Ink Painting Style Transfer via Convolution Neural Network," in *2018 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN)*, 2018, pp. 193–200.

[21] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep Photo Style Transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4990–4998.

[22] X. Wang, G. Oxholm, D. Zhang, and Y.-F. Wang, "Multimodal Transfer: A Hierarchical Deep Convolutional Neural Network for Fast Artistic Style Transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5239–5247.

[23] J. Cui, "Image Style Migration Algorithm Based on HSV Color Model," in *2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, 2022, pp. 111–114.

[24] Y. Liao and Y. Huang, "Deep Learning-Based Application of Image Style Transfer," *Mathematical Problems in Engineering*, vol. 2022, no. 1, p. 1693892, 2022.

[25] X. Liu, X. Li, M.-M. Cheng, and P. Hall, "Geometric Style Transfer," *arXiv preprint arXiv:2007.05471*, 2020.

[26] X. Han, Y. Wu, and R. Wan, "A Method for Style Transfer from Artistic Images Based on Depth Extraction Generative Adversarial Network," *Applied Sciences*, vol. 13, no. 2, p. 867, 2023.

[27] C. Lv, D. Zhang, S. Geng, Z. Wu, and H. Huang, "Color Transfer for Images: A Survey," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 20, no. 8, pp. 1–29, 2024.

[28] L. Bao, K. Panetta, and S. Agaian, "Fast Color Transfer for Camouflage Applications," in *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, 2017, pp. 1–5.

[29] S. Liu, "An Overview of Color Transfer and Style Transfer for Images and Videos," *arXiv preprint arXiv:2204.13339*, 2022.

[30] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color Transfer Between Images," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34–41, 2001.

[31] A. Abadpour and S. Kasaei, "An Efficient PCA-Based Color Transfer Method," *Journal of Visual Communication and Image Representation*, vol. 18, no. 1, pp. 15–34, 2007.

[32] X. An and F. Pellacini, "User-Controllable Color Transfer," in *Computer Graphics Forum*, 2010, pp. 263–271.

[33] B. Arbelot, R. Vergne, T. Hurtut, and J. Thollot, "Local Texture-Based Color Transfer and Colorization," *Computer & Graphics*, vol. 62, pp. 15–27, 2017.

[34] B. Xu, X. Liu, C. Lu, T. Hong, and Y. Zhu, "Transferring the Color Imagery from an Image: A Color Network Model for Assisting Color Combination," *Color Research and Application*, vol. 44, no. 2, pp. 205–220, 2019.

[35] C. Gu, X. Lu, and C. Zhang, "Example-Based Color Transfer with Gaussian Mixture Modeling," *Pattern Recognition*, vol. 129, p. 108716, 2022.

[36] X. Xiao and L. Ma, "Gradient-Preserving Color Transfer," in *Computer Graphics Forum*, 2009, pp. 1879–1886.

[37] J. Lee, H.-Y. Son, G. Lee, J. Lee, S.-H. Cho, and S. Lee, "Deep Color Transfer Using Histogram Analogy," *Visual Computer*, vol. 36, pp. 2129–2143, 2020.

[38] J. Yin, Y.-C. Huang, B.-H. Chen, and S.-Z. Ye, "Color Transferred Convolutional Neural Networks for Image Dehazing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 3957–3967, 2019.

[39] D. Liu, Y. Jiang, M. Pei, and S. Liu, "Emotional image color transfer via deep learning," *Pattern Recognition Letters*, vol. 110, pp. 16–22, Jul. 2018, doi: 10.1016/j.patrec.2018.03.015.

[40] M. Zhang, J. Liao, and J. Yu, "Deep exemplar-based color transfer for 3d model," *IEEE Trans Vis Computers & Graphics*, vol. 28, no. 8, pp. 2926–2937, 2020.

[41] Q. C. Tian and L. D. Cohen, "Histogram-Based Color Transfer for Image Stitching," *Journal of Imaging*, vol. 3, no. 3, p. 38, 2017.

[42] Y. Qian, D. Liao, and J. Zhou, "Manifold Alignment Based Color Transfer for Multiview Image Stitching," in *Proceedings of the IEEE International Conference on Image Processing*, 2013.

[43] Z. Ding, P. Li, Q. Yang, S. Li, and Q. Gong, "Regional Style and Color Transfer," in *2024 5th International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, 2024, pp. 593–597.

[44] I. Luengo, E. Flouty, P. Giataganas, P. Wisanuvej, J. Nehme, and D. Stoyanov, "SurReal: Enhancing surgical simulation realism using style transfer," *arXiv preprint arXiv:1811.02946*, 2018.

[45] W. H. Png, Y. Aun, and M. L. Gan, "FeaST: Feature-guided Style Transfer for high-fidelity art synthesis," *Computers & Graphics*, p. 103975, 2024.

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 92-98
DOI: 10.36287/ijmsit.8.2.8
Received: October 16, 2024; Accepted: November 28, 2024

Metin Sınıflandırmaya Karşı Kriptografi Yöntemlerinin Kullanılması

Ahmet Emre ERGÜN [1*], Özgü CAN [2]

[1*] *İzmir Kâtip Çelebi Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye (ahmetemreergun95@gmail.com) (ORCID: 0000-0002-3025-5640)*

[2] *Ege Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye (ozgu.can@ege.edu.tr) (ORCID: 0000-0002-8064-2905)*

Özet **–** Bu makale, makine öğrenmesi sınıflandırma algoritmalarına karşı verilerin gizliliğini sağlamak için kriptografik tekniklerin nasıl kullanılabileceğini araştırmaktadır. Çalışma, ruh sağlığı sorunlarıyla ilgili metin ve etiket sütunlarını içeren Mental Health Corpus veri kümesine odaklanmaktadır. Metinleri sınıflandırmak için Rastgele Orman (*Random Forest*, RF), Karar Ağacı (*Decision Tree*, DT) ve Destek Vektör Makinesi (*Support Vector Machine*, SVM) sınıflandırma algoritmaları kullanılmıştır. Sınıflandırma doğruluğunu azaltmak için ise kriptografi yöntemi olan karakter kaydırma (*shift*) uygulanmaktadır. Sonuçlar, karakter kaydırmalarının sınıflandırıcı doğruluğunu büyük ölçüde azalttığını, 1 karakter kadar küçük kaydırmaların tüm modellerde doğruluğu %30'dan fazla azalttığını göstermektedir. Bulgular, kriptografik yöntemlerin, özellikle hassas bilgilerin söz konusu olduğu çeşitli alanlarda veri gizliliğini ve güvenliğini artırma potansiyelini göstermektedir.

Anahtar Kelimeler – gizlilik, makine öğrenmesi, kriptografi, kaydırma, ikame

Atıf: Ergün A.E., Can, Ö. (2024). Metin Sınıflandırmaya Karşı Kriptografi Yöntemlerinin Kullanılması. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 92-98.

Using Cryptography Methods Against Text Classification

Abstract – This article explores how cryptographic techniques can be used to ensure the confidentiality of data against machine learning classification algorithms. The study focuses on the Mental Health Corpus dataset, which contains text and tag columns related to mental health issues. Random Forest (RF), Decision Tree (DT) and Support Vector Machine (SVM) classification algorithms were used to classify the texts. To reduce classification accuracy, character shift, which is a cryptography method, is applied. Results show that character shifts greatly reduce classifier accuracy, with shifts as small as 1 character reducing accuracy by more than 30% across all models. The findings demonstrate the potential of cryptographic methods to increase data confidentiality and security in a variety of areas, especially where sensitive information is involved.

Keywords – *confidentiality, machine learning, cryptography, shift, substitution*

Citation: Ergün A.E., Can, Ö. (2024). Using Cryptography Methods Against Text Classification. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 92-98.

I. GİRİŞ

Büyük veri ve yapay zeka çağında, büyük veri kümelerini analiz etme ve bunlardan içgörü elde etme becerisi giderek daha değerli hale gelmiştir. Bu kabiliyet, veriye dayalı kararların önemli ilerlemelere ve verimliliklere yol açabileceği sağlık, finans ve pazarlama gibi alanlar için derin etkilere sahiptir [14], [17]. Bununla birlikte, özellikle ruh sağlığı kayıtları gibi hassas bilgilerle uğraşırken güvenlik riski artmaktadır. Bu tür verilere yetkisiz erişim, ayrımcılık, damgalama ve kimlik hırsızlığı gibi ciddi etik, yasal ve sosyal sonuçlara yol açabilmektedir [10], [12]. Kriptografik teknikler bu risklere karşı bir çözüm sunmaktadır. Kriptografi, verileri yetkisiz taraflar için anlamsız hale getirecek şekilde gizleyerek hassas bilgilerin istismar edilmesini önleyebilmektedir.

Ruh sağlığı verileri gibi hassas veriler mahremiyet açısından önemlidir. Sınıflandırma algoritlamaları kullanılarak hassas veriler sınıflandırılabilir ve bu verilere sahip kişilerin ruh sağlığı sorunu olup olmadığı analiz edilebilmektedir. Hassas metinler kriptografik yöntemler kullanılarak anlamsız hale getirilebilmektedir. Bu çalışma, metinsel verileri sınıflandırma saldırılarından korumak için basit ama etkili bir teknik olan kriptografik kaydırmaların (*substitution*) uygulanmasını araştırmaktadır. Karakter kaydırmaları ile, metinleri sistematik olarak değiştirerek, bu kriptografik yöntemlerin yaygın sınıflandırma algoritmalarının performansını nasıl

düşürebileceğini belirlemek ve böylece veri gizliliğini ve güvenliğini artırmak amaçlanmaktadır.

Bu çalışmadaki deneyler, ruh sağlığı sorunlarıyla ilgili metin ve etiket sütunlarını içeren Mental Health Corpus veri kümesine odaklanmaktadır. Veri seti İngilizce metinlerden oluştuğu için çalışmada kullanılan kaydırma yöntemi İngiliz alfabesine göre yapılmıştır. İngiliz alfabesinde toplam 26 harf bulunmaktadır. Bir harf 1 kere kaydırılırsa alfabedeki sıraya göre bir sonraki harfi almaktadır. Eğer bir harfe 0 veya 26 kere kaydırma yapılırsa kendisine geri döner ve harf değişmez. Bu sebeple 0 veya 26 harf kaydırma uygulanması güvenliği sağlamak açısından anlamsız olmaktadır. Bu çalışmadaki deneylerde kullanılan veri setinin %80'i eğitim, %20'si test verisi olarak kullanılmıştır. Deneylerdeki kaydırma işlemi 0'dan 25'e kadar olan sayılarda yapılmıştır. Bu sebeple, RF, DT ve SVM sınıflandırma algoritmalarının performansları 26 farklı deney üzerinde denenmiştir. Sonuçlar 4 farklı metrikte kıyaslanmıştır. Çalışmada, SVM algoritmasının DT ve RF algoritmalarına göre daha etkin olduğu görülmüştür.

Bu araştırmanın temel amacı, Mental Health Corpus veri kümesi üzerinde makine öğrenmesi sınıflandırıcılarının doğruluğunu azaltmada kriptografik yöntemlerin, özellikle de karakter kaydırmanın etkinliğini değerlendirmektir. Bu çalışma, karakter kaydırma yoluyla metinleri sistematik olarak değiştirerek, bu kriptografik tekniklerin yaygın sınıflandırma algoritmalarının performansını ne ölçüde düşürebileceğini belirlemeyi amaçlamaktadır. Nihai hedef, hassas verileri yetkisiz makine öğrenmesi analizinden koruyabilecek stratejiler geliştirmek ve böylece veri gizliliğini ve güvenliğini artırmaktır.

Bu çalışmadaki bölümler aşağıdaki şekilde yapılandırılmıştır: İkinci bölümde, mevcut literatürdeki çalışmalar ve ilgili araştırmalar gözden geçirilecektir. Üçüncü bölümde, araştırmada kullanılan metodoloji ve deneysel metrikler detaylandırılacaktır. Dördüncü bölümde, deneylerin sonuçları ve bulgular tartışılacaktır. Beşinci bölümde, elde edilen sonuçlar ışığında genel bir değerlendirme yapılacaktır. Altıncı bölümde, çalışmanın sonuçları ve gelecek araştırmalar için öneriler sunulacaktır. Son olarak çalışmada başvurulan kaynaklar listelenecektir.

II. LİTERATÜR

Literatürdeki bazı çalışmalar, ruh sağlığı bozuklukları için metin sınıflandırmasının kullanımını araştırmıştır. Sarno ve arkadaşları [16] ile Ameer ve arkadaşları [1], sosyal medya verilerine çok sınıflı sınıflandırma algoritmaları uygulamış, Sarno ve arkadaşları Mekanik Kontrol Tabanlı Makine Öğrenmesi (*Mechanical Control-Based Machine Learning*, MCML) algoritmasını kullanarak yüksek doğruluk elde ederken Ameer ve arkadaşları derin öğrenme ve transfer öğrenme modellerine odaklanmıştır. Abusaa ve arkadaşları [4], yazıya dökülmüş konuşma örneklerine dayanarak ruh sağlığı sorunlarını sınıflandırmak için makine öğrenmesi tekniklerini kullanmış ve şizofreni için yüksek doğruluk elde etmiştir. Ive ve arkadaşları [8], ruh sağlığıyla ilgili sosyal medya metinlerini sınıflandırmak için dikkat mekanizmalarına sahip hiyerarşik bir sinir modeli sunmuş ve geleneksel yöntemlere kıyasla daha iyi sonuçlar elde etmiştir. Bu çalışmalar toplu olarak ruh sağlığı analizinde metin sınıflandırmanın potansiyelini vurgulamaktadır.

Metin sınıflandırma, metin madenciliğinin önemli bir yönüdür ve bu amaçla çeşitli yöntemler ve sınıflandırıcılar kullanılmaktadır [7]. Zhang ve diğerleri [18], metin sınıflandırmada SVM ve Geri Yayılım Sinir Ağının (*Back Propagation Neural Network,* BPNN) performansını karşılaştırmış ve SVM'nin çok sınıflı sınıflandırmada daha iyi performans gösterdiğini bulmuştur. Kamruzzaman ve diğerleri [9], veri madenciliği kullanarak metin sınıflandırması için daha az eğitim gerektiren ve özellikleri türetmek için kelime ilişkilendirme kurallarını kullanan yeni bir algoritma tanıtmıştır. Arunachalam ve diğerleri [2] Bayesian sınıflandırması, Latent Dirichlet Allocation (*Gizli Dirichlet Tahsisi,* LDA) sınıflandırması, Dinamik Ontoloji Sınıflandırması ve Genetik Algoritma dahil olmak üzere duygu kutupluluğu tespiti için metin sınıflandırma tekniklerini tartışmıştır. Bu çalışmalar, metin sınıflandırmada kullanılan çeşitli yöntem ve tekniklere kapsamlı bir genel bakış sağlamaktadır.

Son araştırmalar, kaydırma ikamesi (*shifting substitution*) yoluyla metin şifrelemenin güvenliğini artırmak için çeşitli yöntemler önermiştir. Verma ve arkadaşları [13], Sezar şifresinin küçük bir dönüşümü olan modellenmiş kaydırma şifresini sunmaktadır. Shareef ve diğerleri [15], karakterleri bir anahtar değerine göre yeniden düzenleyen ve yaygın kriptografi saldırılarına karşı dirençli hale getiren bir şifreli metin kaydırma algoritması tasarlayarak bu tekniği daha da geliştirmiştir. Ambulkar [6], metin şifreleme için (Bilgi Değişimi İçin Amerikan Standart Kodlama Sistemi) (*American Standard Code for Information Interchange*, ASCII) değerleri oluşturmak üzere genetik algoritmalarla birlikte çoklu ikame (*shift*) yöntemlerinin kullanımını araştırmış ve şifreleme sürecine ekstra bir karmaşıklık katmanı eklemiştir. Bu çalışmalar toplu olarak, değişen ikame yöntemleri yoluyla metin şifrelemeyi geliştirme potansiyelini göstermektedir.

Metin manipülasyonunda ikame ve kaydırma tekniklerinin kullanımı çeşitli alanlarda yaygın bir uygulamadır. Borowiak ve diğerleri [3], karmaşık metin değiştirmeleri için PRXCHANGE işlevindeki düzenli ifadelerin gücünü vurgulamıştır. Li ve diğerleri [11], mobil cihazlarda metin revizyonu için değiştirme tabanlı bir teknik olan Swap'ı tanıtmış ve hassas işaret kontrolü ve tekrarlayan geri tuşuna basma ihtiyacını azaltmıştır. Son olarak, Pal ve diğerleri [5], metin mesajlarının güvenliğini sağlamak için kaydırma şifresi ve ikame şifresi gibi klasik kriptografi yöntemlerinin kullanımını araştırmıştır. Bu çalışmalar toplu olarak, metin manipülasyonu, çevirisi ve güvenliğinde kaydırma tekniklerinin öneminin altını çizmektedir.

III.METODOLOJİ

A. Veri Seti

Bu çalışmada kullanılan Mental Health Corpus veri kümesi, her biri bir metin ve buna karşılık gelen bir etiket içeren 27.977 girdiden oluşmaktadır. Metinler anksiyete, depresyon ve diğer ilgili durumlar gibi çeşitli ruh sağlığı sorunlarını tartışan bireylerin gönderilerini içermektedir. Veri kümesinde 0 ve 1 olmak üzere iki etiket bulunmaktadır. 0 ruh sağlığı sorunu olduğunu temsil eder, 1 ise ruh sağlığı sorunu olmadığını temsil etmektedir. Bu veri kümesi, yetkisiz erişim ve analizden korunması gereken hassas bilgiler içerdiğinden bu çalışma için özellikle uygundur.

B. Makine Öğrenmesi Algoritmaları

Saldırı senaryosunu simüle etmek için üç popüler makine öğrenmesi sınıflandırma algoritması kullanılmıştır:

- Rastgele Orman (RF): Eğitim sırasında birden fazla karar ağacı oluşturan ve sınıflandırma için sınıfların modunu çıkaran bir topluluk öğrenme yöntemidir. Bu algoritma sağlamlığı ve yüksek doğruluğu ile bilinir, bu da onu kriptografik tekniklerin etkinliğini test etmek için uygun bir aday haline getirmektedir.
- Karar Ağacı (DT): Kararların ağaç benzeri bir grafiğini ve tesadüfi olay sonuçları da dahil olmak üzere olası sonuçlarını kullanan bir modeldir. Karar ağaçlarının yorumlanması ve anlaşılması kolaydır, bu nedenle sınıflandırma görevlerinde sıklıkla kullanılırlar.
- Destek Vektör Makinesi (SVM): Bir veri kümesini sınıflara en iyi şekilde ayıran hiper düzlemi bularak sınıflandırma için verileri analiz eden denetimli bir öğrenme modelidir. SVM'ler yüksek boyutlu uzaylarda etkilidir ve metin sınıflandırma görevleri için yaygın olarak kullanılmaktadır.

Bu algoritmalar, metin sınıflandırma görevlerinde yaygın kullanımları ve etkinlikleri nedeniyle seçilmiştir. Çalışma, kriptografik tekniklerin bu sınıflandırıcılar üzerindeki etkisini değerlendirerek, yöntemlerin etkinliğinin kapsamlı bir değerlendirmesini sağlamayı amaçlamaktadır.

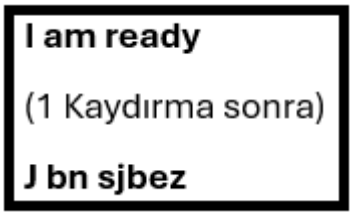
C. Kriptografi Yöntemleri

Bu çalışmada uygulanan kriptografik teknik karakter kaydırmadır. Bu, her bir karakteri alfabede sabit sayıda pozisyon kaydırarak metinleri değiştirmeyi içermektedir. Örneğin, 1'lik bir kaydırma 'a'yı 'b'ye, 'b'yi 'c'ye dönüştürür ve bu böyle devam eder. Tablo 1'de 1'er kaydırma uygulandığında harflerin alacağı yeni harfler gösterilmiştir.

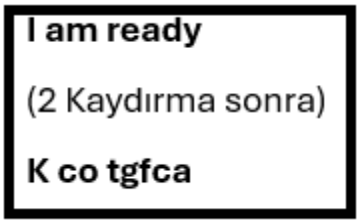
Tablo 1. *1'er Harf Kaydırmaya Göre Alfabetik Karşılık*

1 Kaydırmadan Önce	**1 Kaydırmadan Sonra**
a	b
b	c
c	d
d	e
e	f
f	g
g	i
i	j
j	k
k	l
l	m
m	n
n	o
o	p
p	q
q	r
r	s
s	t
t	u
u	v
v	w
w	x
x	y
y	z
z	a

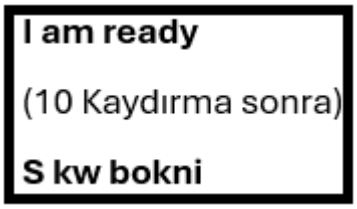
Şekil 1'de 1'er, 2'şer ve 10'ar kaydırma yapılmadan önce ve sonraki metinler gösterilmiştir. Metindeki her harf alfabetik olarak kaydırma sayısı kadar harf sonraki harfi almıştır.



a) 1'er Harf Kaydırma



b) 2'şer Harf Kaydırma



c) 10'ar Harf Kaydırma

Şekil 1. *Metindeki Harfleri Kaydırma*

Çalışmada sınıflandırma doğruluğu üzerindeki etkilerini değerlendirmek için 0 ila 25 arasında değişen kaydırmalar test edilmiştir. Bu basit ama etkili metin değiştirme yöntemi, farklı karakter kaydırma derecelerinin sınıflandırma algoritmalarını nasıl karıştırabileceğini değerlendirmek için kullanılmıştır.

D. Ölçü Metrikleri

Kriptografik yöntemlerin etkinliği dört temel ölçüt kullanılarak ölçülmüştür:

- Doğruluk: İncelenen toplam veri sayısı içindeki doğru sonuçların oranını ifade eder. Bu oran, hem Doğru Pozitifler (*True Positives*, TP), modelin doğru bir şekilde "pozitif" olarak tahmin ettiği durumlar hem de Doğru Negatifler (*True Negatives*, TN), modelin doğru bir şekilde "negatif" olarak tahmin ettiği durumlar göz önünde bulundurularak hesaplanır. Ayrıca, Yanlış Pozitifler (*False Positives*, FP), modelin "pozitif" olarak tahmin ettiği ancak gerçekte "negatif" olan durumlar ve Yanlış Negatifler (*False Negatives*, FN), modelin "negatif" olarak tahmin ettiği ancak gerçekte "pozitif" olan durumlar bu değerlendirmede dikkate alınır. Doğruluk, sınıflandırıcının etkinliğinin genel bir ölçüsünü sağlamaktadır. Doğruluk oranı hesaplama formülü (1)'de belirtilmiştir.

$$Doğruluk = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- Kesinlik (*Precision*): Pozitif olarak sınıflandırılan toplam vaka sayısı içinde doğru pozitiflerin oranıdır. Kesinlik, sınıflandırıcının ilgili vakaları tanımlamadaki performansı hakkında fikir vermektedir. Kesinlik oranı hesaplama formülü (2)'de belirtilmiştir.

$$Kesinlik = \frac{TP}{TP + FP} \quad (2)$$

- Geri Çağırma (*Recall*): Gerçek pozitiflerin toplam sayısı içinde gerçek pozitiflerin oranıdır. Geri çağırma, sınıflandırıcının tüm ilgili verileri belirleme yeteneğini göstermektedir. Geri çağırma hesaplama formülü (3)'te belirtilmiştir.

$$Geri\ Çağırma = \frac{TP}{TP + FN} \quad (3)$$

- F1-Skoru: Kesinlik ve geri çağırmanın harmonik ortalamasıdır ve iki metrik arasında bir denge sağlamaktadır. F1-Skoru özellikle sınıf dağılımı dengesiz olduğunda kullanışlıdır. F1-Skoru hesaplama formülü (4)'te belirtilmiştir.

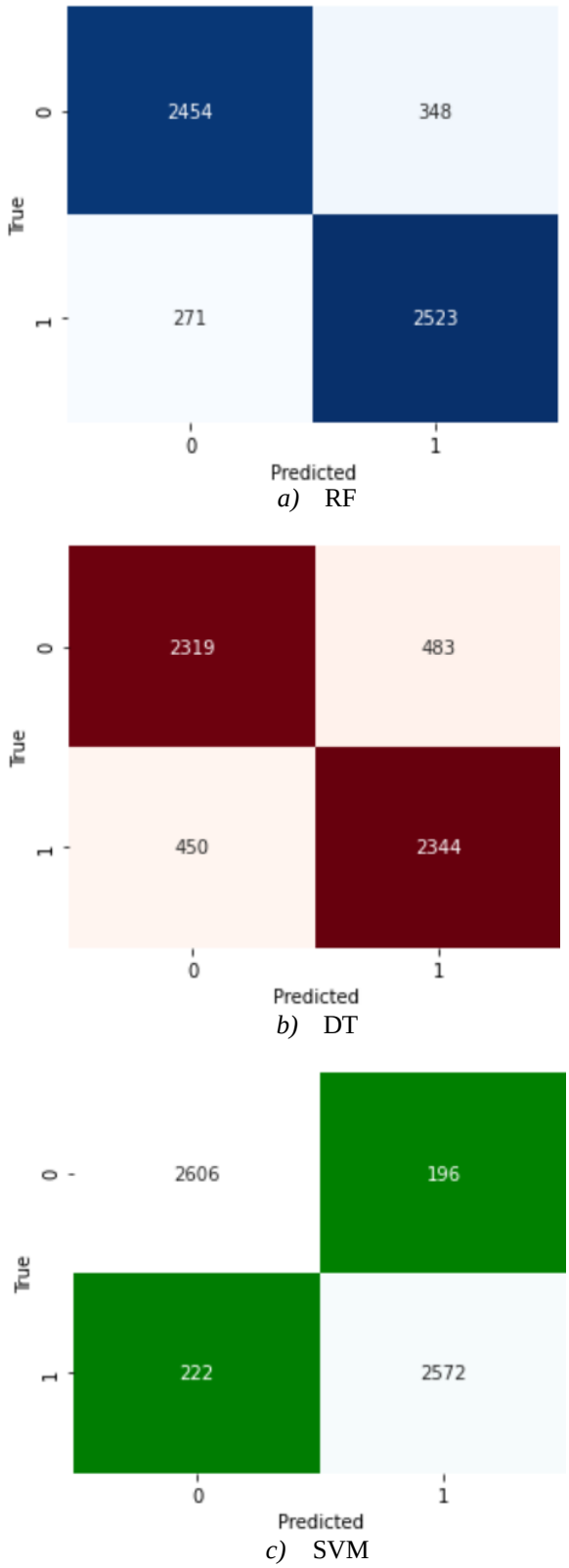
$$F1 - Skoru = 2.\frac{Kesinlik.Geri\ Çağırma}{Kesinlik + Geri\ Çağırma} \quad (4)$$

Bu metrikler, sınıflandırıcıların performansının kapsamlı bir değerlendirmesini sağlayarak kriptografik tekniklerin etkisinin ayrıntılı bir şekilde değerlendirilmesine olanak tanımaktadır.

IV. DENEY SONUÇLARI

Deneysel çalışmalar kapsamında 3 farklı makine öğrenmesi algoritması ve 26 farklı kaydırma durumu denenmiştir. Sonuçlar, RF, DT ve SVM sınıflandırıcılarının farklı kaydırmalardaki performansını göstermektedir. Değiştirilmemiş bir veri kümesi için (kaydırma 0), RF %88,94 doğruluk, %88,97 kesinlik, %88,94 geri çağırma ve %88,94 F1-Skoru elde etmiştir. DT %83,33 doğruluk, %83,33 hassasiyet, %83,33 geri çağırma ve %83,33 F1-Skoru elde etmiştir. SVM %92,53 doğruluk, %92,53 kesinlik, %92,53 geri çağırma ve %92,53 F1-Skoru elde etmiştir.

Harfler 1 karakter kaydırıldığında, tüm sınıflandırıcıların performansı önemli ölçüde düşmektedir. RF'nin doğruluğu %50,09'a, kesinliği %55,03'e, geri çağırma oranı %50,09'a ve F1-Skoru %33,52'ye düşmektedir. DT'nin doğruluğu %52,57'ye, kesinliği %62,92'ye, geri çağırma oranı %52,57'ye ve F1-Skoru %40,60'a düşmektedir. SVM'nin doğruluğu %52,36'ya, hassasiyeti %62,56'ya, geri çağırma oranı %52,36'ya ve F1-Skoru %40,11'e düşmektedir. Şekil 2'de makine öğrenmesi algoritmalarının 0 kaydırmada karışıklık matrisleri gösterilmektedir. SVM'in DT'ye RF'ye göre daha çok doğru pozitif ve doğru negatif sayısına sahip olduğu görülmektedir.



a) RF

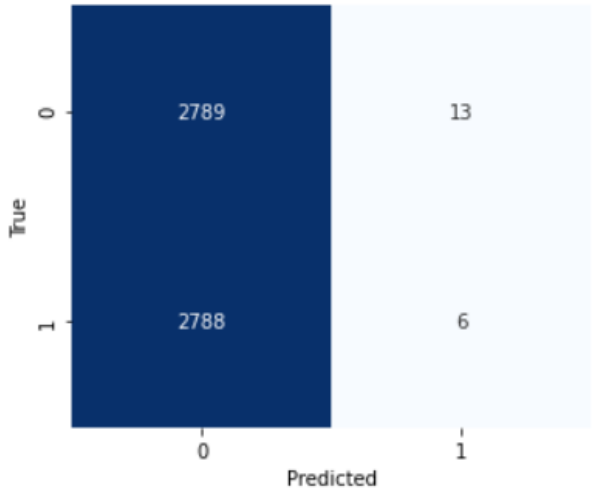
b) DT

c) SVM

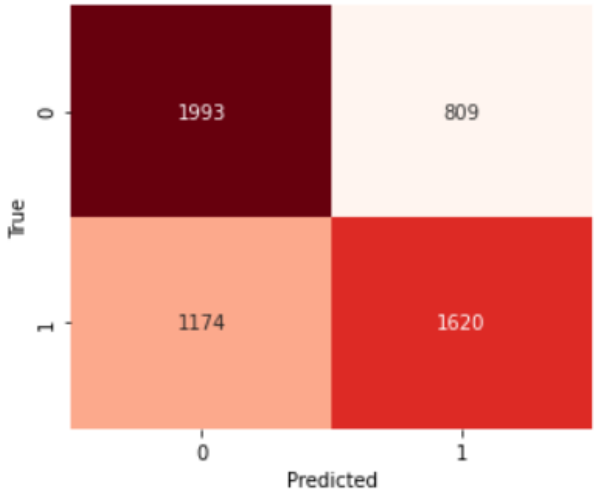
Şekil 2. *0 Kaydırma İçin Karışıklık Matrisleri*

Daha fazla kaydırma da sınıflandırıcıların performansını düşürmektedir. Örneğin, 5 karakterlik bir kaydırma ile RF %51,55 doğruluk, %68,13 kesinlik, %51,55 geri çağırma ve %37,11 F1-Skoru elde etmiştir. DT %51,89 doğruluk, %68,69 kesinlik, %51,89 geri çağırma ve %37,87 F1-Skoru elde etmektedir. SVM %52,11 doğruluk, %62,94 kesinlik, %52,11 geri çağırma ve %39,32 F1-Skoru elde etmektedir.

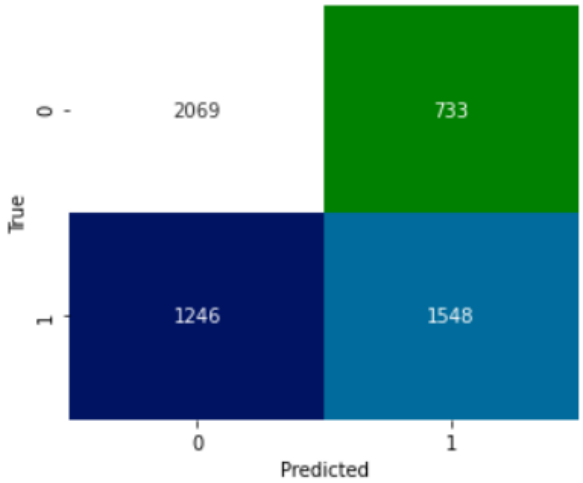
10 karakterlik bir kaydırmada DT'nin doğruluğu %64,56'ya, kesinliği %64,81'e, geri çağırma oranı %64,56'ya ve F1-Skoru %64,41'e düşmüştür. SVM %64,64 doğruluk, %65,14 kesinlik, %64,64 geri çağırma ve %64,33 F1-Skoru elde etmiştir. RF'nin performansı %49,95 doğruluk, %40,81 kesinlik, %49,95 geri çağırma ve %33,55 F1-Skoru ile nispeten sabit kalmaktadır. Şekil 3'te makine öğrenmesi algoritmalarının 10 kaydırmada karışıklık matrisleri gösterilmektedir. SVM ve DT'nin RF'ye göre daha çok doğru pozitif ve doğru negatif sayısına sahip olduğu görülmektedir.



a) RF



b) DT



c) SVM

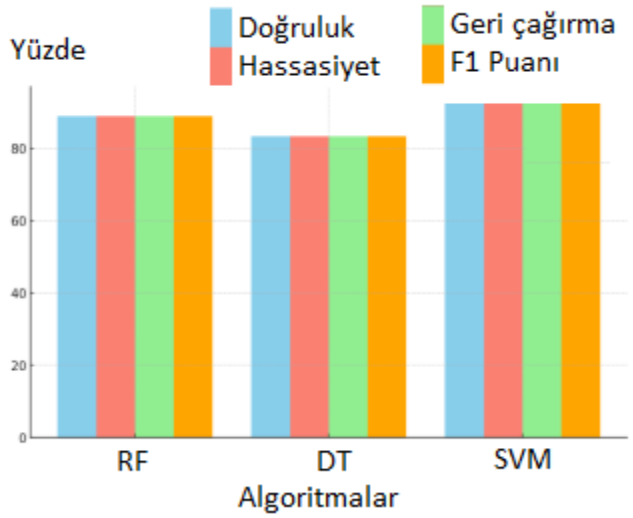
Şekil 3. 1*0 Kaydırma İçin Karışıklık Matrisleri*

Tablo 2'de 0'dan 25'e kadar kaydırmaya karşı makine öğrenmesi algoritmalarının performansları gösterilmektedir. SVM algoritması kaydırma olan ve olmayan durumlarda en yüksek doğruluk oranlarına sahiptir. Kaydırma kullanılan durumlarda her algoritmanın doğruluk oranı yaklaşık olarak %30 oranında düşürülmüştür. Genel olarak, kaydırma arttıkça sınıflandırıcıların performansı dalgalanmakta, ancak değiştirilmemiş veri kümesindeki performanslarından daha düşük kalmaktadır. Bu, karakter kaydırmanın makine öğrenmesi sınıflandırıcılarının doğruluğunu etkili bir şekilde azaltabileceğini ve hassas metinsel verileri yetkisiz sınıflandırma girişimlerinden koruyabileceğini göstermektedir.

Aşağıda Şekil 4'te, RF, DT ve SVM algoritmalarının kaydırmasız durumdaki doğruluk oranları karşılaştırılmaktadır. Kaydırmasız durumda, SVM %92,53 doğruluk ile en iyi performansı sergileyerek, doğrusal olmayan verileri sınıflandırma konusundaki güçlü yeteneğini ortaya koymaktadır. RF ise %88,94 doğrulukla oldukça etkili bir performans sunmaktadır; ağaç tabanlı yapısı, verinin önemli desenlerini ve ilişkilerini öğrenerek doğru sınıflandırmalar yapılmasını sağlamaktadır. DT'nin doğruluğu %83,33 ile RF ve SVM algoritmalarının gerisinde kalmaktadır; bu durum, karar ağaçlarının veri üzerinde aşırı uyum yaparak genelleme yeteneklerini zayıflatmasından kaynaklanmaktadır. Bununla birlikte, üç algoritma için de F1 skoru, hassasiyet, geri çağırma ve doğruluk oranları birbirine çok yakındır, bu da modellerin genel olarak iyi çalıştığını ve veriye uygun sınıflandırmalar gerçekleştirdiğini göstermektedir.

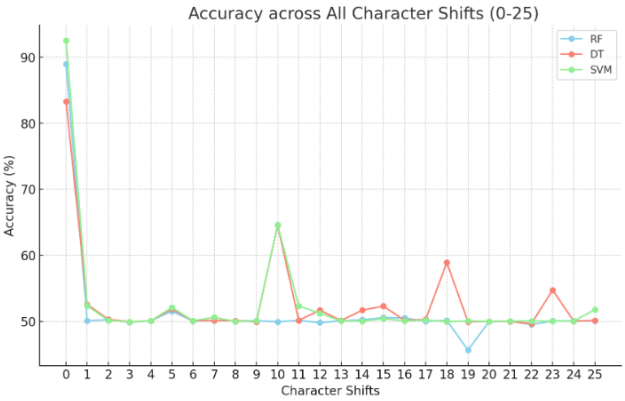
Tablo 2. *Algoritmaların Kaydırmalara Göre Doğruluk Oranları*

Kaydırma	RF Doğruluk Oranı	DT Doğruluk Oranı	SVM Doğruluk Oranı
0	%88,94	%83,33	**%92,53**
1	%50,09	%52,57	%52,36
2	%50,25	%50,30	%50,14
3	%49,89	%49,95	%49,93
4	%50,09	%50,07	%50,07
5	**%51,55**	%51,89	%52,11
6	%50,05	%50,09	%50,09
7	%50,13	%50,13	%50,63
8	%50,07	%50,11	%49,98
9	%50,09	%49,95	%50,07
10	%49,95	**%64,56**	**%64,64**
11	%50,13	%50,16	%52,34
12	%49,82	%51,70	%51,23
13	%50,07	%50,13	%50,07
14	%50,25	%51,70	%50,04
15	%50,57	%52,31	%50,41
16	%50,52	%50,16	%50,07
17	%50,02	%50,30	%50,21
18	%50,13	%58,92	**%49,96**
19	**%45,66**	%49,95	%50,07
20	%49,98	%50,00	%49,98
21	%50,05	%50,00	%50,04
22	%49,52	**%49,66**	%50,04
23	%50,07	%54,75	%50,04
24	%50,07	%50,05	%50,05
25	%50,07	%50,13	%51,79



Şekil 4. *Kaydırmasız Durumda Algoritmaların Performansı*

Şekil 5'te, RF, DT ve SVM algoritmalarının 0'dan 25'e kadar tüm kaydırma durumlarındaki doğruluk oranları karşılaştırılmaktadır. 1 kaydırmadan itibaren tüm algoritmaların doğruluğu hızla düşerek %50 seviyelerine gerilemiş, ancak bazı kaydırma noktalarında sınırlı da olsa farklılaşmalar gözlenmiştir. Özellikle 10. kaydırmada DT ve SVM doğruluk oranları %64 seviyelerine yükselmiş, diğer kaydırmalarda ise performans daha stabil bir şekilde düşük kalmıştır. Bu durum, kaydırmanın algoritmaların sınıflandırma performansı üzerinde dalgalı bir etki yarattığını ve genel olarak doğruluk oranlarının kaydırmasız duruma kıyasla ciddi ölçüde düştüğünü göstermektedir.



Şekil 5. *25 Karakterlik Kaydırmaya Kadar Doğruluk Oranları*

V. TARTIŞMA

Bu çalışmanın bulguları, hassas bilgilerin makine öğrenmesi tabanlı saldırılardan korunmasında kriptografik tekniklerin potansiyelini vurgulamaktadır. Ruh sağlığı bağlamında, mahremiyetin korunması özellikle önemlidir. Ruh sağlığı kayıtları, ifşa edilmesi halinde ayrımcılığa, damgalanmaya ve önemli kişisel sıkıntılara yol açabilecek son derece kişisel ve hassas bilgiler içerebilmektedir. Metinsel verilere karakter kaydırmaları uygulandığında, sınıflandırıcıların doğruluğu, kesinliği, geri çağırması ve F1 puanları önemli ölçüde azalır ve böylece verileri doğru bir şekilde sınıflandırma oranları düşürülmektedir. Bu yaklaşım, veri gizliliğini artırmanın basit ama etkili bir yolunu sunmaktadır.

Karakter kaydırma gibi kriptografik teknikler, metinsel verileri korumak için basit ancak güçlü bir yöntem sunmaktadır. Bu yöntemler, bir metindeki karakterleri değiştirerek, makine öğrenmesi sınıflandırıcılarının doğru tahminler yapmakta zorlanacağı noktaya kadar içeriği gizleyebilmektedir. Bu çalışma, basit kaydırmaların bile sınıflandırıcı performansını büyük ölçüde azaltabileceğini ve bu sayede veri gizliliğini artırmak için uygun bir seçenek olduğunu göstermektedir. Çalışmada veri setindeki her harf alfabede alabileceği tüm harflere kaydırılmış ve harf olarak alabileceği tüm değerlere göre kıyaslanmıştır. Bu da, yaygın kullanılan makine öğrenmesi algoritmaları olan RF, DT ve SVM sınıflandırıcılarının performans kıyaslamasına çok çeşitli bir perspektif katmıştır.

Karakter kaydırma, basitliği ve uygulama kolaylığı nedeniyle özellikle caziptir. Daha karmaşık kriptografik yöntemlerin aksine, sofistike algoritmalar veya kapsamlı hesaplama kaynakları gerektirmez. Bu da onu kişisel verilerin korunmasından kurumsal veri güvenliğine kadar geniş bir uygulama yelpazesi için erişilebilir kılmaktadır.

VI. SONUÇ

Bu çalışma, hassas metin verilerini yetkisiz makine öğrenmesi sınıflandırmasından korumak için kriptografik tekniklerin, özellikle de karakter kaydırmanın etkinliğini göstermektedir. Bu yöntemlerin uygulanması, Rastgele Orman, Karar Ağacı ve Destek Vektör Makinesi gibi yaygın

kullanılan sınıflandırıcıların doğruluğunu ve diğer performans ölçütlerini önemli ölçüde azaltmaktadır. Bu bulgular, kriptografik yöntemlerin çeşitli alanlarda veri gizliliği ve güvenliğini artırma potansiyelinin altını çizmektedir. Bu, özellikle sağlık ve finans gibi veri gizliliğinin çok önemli olduğu alanlarda önemlidir. Genel olarak, bu çalışmanın sonuçları, büyük veri ve yapay zeka çağında hassas verileri korumak için kriptografik yöntemlerin sürekli araştırılması ve geliştirilmesi ihtiyacını vurgulamaktadır. Çalışmadaki deneyler karakter kaydırmanın 26 farklı şekilde yapılması ve bunların sonuçlarının yaygın kullanılan sınıflandırma algoritmaları kullanılarak kıyaslanmasını içermektedir. Bu bağlamda incelenen literatürde benzer çalışma bulunamamıştır.

Gelecekteki çalışmalar kapsamında veri güvenliğini artırmak için birden fazla kriptografik tekniğin birleştirilmesi, yöntemlerin gerçek dünya senaryoları için genelleştirilebilmesi ve canlı ortama alınması hedeflenmektedir. Bu alanlara odaklanarak, gelecekteki araştırmalar hassas verileri yetkisiz makine öğrenmesi analizinden korumak için daha esnek ve etkili stratejilerin geliştirilmesine katkıda bulunabilir.

KAYNAKLAR

[1] S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[2] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.

[4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.

[5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.

[6] (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[7] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/

[8] *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.

[9] "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.

[10] A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

[11] J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.

[12] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Clinical Research

2024, 8 (2), 99-108
DOI: 10.36287/ijmsit.8.2.9
Received: September 20, 2024; Accepted: November 13, 2024

A Multidisciplinary Discussion on the Theory of Relativity and the Mi'raj

Ahmet Efe[1*]

[1*]*Senior Regional Risk Management Officer/International Federation of Red Cross and Red Crescent, Ankara, Türkiye (icsiacag@gmail.com) (ORCID: 0000-0002-2691-7517)*

Abstract – This study delves into the intersection of the Theory of Relativity and the Mi'raj (Ascension) event, drawing upon Said Nursi's philosophical insights into the multifaceted nature of time. Nursi's exploration of temporal relativity illuminates the variability of time perception across different realms, resonating with Einstein's concepts of time dilation and the relativity of simultaneity. By examining the speeds of light, spirit, and imagination, Nursi illustrates that, much like motion varies in the universe, so too does time, thus rationalizing the remarkable physical and spiritual ascension of the Prophet Muhammad (PBUH) within a brief earthly timeframe. Furthermore, Nursi's analogy of multiple clock hands, each measuring different velocities, serves as a mirror to Einstein's space-time continuum, suggesting that the apparent temporal paradox of the Mi'raj can be reconciled through a relativistic framework with a remodeling approach to formulation of the relativity. Ultimately, Nursi's synthesis of spiritual metaphysics and scientific principles offers a distinctive lens to understand the Mi'raj event in the context of modern physics, proposing that the relativity of time provides a coherent explanation for this transcendent journey.

Keywords *– Theory of Relativity, Mi'raj (Ascension), Time Dilation, Said Nursi, Spiritual Metaphysics*

Citation: Efe, A. (2024). A Multidisciplinary Discussion on the Theory of Relativity and the Mi'raj. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 99-108.

I. Introduction

The *Mi'raj* (Ascension) event of the Prophet Muhammad (PBUH) holds profound spiritual significance within Islamic tradition, representing a miraculous journey beyond the material universe to the Divine Presence. While theological discussions have long explored its spiritual dimensions, modern science, particularly the Theory of Relativity, offers new perspectives on understanding this transcendent event. Said Nursi, a renowned Islamic scholar, provided a unique interpretation of the *Mi'raj*, emphasizing the relativity of time and motion, concepts that resonate with Einstein's groundbreaking work.

The Theory of Relativity, introduced by Albert Einstein in the early 20th century, revolutionized our understanding of time and space by showing that both are not absolute but rather interdependent and relative to the observer's speed. Time dilation, a key aspect of relativity, explains how time can pass differently for two observers depending on their relative velocities or gravitational fields. This principle can be employed to shed light on the temporal aspects of the *Mi'raj*, where the Prophet's journey across vast realms is said to have occurred within a short span of earthly time.

This study investigates the applicability of relativity to the *Mi'raj* event by integrating Said Nursi's reflections on time and motion. Nursi's insights suggest that time is not a fixed entity but rather a flexible dimension that varies across different realms of existence, aligning with the scientific understanding of time as a relative phenomenon. By comparing the speeds of various entities—light, spirit, and imagination—Nursi provides a metaphysical framework that can bridge Islamic cosmology with contemporary physics [17].

Said Nursi's approach is employed in this study due to its unique capacity to bridge the realms of spiritual metaphysics and modern scientific thought. His philosophical insights offer a profound understanding of time that aligns with contemporary theories, particularly the Theory of Relativity. By emphasizing the relativity of time and its perception across different dimensions, Nursi allows for a reinterpretation of the Mi'raj event that transcends conventional explanations. His use of analogies, such as the multiple clock hands measuring varying velocities, not only clarifies complex concepts but also resonates with Einsteinian physics, thereby facilitating a dialogue between spirituality and science. This integrative perspective enriches our comprehension of profound phenomena, allowing for a coherent synthesis that honors both the spiritual significance of the Mi'raj and the principles of modern physics.

The purpose of this study is to explore the *Mi'raj* through the lens of relativity, offering a harmonious view that integrates spiritual metaphysics with scientific principles. Through this approach, we aim to propose a coherent explanation for how the Prophet Muhammad (PBUH) could have experienced such an extraordinary journey within the constraints of human perception of time, ultimately contributing to a deeper understanding of the intersection between faith and science.

A. Research Design

This study adopts a literature-based approach, integrating classical Islamic theological perspectives, particularly those of Said Nursi, with modern scientific theories—primarily

Einstein's Theory of Relativity. Since this is not studied yet, our study provides a significant contribution to the literature. The aim is to create a conceptual model that aligns the metaphysical understanding of time during the *Mi'raj* event with the scientific principles of time dilation and relativity. The study will utilize a comparative analysis of theological texts with key works in physics, alongside the development of a hypothetical model that applies relativity to the *Mi'raj* event. Through this dual approach, the research seeks to propose a framework of remodeling formulation of the relativity that allows for the rationalization of the Prophet Muhammad's (PBUH) journey in light of contemporary scientific thought.

The foundation of this research lies in Said Nursi's extensive discussions on the relativity of time and the metaphysical nature of the Prophet's (PBUH) *Mi'raj*. Nursi frequently likened the relativity of time to the varying speeds of light, spirit, imagination, and sound. His analogy of different clock hands measuring time at different speeds offers a powerful metaphor for exploring how time, as experienced by humans, might differ significantly from time as experienced by spiritual or divine entities in his work called "*Mi'raç Risalesi*" in the 31st Words [17].

Parallel to this, the scientific literature on Einstein's Theory of Relativity—particularly the concepts of time dilation and the space-time continuum—provides the framework for understanding how physical time can vary under different conditions, such as high velocities or strong gravitational fields. Time dilation, which describes how time can slow down for an object traveling at speeds approaching the speed of light, offers a scientific basis for reconciling the temporal aspect of the *Mi'raj* with human perception.

In this research, Nursi's philosophical approach to time is modeled alongside Einstein's mathematical framework of relativity. The research discusses how the spiritual journey of the Prophet, which traversed vast realms of existence within minutes, could be explained by principles akin to relativistic time dilation.

B. Hypothesis

The *Mi'raj* event, as described in Islamic tradition, can be conceptually modeled using the Theory of Relativity, particularly the principles of time dilation and space-time, as articulated by Said Nursi, to reconcile the apparent paradox of the Prophet Muhammad's (PBUH) journey occurring within a brief span of earthly time.

C. Assumptions

1. Relativity of Time: Time is not a fixed or absolute entity, but rather a relative dimension that varies depending on speed, gravity, and the nature of the entity experiencing it. This assumption is derived from both the Theory of Relativity and Said Nursi's theological reflections.
2. Dual Nature of the *Mi'raj*: The event involved both the physical and spiritual realms, which may follow different laws of time and motion. This assumption is based on Nursi's argument that the Prophet's body and spirit ascended together, necessitating the application of both metaphysical and physical principles.
3. Variable Speed of Entities: Entities such as light, spirit, and imagination move at different velocities, and these velocities determine their experience of time. This idea is supported by both classical theological texts and modern physics.
4. Time Dilation as Explanatory Tool: The time dilation effect, as explained by the Theory of Relativity, is a valid model to explain how the Prophet could traverse immense distances within a short earthly timeframe.
5. Metaphorical Interpretation of Ascension: While the *Mi'raj* event is considered miraculous, it is assumed that certain aspects can be explored through metaphorical or symbolic interpretations, allowing scientific concepts to inform theological understanding.

D. Limitations

1. Non-empirical Nature: The study is primarily theoretical and relies on the intersection of religious and scientific thought without empirical data. The *Mi'raj* is a unique, non-replicable event that exists within the domain of belief, limiting the applicability of scientific models to it.
2. Conceptual Barriers: The integration of modern physics with metaphysical phenomena poses inherent challenges. Some aspects of the *Mi'raj* may transcend human understanding and cannot be fully explained through scientific principles alone.
3. Cultural and Religious Sensitivity: The interpretation of the *Mi'raj* using scientific models, while intended to enhance understanding, may not resonate with all Islamic scholars or followers, who may prefer a purely theological or spiritual explanation.
4. Scope of Relativity: The Theory of Relativity primarily deals with physical phenomena, while the *Mi'raj* encompasses spiritual and metaphysical elements that may not adhere strictly to the laws of physics as we know them.
5. Lack of Mathematical Rigor: While the study uses the principles of time dilation, it does not involve complex mathematical modeling, as the focus is on conceptual alignment rather than precise calculations.

By addressing these limitations, the study aims to explore the potential intersections between faith and science, without asserting definitive conclusions on matters that extend beyond empirical investigation.

II. Materials and Method

To model the *Mi'raj* event using relativity, the following components are integrated:

- Relativistic Time Dilation: By drawing on Einstein's equations for time dilation [10], where time slows down significantly as an object approaches the speed of light, we hypothesize that the Prophet's journey to the farthest reaches of the universe could occur in a relatively short time from the Earthly perspective, as his body and spirit moved at a speed far beyond human experience.
- Nursi's Multi-Level Time Approach: Nursi's metaphor of clock hands moving at different speeds provides a conceptual model for how different entities—body, spirit, and imagination—experience time. This model is mapped onto the relativistic framework, with the Prophet's body moving through physical space-time, and his spirit ascending to divine realms, where time as we understand it may cease to exist altogether.

- Parallel Realms and Time Perception: Drawing on the metaphysical nature of the *Mi'raj*, the study proposes that different realms (e.g., physical universe, spiritual realms, and the Divine Presence) experience time in radically different ways, and these variances can be compared to the space-time continuum described by Einstein.

A. Literature Discussions on Relativity

The Theory of Relativity, developed by Albert Einstein in the early 20th century, revolutionized the way we understand time, space, and gravity. Relativity, particularly time dilation and space-time, has been explored extensively in both scientific and philosophical contexts. Below is an analysis of key literature on relativity, connecting it to broader discussions relevant to the *Mi'raj* event, as well as metaphysical interpretations from Islamic scholars like Said Nursi.

1. The Special Theory of Relativity and Time Dilation

Einstein's Special Theory of Relativity (1905) introduced the idea that time and space are not absolute, but rather relative to the observer's frame of reference [10]. One of the most well-known implications of this theory is time dilation, where time slows down for objects moving at speeds approaching the speed of light.

For example, according to the theory, if an object travels at nearly the speed of light, time for that object will pass more slowly relative to a stationary observer. This phenomenon is captured in Einstein's equation:

$$t' = \frac{t}{\sqrt{1 - \frac{v^2}{c^2}}}$$

where t′ is the time observed in the moving frame, t is the time in the stationary frame, v is the velocity of the moving object, and ccc is the speed of light. This equation mathematically justifies the claim that an individual traveling at high speeds could experience a shorter period of time relative to someone stationary. This is central to the conceptual model in this study, where the *Mi'raj* event can be understood through time dilation.

Einstein's theory can be vividly illustrated through thought experiments. One of the most iconic is the "twin paradox," which explores the effects of time dilation in the context of two twins: one remains on Earth, while the other travels on a high-speed journey through space. Upon returning, the traveling twin would have aged more slowly due to time dilation, thus being younger than the twin who remained on Earth. While seemingly paradoxical, this scenario adheres strictly to the principles of special relativity and has analogies in real-world observations.

It is crucial to emphasize that time dilation is not a hypothetical abstraction but a measurable phenomenon with tangible consequences. One of the clearest examples is found in the realm of particle physics. Muons, unstable subatomic particles generated in the Earth's upper atmosphere by cosmic rays, have a very short lifespan in their own frame of reference. However, due to their high velocity, they experience time dilation, allowing them to survive longer and travel much further than they would if time passed at the same rate as it does for stationary observers on Earth [19]. Without time dilation, most muons would decay before reaching the Earth's surface, but experiments consistently show that their lifespan is extended by the effects of relativity.

The implications of time dilation stretch beyond the technicalities of physics into the realms of philosophy and metaphysics. In classical physics, time was perceived as uniform and independent, flowing identically for all observers in all situations. This Newtonian "absolute time" formed the bedrock of scientific and philosophical thought for centuries. However, with the advent of relativity, time's sovereignty was dethroned, revealing a universe where temporal experience is intertwined with an observer's motion.

This revolutionary perspective forces us to reconsider the nature of reality itself. If time is not a constant, but rather contingent on an observer's frame of reference, then the classical notion of a singular, objective reality comes into question. Events that are simultaneous for one observer may not be for another. This relativity of simultaneity, inherent in Einstein's theory, suggests a universe where there is no absolute "now," but rather a web of relative temporal experiences [5]. The implications for causality and determinism are profound, calling into question deeply held notions of how events unfold in time.

2. General Relativity and Gravitational Time Dilation

General Relativity (1915) expanded upon the Special Theory by incorporating gravity, describing it not as a force but as a curvature in space-time caused by mass and energy. A key prediction of this theory is gravitational time dilation—the idea that time passes more slowly in stronger gravitational fields. This phenomenon has been experimentally confirmed, such as in the famous Hafele–Keating experiment (1971), where atomic clocks on airplanes showed measurable time differences relative to clocks on the ground [12].

Einstein's general relativity posits that massive objects, such as planets and stars, warp the fabric of spacetime around them. This curvature affects the paths of objects in their vicinity and alters the flow of time. According to the theory, the stronger the gravitational field, the slower time moves relative to an observer situated in a weaker gravitational field [10, 11]. This concept has been experimentally validated in numerous contexts, most notably in the famous Pound-Rebka experiment [20], which measured the frequency shift of gamma rays emitted from a tower at Harvard University. The results confirmed that time indeed passes more slowly in a gravitational field, demonstrating the practical implications of Einstein's theoretical framework.

Gravitational time dilation can be quantitatively described by the equation:

$$T' = T\sqrt{1 - \frac{2GM}{c^2 r}}$$

Where T′ is the time interval measured by an observer in a gravitational field, T is the time interval measured by an observer far from the mass, G is the gravitational constant, M is the mass of the object creating the gravitational field, c is the speed of light, and r is the radial coordinate of the observer from the center of the mass [14]. This equation encapsulates the essence of gravitational time dilation: the closer one is to a massive object, the slower time progresses relative to an observer positioned far away from the gravitational influence.

Numerous empirical studies support the principles of gravitational time dilation. The Global Positioning System (GPS) satellites, for instance, require adjustments for both special and general relativistic effects to maintain their accuracy. Due to their high velocities and weaker gravitational

field compared to Earth's surface, the clocks aboard these satellites tick faster than those on the ground. Without accounting for this difference, GPS systems would yield errors of several kilometers per day [3]. This practical application of general relativity underscores the theory's significance beyond theoretical physics, impacting technologies that billions of people rely on daily.

The ramifications of gravitational time dilation extend into philosophical discussions about the nature of time itself. Traditional Newtonian physics treats time as an absolute entity, ticking uniformly regardless of external factors. In contrast, general relativity invites a more nuanced view, where time becomes relative and context-dependent. This shift raises profound questions about the nature of reality: Is time an intrinsic property of the universe, or is it fundamentally intertwined with the gravitational context? Such inquiries align with contemporary debates in both physics and philosophy, particularly concerning the interpretations of time in quantum mechanics [5, 8].

Despite the robust experimental validation of gravitational time dilation, challenges remain in reconciling general relativity with quantum mechanics. The theory does not address the phenomena that occur at singularities, such as those found in black holes, where gravitational effects become extreme, and conventional understandings of time and space cease to function coherently[18]. Furthermore, the reconciliation of general relativity with quantum theories is a focal point of ongoing research, as scientists explore frameworks like string theory and loop quantum gravity [22].

Therefore, the exploration of general relativity and gravitational time dilation has significantly altered our comprehension of the universe. From its theoretical underpinnings to its practical applications in technologies such as GPS, the implications of this theory are profound. Moreover, the philosophical questions it raises about the nature of time challenge traditional notions, inviting a re-evaluation of our understanding of reality itself. As research continues to unfold, particularly at the intersection of general relativity and quantum mechanics, the quest for a unified understanding of time may very well redefine our conceptual landscape in physics.

In the context of the *Mi'raj*, this principle suggests that different realms or dimensions—such as the material universe and spiritual realms—could experience time differently, as the *Mi'raj* involves traversing not only space but various metaphysical planes. The curvature of space-time in regions of intense gravitational fields, such as near black holes, offers parallels for understanding how different levels of existence might experience vastly different rates of time progression.

3. The Relativity of Time in Philosophical and Religious Contexts

The concept of time has been a profound subject of inquiry across various fields, including physics, philosophy, and religion. In contemporary discourse, the theory of relativity posits that time is not a universal constant but varies depending on the observer's velocity and gravitational field. This scientific framework invites philosophical and religious reflections, particularly regarding the nature of time as experienced in spiritual contexts. This discussion aims to explore the relativity of time through philosophical lenses and religious narratives, considering how these perspectives intersect with and diverge from scientific understandings.

In philosophical literature, time's relativity has also been discussed in non-physical terms. Henri Bergson [6] argued that time, as experienced by consciousness, is fundamentally different from time as measured by clocks. He proposed that "duration" (lived time) is a qualitative experience that varies depending on mental and spiritual states, whereas the scientific conception of time is quantitative and measurable. Bergson's ideas align with Said Nursi's reflections on the relativity of time, particularly in his analogies of how different beings experience time. Nursi emphasizes that time in the spiritual realm moves differently from time in the physical world. He draws on the Qur'anic description of time, stating that "a day with your Lord is like a thousand years" (Qur'an 22:47). This aligns with Bergson's notion that subjective experience can transcend the linear progression of time.

Philosophers have grappled with the nature of time for centuries. Early thinkers such as Aristotle [2] viewed time as a measure of change, closely tied to motion and events in the physical world. In contrast, Immanuel Kant proposed that time is not an external reality but rather a form of human intuition that structures our experiences (Kant, 1781). This notion of time as a subjective experience aligns intriguingly with relativistic physics, where time becomes dependent on the observer's frame of reference.

The modern philosopher Henri Bergson argued for a distinction between *measured time* (the quantitative time of clocks) and *lived time* (the qualitative experience of duration). He contended that true understanding of time must account for the rich, flowing nature of human experience, which cannot be fully captured by mathematical equations [6]. This perspective invites a reconsideration of time in both philosophical and spiritual contexts, suggesting that lived experiences of time may diverge from scientific measures.

In many religious traditions, time is understood not merely as a linear progression but as a dynamic interplay between the temporal and the eternal. For instance, in Christianity, time is often viewed through the lens of salvation history, where past, present, and future converge in the divine plan. Augustine of Hippo famously reflected on time as a mystery, suggesting that the past exists in memory, the future in expectation, and the present as a fleeting moment [4]. His reflections resonate with the philosophical inquiries of relativity, highlighting the subjective dimensions of temporal experience.

Islamic thought, particularly as articulated by Said Nursi, offers a rich exploration of time's relativity. Nursi's insights into the Miraj (Ascension) exemplify the intersection of time and spirituality. He argues that spiritual journeys transcend conventional time constraints, allowing for experiences that defy the physical limitations of distance and duration [17]. This perspective suggests that time may operate differently within spiritual realms, echoing the principles of relativity where subjective experiences of time can vary dramatically.

The theory of relativity, introduced by Albert Einstein, revolutionized our understanding of time and space, establishing that time is not an absolute entity but is intertwined with the fabric of the universe. Einstein's theory asserts that the passage of time can differ based on relative velocity and gravitational influence [10, 11]. This scientific perspective aligns intriguingly with the philosophical and religious interpretations of time, inviting questions about the nature of reality itself.

For instance, in moments of heightened awareness or spiritual transcendence, individuals often report altered

perceptions of time, experiencing what is sometimes referred to as "timelessness." Such phenomena can be likened to the effects of relativistic time dilation, where time appears to stretch or contract based on the observer's state of being. This analogy opens avenues for interdisciplinary dialogue between science, philosophy, and spirituality, suggesting that time's relativity might encompass not only physical realities but also subjective experiences shaped by consciousness.

The relativity of time presents a fertile ground for exploration across philosophical and religious contexts. While scientific discourse offers a framework for understanding time as a variable phenomenon, philosophical reflections deepen our comprehension of lived experiences, and religious narratives enrich our understanding of the interplay between the temporal and the eternal. By engaging with these diverse perspectives, we uncover a multifaceted view of time that acknowledges both its measurable aspects and its profound implications for human existence and spiritual experience.

4. Islamic Perspectives on Time and the *Mi'raj*

The Mi'raj, or the Night Ascension of the Prophet Muhammad (PBUH), represents a significant event in Islamic tradition, encapsulating profound theological, metaphysical, and temporal dimensions. This miraculous journey, as described in various Hadith and historical texts, prompts a re-examination of the nature of time from an Islamic perspective. Notably, the event challenges conventional understanding of time as linear and absolute, aligning with certain principles of relativity while integrating spiritual insights unique to Islamic thought. This paper seeks to explore the multifaceted Islamic perspectives on time in relation to the Mi'raj, drawing upon classical and contemporary scholarship.

Islamic scholars have long debated the concept of time, viewing it through both cosmological and spiritual lenses. The Qur'an frequently alludes to time as a creation of Allah, emphasizing its transitory nature. For instance, Surah Al-'Asr (103:1-3) underscores the importance of time in human affairs and the necessity of righteous deeds within it. Such references suggest that time is not merely a measure of duration but is intertwined with moral and spiritual accountability.

The Mi'raj, or the Night Ascension of the Prophet Muhammad (PBUH), is a significant event in Islamic tradition, symbolizing a profound spiritual journey and connection with the divine. This event is supported by various references in the Qur'an and Hadith, providing a foundation for its theological and metaphysical implications. The Qur'an begins by mentioning the journey: *"Glory be to Him who took His Servant by night from Al-Masjid Al-Haram to Al-Masjid Al-Aqsa, whose surroundings We have blessed, to show him of Our signs. Indeed, He is the Hearing, the Seeing."* This verse establishes the divine origin of the journey, emphasizing the transition from the Sacred Mosque in Mecca to the Farthest Mosque in Jerusalem, and indicates the significance of the signs shown to the Prophet during this experience. (Koran, Surah Al-Isra, 17:1). In the verses, the Qur'an narrates the Prophet's encounter with divine realities during the Mi'raj. It states, *"And he saw him another time, at the Sidrat al-Muntaha (the Lote Tree of the Utmost Boundary), near it is the Paradise of Refuge."* These verses underscore the spiritual elevation and the profound visions the Prophet experienced, which serve to affirm the extraordinary nature of the Mi'raj. (Koran, Surah Al-Najm, 53:13-18).

In the hadith collections, it is narrated that the Prophet Muhammad (PBUH) described his experience in detail, mentioning that he was taken through the heavens and met various prophets. He said, *"I was taken up to the heavens, and I met Adam, then I met Moses, then I met Jesus..."* (Sahih Bukhari). This narrative highlights the interconnectedness of prophetic missions and emphasizes the Prophet's unique status among them. Another hadith records the Prophet stating, *"During the night of Mi'raj, I was shown my ummah (community) and I was shown the status of my Lord."* (Sahih Muslim). This highlights the event's importance not only as a personal journey for the Prophet but also as a moment of intercession and representation for his followers. Therefore, the Mi'raj serves as a pivotal moment in Islamic theology, illustrating the concept of divine closeness and the unique status of the Prophet Muhammad (PBUH) as a mediator between Allah and humanity. The Qur'anic emphasis on the journey's miraculous nature encourages believers to understand time and space as relative constructs, aligning with Nursi's philosophical interpretations.

The event also reinforces the significance of prayer in Islam. Following the Mi'raj, the five daily prayers were instituted, marking a direct link between the divine experience and the practices that govern Muslim life. This connection underscores the necessity of maintaining a spiritual relationship with Allah, emphasizing the continuity of divine communication.

Prominent Islamic philosophers, such as Ibn Sina (Avicenna) and Al-Ghazali [1], contributed significantly to the discourse on time. Ibn Sina conceptualized time as a measure of motion, asserting that it exists only as a result of change in the physical world [15]. In contrast, Al-Ghazali [1] presented a more theological viewpoint, positing that time is a creation of God, thus placing divine will at the center of its significance [1]. These classical perspectives provide a foundation for understanding how time is perceived in Islamic metaphysics, emphasizing its non-material, divinely governed essence.

The Mi'raj is described in Islamic sources as an extraordinary journey undertaken by the Prophet Muhammad (PBUH) during which he ascended through the heavens, meeting various prophets and ultimately coming into the presence of Allah. This event raises critical questions regarding the nature of time experienced during the journey.

Said Nursi, a prominent Islamic thinker, argued that time is relative and can be experienced differently depending on one's spiritual state[17]. He posited that in spiritual realms, time is not bound by the same constraints as in the physical world. This notion resonates with the experience of the Prophet during the Mi'raj, where the temporal limits of earthly existence were transcended.

Nursi's framework suggests that during the Mi'raj, the Prophet's spiritual elevation allowed him to traverse vast distances in an instant, akin to the relativistic concept of time dilation, where time is perceived differently based on one's frame of reference. This aligns with Einstein's theory that time can slow down or speed up relative to speed and gravity, further illustrating the potential for varied experiences of time.

The Mi'raj serves not only as a physical journey but also as a theological cornerstone that reaffirms the relationship between time, space, and divine reality. The event exemplifies several key theological principles in Islam:

1. Divine Omnipotence: The Mi'raj illustrates Allah's power over time and space, reinforcing the belief that divine intervention can manifest beyond the limitations of the physical universe [5].

2. Temporal Versus Eternal: The journey highlights the contrast between temporal existence and eternal reality. The Prophet's experience symbolizes the soul's capacity to transcend worldly limitations, inviting believers to reflect on the spiritual dimensions of time [19].
3. Moral Accountability: By experiencing the divine presence and receiving the command of prayer during the Mi'raj, the Prophet emphasizes the importance of time management in the pursuit of spiritual fulfillment and moral conduct [1].

Islamic scholars have long debated the nature of time in the context of metaphysical events like the *Mi'raj*. Said Nursi's interpretation, as discussed in his magnum opus *Risale-i Nur*, applies a metaphysical framework to understand the Prophet Muhammad's (PBUH) journey. Nursi's analogy of clock hands moving at different speeds is critical to understanding his approach. He suggests that just as light and spirit move at different velocities, allowing them to transcend earthly limitations, so too could the Prophet's body and soul during the *Mi'raj*. By drawing on these metaphors, Nursi opens a path for reconciling theological beliefs with modern physics.

The exploration of Islamic perspectives on time through the lens of the Mi'raj reveals a rich tapestry of theological and philosophical insights. It challenges conventional notions of time, suggesting that spiritual elevation and divine interaction can profoundly alter our perception of temporal reality. By bridging classical Islamic thought with contemporary scientific understanding, this analysis underscores the dynamic interplay between faith, time, and existence.

B. Interdisciplinary Approaches to Time in Science and Religion

Contemporary discussions in both physics and theology are increasingly interdisciplinary. John Polkinghorne, a physicist-turned-theologian, argues for a dialogue between science and religion, suggesting that concepts like the relativity of time need not conflict with religious metaphysics. He advocates for a complementary approach, where both scientific and spiritual understandings of time coexist [19].

In his works, Polkinghorne explores how the Theory of Relativity might offer insights into theological concepts like eternity and the timelessness of God. While science deals with the relative nature of time within the universe, theology considers how God, existing outside time, can engage with temporal creation. This dual approach echoes the synthesis attempted in this study, where time in the context of the *Mi'raj* is explored as both a physical and metaphysical reality.

The dialogue between science and religion regarding time raises essential questions about the nature of reality. Can the scientific understanding of time as a physical dimension coexist with the religious view of time as a spiritual experience? Some scholars argue that both perspectives can be harmonized, suggesting that science explains the mechanics of time, while religion addresses the meaning and purpose behind it [5].

The literature on relativity, particularly time dilation and space-time, provides a robust framework for understanding the *Mi'raj* event in scientific terms. Said Nursi's metaphysical discussions of time align with these scientific principles, offering a unique opportunity to model the Prophet's (PBUH) journey as one that transcends the typical constraints of space and time. This interdisciplinary dialogue between science and theology enriches both domains, offering deeper insights into the nature of reality, both physical and spiritual.

The concept of time has long been a subject of fascination and inquiry across various fields, including physics, philosophy, psychology, and theology. An interdisciplinary approach to the study of time is not merely beneficial; it is essential for a comprehensive understanding of its multifaceted nature. This discussion elucidates the necessity of integrating insights from diverse disciplines to construct a holistic view of time, considering its implications for human experience, scientific inquiry, and philosophical exploration.

1. Diverse Interpretations of Time

Time is perceived and interpreted differently across disciplines. In physics, particularly in Einstein's theory of relativity, time is treated as a dimension intertwined with space, forming the fabric of spacetime [11]. This scientific perspective contrasts sharply with philosophical discussions, where time is often viewed as a mental construct, influenced by human perception and consciousness [6]. Such divergent interpretations raise important questions: How do these differing views shape our understanding of reality? Can a unified theory of time emerge from the synthesis of these perspectives?

For example, Bergson [6] emphasized the qualitative experience of time (durée) as opposed to the quantitative measurement of time (chronos) in physics. This philosophical standpoint invites a critical examination of how subjective experiences, such as emotions and memories, influence our understanding of time. The qualitative aspects of time become particularly relevant in fields such as psychology, where temporal perception plays a crucial role in cognitive processes and emotional states [7].

2. Temporal Dynamics in Human Experience

Human experience is deeply rooted in the perception of time. Psychological research indicates that our perception of time can be influenced by various factors, including age, emotional state, and context [13]. For instance, studies have shown that time seems to pass more slowly during moments of heightened emotional intensity, such as fear or joy [9]. This suggests that time is not merely a fixed quantity but a dynamic experience shaped by our mental and emotional states.

Incorporating insights from psychology into the discourse on time can enrich our understanding of its impact on human behavior and decision-making. As temporal perception is subjective, a multidisciplinary approach allows us to investigate how individuals and cultures conceptualize time, which in turn can inform practices in education, healthcare, and social policy.

3. Theological and Spiritual Dimensions of Time

The exploration of time extends into theological and spiritual realms, where different religious traditions offer unique insights into its nature and significance. For instance, in Islamic thought, time is often viewed as a linear journey towards the Divine, with profound implications for moral and ethical living (Nursi, 1950). This perspective encourages believers to reflect on the transient nature of worldly existence and the eternal reality of the hereafter.

Integrating theological perspectives into the study of time can illuminate how cultural and spiritual beliefs shape temporal concepts. This intersection is vital for fostering a more inclusive understanding of time that respects and acknowledges the diversity of human experience.

4. Scientific Advancements and Temporal Measurement

Advancements in science have also prompted a reevaluation of time's nature. The advent of quantum mechanics, for instance, has introduced concepts of time that challenge classical notions. Quantum entanglement raises questions about the simultaneity of events, suggesting that time may not be a linear progression but rather a more complex, interconnected phenomenon [16].

Such scientific developments necessitate an interdisciplinary discourse that includes philosophical inquiry into the implications of these findings. How do emerging scientific theories reshape our understanding of time? What philosophical implications arise from the quantum view of reality? Addressing these questions requires collaboration among physicists, philosophers, and other scholars.

5. Societal Implications of Time Perception

The societal implications of our understanding of time are profound. In a fast-paced world characterized by technological advancement and instant communication, the perception of time is often compressed. This phenomenon can lead to increased stress and anxiety, as individuals grapple with the demands of modern life [21].

An interdisciplinary approach to time can provide insights into how societal structures and cultural practices influence temporal perception. By examining how different cultures conceptualize and prioritize time, we can better understand the underlying values that shape human behavior and social organization.

Time perception, the subjective experience of time, plays a crucial role in shaping human behavior, social interactions, and cultural constructs. As a multifaceted phenomenon, it is influenced by psychological, biological, and sociocultural factors. This discussion explores the societal implications of time perception, focusing on its effects on social norms, productivity, mental health, and cultural practices. Time perception is deeply rooted in psychological constructs, which vary across individuals and cultures. According to Eagleman (2005), the brain constructs a model of time that is not merely a passive reflection of external reality but is actively shaped by cognitive processes and emotional states. For instance, time may appear to "fly" during enjoyable activities, while it can feel excruciatingly slow during periods of distress or boredom. This malleability of time perception can influence social interactions; individuals may perceive the same social situations differently based on their emotional states, leading to misunderstandings or conflicts (Lamm, Batson, & Decety, 2007).

Culturally constructed notions of time—whether linear or cyclical—significantly affect societal norms. Western cultures often view time linearly, emphasizing punctuality, efficiency, and future orientation. This perspective fosters a fast-paced lifestyle that prioritizes productivity and achievement, leading to societal pressures to conform to these ideals (Levine, 1997). Conversely, cultures that embrace cyclical notions of time, such as many Indigenous and Eastern cultures, may prioritize relationships, communal activities, and the natural rhythms of life, often leading to a more relaxed approach to time management (Hofstede, 2001). These contrasting perceptions of time can create tensions in multicultural settings. For example, in a workplace with diverse cultural backgrounds, differing attitudes toward deadlines and punctuality can lead to frustration and miscommunication among team members (Trompenaars & Hampden-Turner, 2012).

The societal emphasis on time efficiency has profound implications for productivity and economic performance. In capitalist societies, time is often equated with money, leading to the commodification of time and the establishment of rigorous work schedules (Graeber, 2018). This perspective fosters a culture where long hours and relentless work are valorized, potentially leading to burnout and diminished mental health (Maslach & Leiter, 2016). Moreover, the rise of technology has accelerated the pace of life, creating a paradox where individuals are constantly connected yet may feel increasingly isolated and pressured. The expectation of immediate responses in digital communication can distort time perception, causing anxiety and stress, particularly among younger generations (Twenge, 2019).

Time perception also plays a critical role in mental health. Distorted time perception is frequently observed in individuals with anxiety disorders, depression, and PTSD (Sierra et al., 2013). For example, individuals with anxiety may experience a heightened awareness of time, leading to feelings of urgency and distress. Conversely, those with depression may perceive time as dragging, which can exacerbate feelings of hopelessness (Riemann et al., 2010).

Understanding these variations in time perception can inform therapeutic practices. For instance, mindfulness-based interventions often encourage individuals to focus on the present moment, effectively reshaping their relationship with time and potentially alleviating symptoms of anxiety and depression (Kabat-Zinn, 1990). Cultural practices and rituals also reflect diverse understandings of time. In many societies, rituals serve to structure time, marking significant life events such as births, marriages, and deaths. These rituals reinforce social bonds and community identity, illustrating how time perception is intertwined with cultural values and practices (Turner, 1969).

Furthermore, the impact of globalization has led to the blending of various cultural perceptions of time. As societies interact, hybrid forms of time perception emerge, influencing everything from work practices to social relationships. This cultural exchange can enrich societies but may also lead to tensions as differing values regarding time coexist (Appadurai, 1996). Therefore, time perception is a complex construct with profound societal implications. It shapes individual experiences, influences cultural norms, impacts productivity, and affects mental health. As societies continue to evolve, understanding the nuances of time perception will be crucial in fostering effective communication, promoting well-being, and navigating the challenges of an increasingly interconnected world.

The need for an interdisciplinary nature of time is evident in the complexity of its interpretations, the dynamics of human experience, and the cultural and societal implications of temporal perception. By fostering collaboration among diverse fields, we can cultivate a more nuanced understanding of time that encompasses its scientific, philosophical, psychological, and spiritual dimensions. This holistic approach not only enriches academic discourse but also has practical implications for how we navigate our lives in relation to time.

III. RESULTS

In light of Said Nursi's discussions on time and relativity, particularly in relation to the event of Miraj (Ascension), we can conceptualize a function that relates the subjective

experience of time to physical movement across different levels of existence. Nursi's (1950) ideas below align well with the theory of relativity, especially when it comes to the perception of time across different frames of reference [17]:

How then should the motion at the speed of spirit of his subtle body, which followed his exalted spirit during the Ascension, seem contrary to reason?

Furthermore, it sometimes happens that on sleeping for ten minutes you are subject to a year's-worth of different states. And even, if the words spoken and heard during a dream lasting one minute were collected, for them to be spoken and heard in the waking world, a day or even longer, would be necessary. That means a single period is relative; it may seem like one day to one person and like a year to another. Consider the meaning of this by means of a comparison. Let us imagine a clock which measures the speed of the movement displayed by man, cannonballs, sound, light, electricity, spirit, and imagination. The clock has ten hands. One shows the hours while another counts the minutes in a sphere sixty times greater. Another hand counts the seconds in a sphere sixty times greater than the previous one, and yet others each count regularly decreasing fractions to a tenth of a second in vast spheres that regularly increase sixty times. Let us suppose the circles described by the hand counting hours was the size of our clock, so that of the hand counting tenths of a second would have to be the size of the annual orbit of the earth, or even larger. Now, let us suppose there are two people. One of them is as though mounted on the hour-hand and observes according to its motion while the other is on the hand counting tenths of a second. There will be an enormous difference, as great as the relation between our clock and the annual orbit of the earth, as regards the things observed by these two individuals in the same period. Thus, since time is like a hue, shade, or ribbon of motion, a rule that is in force in motion is also in force in time. And so, although the things we observe in the period of one hour would be equalled in amount by the conscious individual mounted on the hour-hand of the clock, like the one mounted on the hand counting tenths of a second, God's Noble Messenger (Peace and blessings be upon him) mounted Buraq of Divine Assistance and in the same space of time, in that specified hour, like lightning traversed the entire sphere of contingency, saw the wonders of the outer aspects of things and the aspects which look to their Creator, ascended to the point of the sphere of necessity, was honoured with Divine conversation and favoured with the vision of Divine beauty, received his decree, and returned to his duty. It was possible for this to happen, and it did happen.

And again, it comes to mind that you would say:" Yes, so it could happen, it is possible. But everything possible does not occur, does it? Is there anything else like this so that it can be accepted? How can the occurrence of something to which there are no similar cases be posited through only probability?" To which we would reply:

There are so many similar cases to it that they cannot be enumerated. For example, anyone who possesses sight can ascend with his eyes from the ground to the planet Neptune in a second. Anyone who has knowledge can mount the laws of astronomy with his intellect and travel beyond the stars in a minute. Anyone who has belief can, by mounting his thought on the action and pillars of the obligatory prayers, through a sort of Ascension, leave the universe behind and go as far as the Divine presence. Anyone who sees with his heart and any saint of perfection can, through his spiritual journeying, traverse in forty days the Divine Throne and the sphere of the Divine Names and attributes. And certain persons, even, like Shaykh Geylani and Imam-i Rabbani, truthfully recorded their spiritual ascensions as far as the Throne, which lasted a minute. Furthermore, there is the coming and going of the angels, which are luminous bodies, from the Divine Throne to the earth and from the earth to the Throne in a short period of time. And the people of Paradise ascend to the gardens of Paradise from the plain of resurrection in a short space of time.

Here's a simplified function formula inspired by both Einstein's theory of relativity and Nursi's arguments on time [17]:

$$T_{obs} = \frac{T_0}{\sqrt{1 - \frac{v^2}{c^2}}} \cdot f(s)$$

Where:

- T_{obs} = Subjective time experienced by the observer (in the spiritual or higher realm, based on Nursi's idea that time can vary by perception).
- T_0 = Proper time (the time experienced by a stationary observer, or the time of the physical world).
- V = Velocity of the observer or entity (which, in the context of the Miraj, can be related to the speed of spiritual or non-physical entities like light, spirit, or imagination).
- C = The speed of light in a vacuum.
- $f(s)$ = A function of spirituality or "subtlety" (inspired by Nursi's idea that spiritual realities can bypass the normal restrictions of physical laws).

Relativistic Time Dilation: The term

$$\frac{T_0}{\sqrt{1-\frac{v^2}{c^2}}}$$

is derived from Einstein's special relativity, where time slows down as an object approaches the speed of light. This component models the relativity of time as discussed by Nursi, particularly when applied to spiritual experiences that transcend normal physical constraints.

Spiritual Component f(s): This is a proposed multiplier based on Nursi's idea that time perception is not only affected by physical motion but also by the level of spiritual reality. For example, in dreams or spiritual ascensions like the Miraj, a few seconds might seem like years, suggesting a subjective time dilation that's beyond physical constraints.

f(s) could be modeled as:

$$f(s) = \frac{1}{1 + \alpha s}$$

Where:

- s represents the degree of spiritual elevation or proximity to divine reality.
- α is a constant that controls the extent of spiritual influence on time perception.

This formula attempts to incorporate both physical and metaphysical aspects of time, where:

- Time dilation occurs not only due to physical velocity (as in relativity) but also due to spiritual or metaphysical realities.
- Nursi's argument that time is experienced differently by different beings (e.g., angels, saints, and prophets) due to their spiritual closeness to divine realities is captured in the f(s) term, which adjusts the total time experienced.

This blend of modern physics with metaphysical time perception provides a unique framework to discuss the Miraj in both scientific and spiritual contexts. Describe in detail the materials and methods used when conducting the study. The citations you make from different sources must be given and referenced in references.

IV. DISCUSSION

The modeling of the Mi'raj through Said Nursi's philosophical approach offers several significant impacts on our understanding of this profound event. By applying concepts from the Theory of Relativity, the modeling bridges the gap between scientific understanding and spiritual beliefs. Nursi's interpretation suggests that the miraculous aspects of the Mi'raj can coexist with modern scientific principles, allowing believers to appreciate the event without conflicting with rational thought.

The modeling encourages a reevaluation of conventional perceptions of time and space. By illustrating how different realms may perceive time differently, it provides a framework for comprehending the extraordinary nature of the Mi'raj. This perspective aligns with contemporary discussions in physics, inviting a dialogue between Islamic theology and scientific inquiry.

Nursi's approach fosters a deeper exploration of spiritual metaphysics, suggesting that the Mi'raj is not merely a physical journey but also a profound spiritual experience. This understanding emphasizes the significance of personal

spiritual growth and connection to the divine, inspiring believers to engage more deeply in their faith.

By framing the Mi'raj in the context of relativity and prophetic narratives, the modeling underscores the interconnectedness of the missions of various prophets. This holistic view enriches the understanding of Islamic history and the continuity of divine guidance, encouraging a sense of unity among believers.

The implications of the modeling extend to Islamic practices, particularly in relation to prayer and spiritual discipline. Understanding the Mi'raj as a transformative event enhances the significance of the five daily prayers instituted during this journey, encouraging adherents to view these rituals as pathways to spiritual elevation.

The application of philosophical and scientific models to religious experiences promotes an environment of intellectual inquiry within the Islamic tradition. This encourages scholars and believers alike to explore the intersections of faith and reason, fostering a more dynamic engagement with their beliefs.

V. CONCLUSION

The exploration of Islamic perspectives on time, particularly through the lens of the Mi'raj, reveals a rich interplay between theological understanding and scientific inquiry. This study illuminates the profound implications of the Mi'raj for the concept of time, showcasing how it challenges linear, absolute perceptions and invites a deeper consideration of temporal dynamics that encompass both physical and metaphysical realms.

In examining the interpretations of classical scholars such as Ibn Sina and Al-Ghazali, alongside the innovative ideas of Said Nursi [17], we observe a transition from a purely material conception of time to one that is inherently tied to spiritual experience and divine reality. Nursi's differentiation between internal and external time underscores the potential for a relative experience of time that resonates with contemporary scientific theories, particularly Einstein's relativity. This convergence highlights a paradigm where spiritual elevation allows for the transcendence of conventional temporal constraints, thereby enriching our understanding of time not merely as a measurement of duration, but as a nuanced dimension influenced by spiritual states.

Furthermore, the interdisciplinary dialogue between science and religion opens avenues for reconciling seemingly disparate views on time. The alignment of Nursi's metaphysical insights with the principles of relativity suggests a holistic framework for understanding time, where the subjective experience of temporal flow can vary across different states of consciousness. This synthesis invites further exploration into how scientific advancements, particularly in quantum mechanics and relativistic physics, might inform theological discourse and vice versa.

In broader societal contexts, this study emphasizes the importance of recognizing diverse cultural and philosophical perspectives on time. As we navigate an increasingly complex and fast-paced world, understanding the implications of time perception can foster greater empathy and collaboration across cultural boundaries. Such insights are particularly relevant in addressing the mental health challenges associated with modern temporal pressures, as well as in enhancing interpersonal dynamics within multicultural environments.

The exploration of time through the lens of the Mi'raj, alongside Said Nursi's theological insights and modern physics, invites a profound rethinking of our understanding of time as a multifaceted phenomenon. This study has aimed to integrate the realms of science and spirituality, proposing a framework that transcends conventional interpretations of time by drawing from both Einstein's theories and Nursi's metaphysical reflections.

By modeling the subjective experience of time (T_{obs}) in relation to physical movement and spiritual elevation, we establish a theoretical foundation that acknowledges the interplay between temporal perception and existential realities. This remodeling emphasizes that time is not merely a linear progression but a dynamic construct influenced by various factors, including velocity, spiritual state, and the nature of the observer's experience. The proposed function serves as a metaphorical bridge, linking physical and metaphysical domains, and underscores how different beings, depending on their spiritual proximity to divine reality, may experience time differently.

In this context, the Mi'raj stands as a pivotal event that exemplifies the extraordinary capabilities of the soul when elevated through divine experience. It illustrates how moments of profound spiritual significance can alter temporal perception, enabling the transcendence of earthly limitations. This aligns with Nursi's assertion that spiritual realities can bypass physical constraints, offering a compelling narrative that enriches our understanding of religious experiences.

Furthermore, the integration of scientific concepts, such as time dilation, into this theological framework provides a robust dialogue between disciplines. It challenges the dichotomy often perceived between faith and reason, suggesting that insights from contemporary physics can enhance our comprehension of ancient spiritual truths. This interdisciplinary approach fosters a holistic view of time that resonates across cultures and beliefs, promoting a deeper appreciation for the complexity of human experience.

Ultimately, remodeling our understanding of time through the Mi'raj not only enriches Islamic thought but also contributes to broader discussions on time in the fields of philosophy, psychology, and science. It invites further research into how varying perceptions of time impact human behavior, ethical considerations, and our relationship with the divine. This synthesis of ideas paves the way for a more nuanced understanding of time that respects its rich tapestry of interpretations while recognizing the universal quest for meaning and connection in both the temporal and spiritual realms.

Future research

As we stand at the intersection of advancing technology and deepening spiritual inquiry, researchers are called to adopt a futuristic approach that transcends traditional boundaries. This entails integrating interdisciplinary methodologies that encompass philosophy, quantum physics, and cognitive science to further explore the dimensions of spiritual experiences like the Mi'raj. By leveraging emerging technologies such as virtual reality and artificial intelligence, researchers can create immersive simulations that allow individuals to experience and reflect on spiritual journeys in a modern context. Additionally, collaborative frameworks that unite scholars from diverse fields—such as theology, neuroscience, and the humanities—can foster rich dialogues that illuminate the complexities of human consciousness and

the divine. Furthermore, harnessing big data analytics to study religious texts and historical narratives can unveil patterns and insights that deepen our understanding of miraculous events. This innovative approach not only seeks to bridge the gap between science and spirituality but also invites a global discourse that enriches the collective human experience, ultimately encouraging a holistic view of existence that is as much about the spiritual as it is about the empirical.

REFERENCES

[1] Al-Ghazali. *The Incoherence of the Philosophers*. Translated by M. E. Marmura, Brigham Young University Press, 2000.

[2] Aristotle. *Physics*. Edited by J. Barnes, vol. 1, The Complete Works of Aristotle, Princeton University Press, 1996, pp. 203-384.

[3] Ashby, N. "Relativity in the Global Positioning System." *Proceedings of the 2003 IEEE/ION Position, Location and Navigation Symposium*, vol. 1, 2003, pp. 1-8. IEEE.

[4] Augustine of Hippo. *Confessions*. Translated by H. Chadwick, Oxford University Press, 397 AD.

[5] Barbour, I. G. *When Science Meets Religion: Enemies, Strangers, or Partners?* HarperOne, 2000.

[6] Bergson, H. *Duration and Simultaneity: With Reference to Einstein's Theory*. George Allen & Unwin Ltd, 1922.

[7] Block, R. A. "The Role of Temporal Experience in the Perception of Time." *Current Directions in Psychological Science*, vol. 12, no. 1, 2003, pp. 1-5. https://doi.org/10.1111/1467-8721.01217.

[8] Callender, C. "What is Time?" *The Oxford Handbook of Philosophy of Time*, Oxford University Press, 2017.

[9] Droit-Volet, S., and S. Gil. "The Effect of Emotion on the Perception of Time." *Emotion*, vol. 9, no. 6, 2009, pp. 883-888. https://doi.org/10.1037/a0017789.

[10] Einstein, A. "Zur Elektrodynamik bewegter Körper." *Annalen der Physik*, vol. 17, no. 10, 1905, pp. 891–921.

[11] Einstein, A. "Die Grundlage der allgemeinen Relativitätstheorie." *Annalen der Physik*, vol. 49, no. 7, 1916, pp. 769-822.

[12] Hafele, J. C., and R. E. Keating. "Around-the-World Atomic Clocks: Predicted Relativistic Time Gains." *Science*, vol. 177, no. 4044, 1972, pp. 166–168. https://doi.org/10.1126/science.177.4044.166.

[13] Hancock, P. A., and D. R. Dyer. "Temporal Perception: The Effects of Cognitive Load on the Perception of Time." *Human Factors*, vol. 43, no. 2, 2001, pp. 253-266. https://doi.org/10.1518/001872001775898908.

[14] Misner, C. W., K. S. Thorne, and J. A. Wheeler. *Gravitation*. W. H. Freeman, 1973.

[15] Nasr, S. H. *Islamic Philosophy from Its Origin to the Present: Philosophy in the Land of Prophecy*. State University of New York Press, 1996.

[16] Norton, J. D. "The Hole Argument." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2008 Edition. https://plato.stanford.edu/archives/fall2008/entries/qm-hole/.

[17] Nursi, S. *The Words*. Sözler Publications, 1950.

[18] Penrose, R. "Gravitational Collapse and Space-Time Singularities." *Physical Review Letters*, vol. 14, no. 3, 1965, pp. 57-59. https://doi.org/10.1103/PhysRevLett.14.57.

[19] Polkinghorne, J. *Belief in God in an Age of Science*. Yale University Press, 1998.

[20] Pound, R. V., and G. A. Rebka. "Gravitational Red Shift in Nuclear Resonance." *Physical Review Letters*, vol. 3, no. 9, 1959, pp. 439-441. https://doi.org/10.1103/PhysRevLett.3.439.

[21] Sennett, R. *The Corrosion of Character: The Personal Consequences of Work in the New Capitalism*. W.W. Norton & Company, 1998.

[22] Smolin, L. *Three Roads to Quantum Gravity*. Basic Books, 2002.

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 109-117
DOI: 10.36287/ijmsit.8.2.10
Received: November 6, 2024; Accepted: December 9, 2024

A Sample Strategic Marketing Application: Patient Segmentation and Channel Analysis with The LRM Model

Mustafa Şehirli[1*], and Samet Aydın [2]

[1*]*SHMYO, Managemnt And Organization Department, University of Health Sciences, İstanbul, Türkiye (mustafasafrani@gmail.com) (ORCID: 0000-0002-4800-0283)*
[2]*International Trade and Logistics Management, Maltepe University, Istanbul, Turkiye (sametaydin@maltepe.edu.tr) (ORCID: 0000-0003-2275-4682)*

Abstract **–** This research aims to develop a new customer segmentation method and to propose strategies for acquiring new customers accordingly. To this end, data from 48,870 patients of a healthcare institution were segmented using the K-Means method. Patients were classified based on their longevity (L), recency (R), and monetary return (M) status and analyzed according to acquisition channels. The findings revealed that the total patients were divided into four distinct clusters. Two clusters containing 2,981 patients, representing 6% of the total, were identified as the ideal segments. While evaluating the clusters, a new indicator based on the Profitability Ratio per Patient was also utilized. The research concluded that the hospital primarily acquires patients through referral channels, with search engines and the website as the second most effective channel. At the same time, social media advertising had a comparatively lesser impact on patient acquisition. Furthermore, it was found that there were no significant differences among customer acquisition channels between the clusters. Recommendations for managers at the end of the study include maintaining more comprehensive customer data, developing profiles for cluster patients for similar sales activities, organizing "refer a friend" campaigns, and conducting evaluations between channel costs and profits.

Keywords *– segmentation, LRFM model, K-means method, clustering, strategic marketing*

Citation: Şehirli, M., Aydın, S. (2024). A Sample Strategic Marketing Application: Patient Segmentation and Channel Analysis with The LRM Model. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 109-117.

I. INTRODUCTION

According to Kotler & Armstrong [16], the art of retaining profitable customers in marketing only aims to acquire or retain some customers. According to the authors, a profitable customer is one whose total revenue over time exceeds the costs incurred while acquiring and serving them; in other words, a customer with a positive Customer Lifetime Value (CLV). In this case, it becomes the most rational approach for companies to identify and invest in customers with high CLV. On the other hand, identifying and distinguishing these customers is a highly challenging and important task. Therefore, during the research phase, which is the first step in the marketing management process, sufficient data must be collected, classified, analyzed, and customers grouped. This process is referred to as segmentation. In the strategic marketing phase, one or more of these segments are targeted, and efforts are made to establish long-term profitable relationships using all marketing mix instruments.

This research has been conducted based on existing customer data related to the segmentation phase. Therefore, this study is not a market analysis but a work that focuses on differentiating, understanding, and clustering existing customers, analyzing them according to acquisition channels, and providing recommendations for customer groups that companies should focus on, including potential customers. Similar studies exist in the literature. In this regard, it cannot be claimed that this study is highly original; however, since segmentation studies are, as noted below, highly subjective and often require posterior approaches, and because the results are produced according to the characteristics of the dataset, it can be said that this study has a degree of originality. For instance, the F variable, which shows the frequency of customer acquisition, was not used in the evaluation, and the study analyzed customers based on the channels through which they arrived. Viewed from this angle, this research serves as an example of the subjective segmentation studies that firms should conduct according to their customer profiles and industry characteristics.

Another significant difference in this research is its perspective, which includes potential customers likely to be acquired. The focus is on retaining existing customers and producing practical outputs that facilitate acquiring new customers. Thus, the research serves as an example for the players in the industry being studied and managers and researchers in any sector.

The literature section presents application examples rather than theoretical explanations of the segmentation concept, focusing on the most commonly used methods and providing more detailed information about them. The research is conducted using models and methods (LRFM and K-Means) commonly used in previous studies and with a considerable amount of actual data (approximately 48,870). The research approach is both a priori and posterior, that is, a hybrid approach. In other words, while segmenting customer data, the

features in a widely accepted method (LRFM) have been revised, and different evaluations based on the dataset's characteristics have been made.

The research findings have been interpreted to serve as an example for industry managers and researchers.

II. Literature Review

A. Strategic Marketing and Customer Segmentation Concept

In today's increasing competition, companies must focus their interests and resources on specific customer groups to capture customers' attention and establish close, long-term relationships. This approach is one of the most strategic methods in marketing. Selecting one or more customer groups after segmentation is referred to as targeting, and the formulation of brand communication aimed at these target customer groups is known as positioning.

The three-phase strategic marketing process initiated from customer data is illustrated in Figure 1.

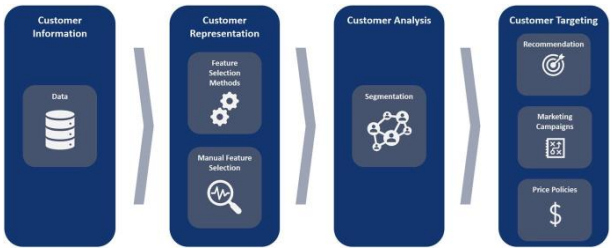


Fig. 1. Strategic Marketing Process Based on Customer Data [1]

This research focuses primarily on segmentation; therefore, the literature review includes information exclusively about this concept. Market segmentation can be defined as the process of dividing a market into customer groups with different needs, characteristics, or behaviours [16]. In other words, segmentation is a strategy for breaking the market into smaller, homogeneous parts to develop products and services tailored to groups with distinct wants and expectations [28].

Through segmentation, it is possible to develop different products for various groups and different marketing strategies and efforts [31]. Segmentation enables companies to focus on a specific subset of consumers that can provide the best service, thereby clarifying the marketing planning process by identifying the needs of particular customer groups for marketing programs [6].

B. Segmentation Criteria and Methods

Various approaches to market segmentation exist, each with its own advantages and disadvantages. The most suitable approach largely depends on the research objective and the type of data available [29]. Generally, analytical marketing strategies, such as data mining, are used to uncover and group customer segments. These strategies utilize demographic, behavioural, and psychographic data to acquire new customers while addressing existing customers' specific needs and wants, enhancing their loyalty [24].

As noted, data is essential for effective segmentation. This data can be categorized into explicit (open) and implicit data. Explicit data refers to customer information, such as demographic details, whereas implicit data pertains to behavioural information, like purchase histories, for accounting purposes. Collecting explicit data can be challenging and potentially misleading due to its constantly changing nature [3]. In contrast, implicit data is generally more accessible and accurate [1].

Despite the growing research and literature on segmentation models, many researchers rely solely on demographic variables to classify consumers, which hinders the discovery of unique patterns, relationships, and latent characteristics [19].

The next phase after obtaining customer data is deciding which customer information will represent the customers and, consequently, be used in segmentation. At this point, there are two approaches: A priori approaches and posterior approaches. The first approach divides the market based on predetermined criteria (demographics, purchasing behaviour, geography, and income). In contrast, the second analyzes the market more thoroughly based on data obtained from the market [11].

In the first approach, the most commonly used models can be summarized as follows [1]:

- RFM (Recency, Frequently., Monetart): Proposed by Hughes [12] , the RFM model is advantageous and widely used because it assesses customer characteristics based solely on three criteria: recency, frequency, and monetary value. It is a powerful and simple model for identifying profitable customers [14] , [27] , [32]. Further details on this model are provided below.
- PCA (Principal Component Analysis): PCA is a dimensionality reduction model that removes features with low information content from consideration.
- PT (Purchase Tree): In this method, customers' products are represented as the leaves of a tree, while product categories are represented as the branches.
- CHAID (Chi-square Automatic Interaction Detector): A technical model based on decision tree techniques.
- CLV (Customer Lifetime Value): A popular metric that focuses on the profit a customer will bring to the company over their lifetime if they remain loyal to the brand.
- DWT (Discrete Wavelet Transform): This method captures location and frequency information.
- GRAPH: Segmentation is based on customer interactions according to location and frequency information.
- MCA (Multiple Correspondence Analysis): MCA allows for a lower-dimensional representation of categorical features.

Alves Gomes and Meisen [1] identified 105 publications analyzing customer behaviour through segmentation methods between 2000 and 2022 and presented a comprehensive study on segmentation techniques. Their research evaluated four stages: data collection, customer representation, customer analysis through segmentation, and customer targeting. According to their findings, customer representation is generally conducted through manual feature selection or RFM analysis. The most commonly used segmentation method is the k-means method, as noted below.

According to Alves Gomes and Meisen [1], RFM has been the most widely used method in the last 20 years (41.9%). There is also a significant amount of research selecting customer data through posterior approaches (47.6%). Figure 2 presents information showing the distribution of methods in this research.

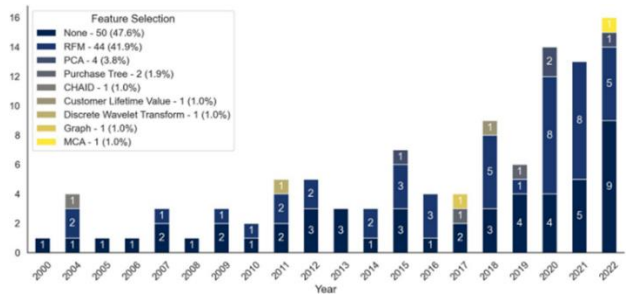


Figure 2: Selection Methods for Customer Representing Features [1]

C. RFM and Its Variants

RFM, as previously defined, is a targeting method that segments customers based on the assumption that newer, more frequent, and higher-spending customers have greater potential. Here, recency refers to the number of days or months since the last purchase; frequency indicates the number of purchases made within a specific period; and monetary value represents the total amount spent [36]. However, RFM has yet to undergo significant improvements over decades. In an actual segment, customers within each segment should respond differently. For example, a customer who has made four purchases may be more likely to make future purchases than a customer with five, which explains why RFM is often considered lower quality compared to methods like CHAID or regression analysis. While RFM is a coarse model based on heuristic perception, advanced methods benefit from statistical frameworks [37].

Indeed, RFM tends to focus excessively on transactional data while overlooking significant differences among customers, such as their values and lifestyles [20]. This can lead to inaccuracies in predicting customer behaviour. Consequently, different parameters have been added to RFM analysis, especially in dynamic sectors like healthcare. For instance, Chang and Tsay [2] introduced length (L), while Yeh et al. [38] added first purchase (T) and customer churn risk (C).

The LRFM model developed by Chang and Tsay [2] segments customers based on the length of their relationship with the company, as illustrated in Figure 3.

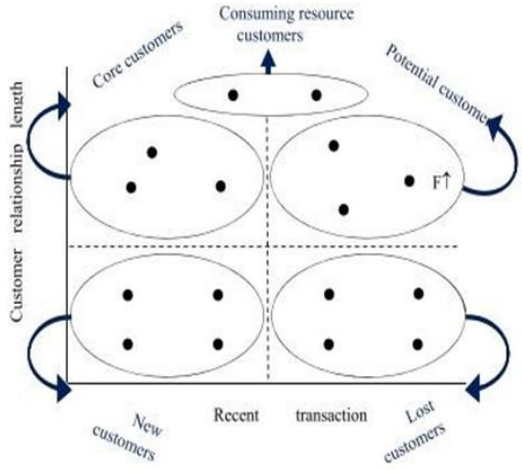


Figure 3: Chang & Tsay [2] LFRM Customer Segmentation Model [18]

As shown in Figure 3, L (length) represents the number of periods from the first purchase to the most recent purchase [36]. Moghaddam et al. [21] introduced the variety (Variety-V) parameter to the model, calling it RFMV, emphasizing the importance of considering product diversity [26].

RFM evaluation is generally conducted using twenty percent segments. For example, the recency (R) score reflects how current a customer's shopping history is; the more recent the purchases, the higher the R-value. Customers in the most current 20% receive a score of 5 points, while those with the oldest purchase dates receive 1 point. Similarly, the frequency (F) and monetary (M) values are also scored based on these twenty percent thresholds [34].

In summary, RFM has advantages such as being simple and comprehensible, offering flexible coding options, and providing insights for predicting future customer behaviour [37].

D. Segmentation Methods

According to the study by Alves Gomes and Meisen [1] , the most commonly used method for segmentation is K-Means, accounting for 39% of the approaches. The following section provides more detailed information about the K-Means method. The distribution of other methods can be found in Figure 4.

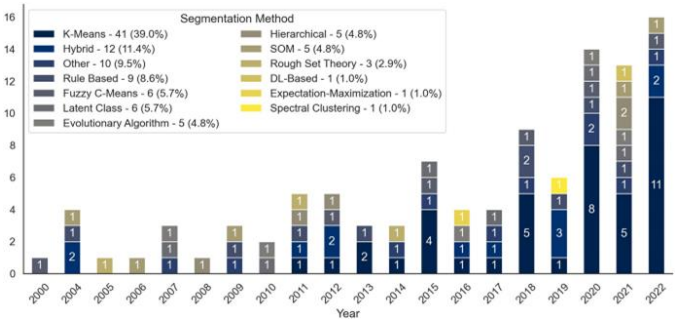


Figure 4: Usage Rates of Segmentation Methods [1]

Outside of the K-Means method, the usage rates for other segmentation approaches—such as time series segmentation, rule-based clustering, Fuzzy C-Means (FCM), Latent Class Analysis (LCA), and Evolutionary Algorithms (EAs)—are quite low, with a total of ten different algorithms showing minimal application. Therefore, it is sufficient to focus solely on the K-Means method:

D.1. K-Means Method

The purpose of the K-Means algorithm is to partition data points into kkk segments, minimizing the distance between the points within each segment. In other words, K-Means clustering is a method aimed at dividing observations into kkk clusters, where each observation belongs to the cluster with the nearest mean. Initially used by MacQueen [17] , the K-Means clustering algorithm is widely applied in various fields such as data mining, statistical data analysis, and other business applications, clustering each observation according to its nearest mean [4]. The Euclidean distance is commonly used in the analysis. However, approximate algorithms such as K-medoids may also be employed due to the NP-hard nature of the underlying optimization problem [17].

The study by Christy et al. [5] observed that the K-Means clustering algorithm consumed less time and reduced the number of iterations compared to Fuzzy C-Means and modified K-Means algorithms. Similarly, in Kanca et al.'s [13] research on 1.9 million unique customer records in the textile sector, the clustering group defined by the fuzzy C-means algorithm (five clusters) provided less in-depth analysis

compared to the clustering group produced by the K-Means algorithm (eight clusters).

E. Segmentation in the Healthcare Sector

According to Nnaji et al. [23], data analytics is revolutionizing the healthcare industry and enhancing customer experience and market penetration. Through data analytics, healthcare providers gain deep insights into patient preferences and behaviours, allowing them to develop effective communication strategies [23]. As providers continue adopting data-driven strategies, the healthcare industry is poised for transformative changes that will benefit providers and patients.

According to Torkzadeh et al. [29], while many studies have addressed market segmentation in the healthcare sector using various methods, there has yet to be a consensus on the best approach. Additionally, no technique has been conducted to compare these methods. Therefore, the authors conducted an extensive study, evaluating 22 articles selected from 239 examined to identify the best techniques and criteria. The authors' findings are presented in Figure 5.

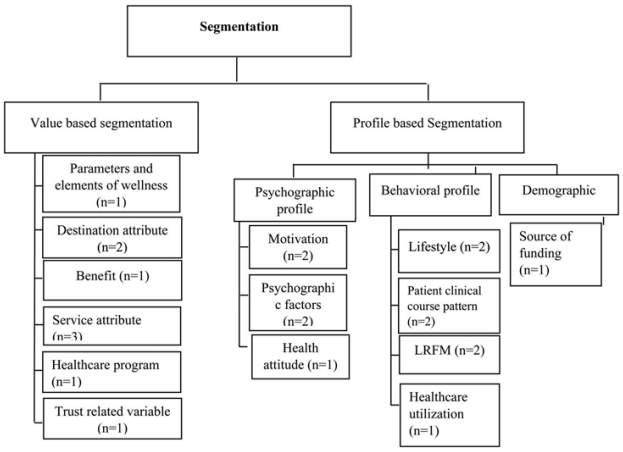


Figure 5: Academic Research on Segmentation in Health Tourism [29]

E.1. Similar Studies

Wei et al. [33] conducted a segmentation study at a dental clinic in Taiwan using the LRFM technique, dividing 2,258 patients into 12 clusters. They developed various suggestions for each cluster, such as offering referral discounts, waiving registration fees, free parking and medical consultations, and giving small gifts alongside medical services. However, they did not present a financial analysis.

In a similar study, Wuu et al. [36] segmented 1,462 patients at a dental clinic into 12 groups. They classified patients using two matrices: customer value and customer relationship. The customer value matrix included treatment frequency and payment amounts, while the customer relationship matrix featured innovation and length variables. Based on these matrices, the authors categorized patients as loyal, those to retain, and those to discard.

Hallab et al. [10] segmented 520 patients in the health sector based on health lifestyle attitudes, which are considered a psychographic variable. They identified this group as significant for health tourism and offered practical recommendations, such as creating smoke-free areas and providing exercise facilities.

Dryglas & Salamaga [8] examined 2,050 tourists at a spa centre in Poland, categorizing them by destination choice motivations (wellness and treatment seekers, treatment seekers, tourism, treatment and wellness seekers). They found that segments differed in socio-demographic, behavioural, and psychographic factors and provided practical suggestions for improving service quality and offering discounts for family packages or during low seasons.

Chen et al. [4] developed a hybrid method using fuzzy logic to support the K-Means algorithm for identifying a hospital's Target Customer Segment (TCS) based on a dataset of 183,947 records.

The results focused on a patient group of 75 individuals and identified 10 key characteristics (Age, Gender, Maximum Monetary Value, Novelty, Frequency, Surgical Chronic Illness, Critical Condition, CT and MRI, and Department) that the hospital should concentrate on. Nakano and Kondo [22] segmented customers based on their entry channel (either online or offline), highlighting differences between the two groups and providing recommendations for managers. Similarly, Konuş et al. [15] proposed a segmentation scheme using Latent-Class Cluster Analysis that considered the communication channel with the company, the stages of information search and purchase, as well as psychographic and demographic individual differences, resulting in three customer segments: multichannel enthusiasts, indifferent shoppers, and store-oriented consumers.

Özkan & Kocakoç [25] segmented patients visiting a children's hospital department using the LRFM method, dividing them into four clusters and offering recommendations for hospital management. Although the fact that the patients were children increased the constraints of the research, practical results were still achieved. Witschel et al. [35] identified that a learned decision tree model was the best descriptive method to capture the essence of clusters following segmentation efforts. Sarıoğlu and İnel [26] conducted three different segmentation analyses—RFM, LRFM, and RFMV—on 228 customers of a biotechnology firm operating under a B2B business model, demonstrating that the RFMV model, which includes the variety (V) parameter, could significantly enhance the firm's customer-centricity and customization capabilities. The authors recommended that different parameters be considered in segmentation studies based on the characteristics of the respective sectors.

III. MATERIALS AND METHOD

This study used data from a dental hospital operating 15 branches in Turkey, covering the period from January 2023 to September 2024. Permissions were obtained by the KVKK (Personal Data Protection Law) to use data from all domestic and international patients served during this period. The data included patient IDs, the relevant branch, the previously explained L, R, and M information, the channel through which the patient arrived at the hospital, and the procedures performed. A total of 48,870 patient records were analyzed. No personal or sensitive data that would create a personal data right was analyzed; patient names and all personal identifiers were coded and stored. Thus, the entire hospital patient profile was segmented..

F. Data Standardization

Prior to analysis, standardization was performed on each dataset. Categorical variables, such as the last visit date of the patient (R - Recency), were converted into numeric values. The first day included in the analysis was assigned a value of

1, and subsequent days were numerically counted in reverse order (with the date one year later coded as 365). This method ensured that patients with the highest R values were those who had visited the hospital most recently.

The Length (L) variable, which shows how long the patient has been receiving services from the hospital, was calculated by subtracting the first visit date from the last visit date. The highest value indicated the most loyal patient. The Monetary (M) variable was standardized to present value due to inflation. To achieve this, the four sales price increase rates (%25, %30, %20, and %20) implemented by the company within two years from the initial registration date were applied backwards to the customer payments using compound interest calculations. Thus, payments made in January 2023 were aligned with those made in September 2024.

G. Data Recording and Analysis

All categorical variables were standardized and recorded as separate variables (Zlength, ZMonetary, Zrecency). However, the Frequency (F) data indicates how often patients visited and was not calculated and excluded from the analysis. The F data could have been more meaningful for the dental hospital context and related more to treatment content outside the patient's control. Therefore, it was left out of the evaluation.

The standardized data were subjected to K-means clustering using the SPSS software. The resulting clusters were analyzed and interpreted based on the information obtained from the hospital, focusing on the channels through which patients arrived and the treatments they received. After the K-means analysis, comments were made using the Pivot Table tool in MS Excel.

IV. RESULTS

During the iterations, the number of individuals in each cluster was monitored. When two clusters were formed, one contained 47,597 individuals, and the other had 1,273. With three clusters, the counts were 46,530 in the first cluster, 2,129 in the second, and 211 in the third. When four clusters were established, the distribution of individuals is shown in Table 1. However, single-member groups emerged when attempting to create five or more clusters, indicating that having more than four would be meaningful.

Table 1. Number of Cases in Each Cluster

Cluster	1	2630.000
	2	22.000
	3	45867.000
	4	351.000
Valid		48870.000
Missing		.000

After conducting 10 iterations for four clusters, the central values for each cluster were obtained, as shown in Table 2.

Table 2. Final Cluster Centers

	Cluster			
	1	2	3	4
L	5	0	3	7
R	314	244	324	314
M	46488	304057	4176	140540

As indicated in Table 3, an ANOVA analysis was conducted to test the significance of the obtained data.

Table 3. Anova Test

	Cluster					
	Mean Square	df	Mean Square	df	F	Sig.
L	6308.230	3	501.817	48866	12.571	.000
R	137646.530	3	35003.238	48866	3.932	.008
M	42172498…	3	49364084.384	48866	85431.420	.000

The results are considered statistically significant since all p-values in Table 3 are less than 0.05. The statistical data for each cluster are presented in Table 4.

Tablo 4. Cluster Data

Cls	**Number Patients**		**L**	**R**	**M (₺)**
1	2630 (%5) Total M: ₺122.263 (%33)	Min	-11	28	26.034
		Max	617	665	98.172
		Av	5	314	46.488
		SD	29	184	17.760
2	22 (%0) Total M: ₺6.689 (%2)	Min	0	31	246.435
		Max	6	630	575.100
		Avg	0	244	304.057
		SD	1	171	69.755
3	45867 (%94) Total M ₺191.540 (%52)	Mi	-16	28	
		Max	833	665	26.023
		Avg	3	324	4.176
		SD	22	187	4.690
4	351 (%1) Total M ₺49.329 (%13)	Min	0	28	98.360
		Max	206	662	240.000
		Avg	7	314	140.540
		SD	29	192	36.470
Av			2,86	323	7.568

When evaluating the clusters, the variable representing the channels through which patients arrived at the hospital (i.e., how patients were acquired) was also considered. For this purpose, Table 5 lists the channels patients came to the hospital.

Table 5. Analysis of All Patients by Arrival Channel

Arrival Channels	**Num**	**%**	**Average M**
Recommendation	21198	43%	₺ 8.050
Google- Website	15654	32%	₺ 6.606
Walk-in Patients	9343	19%	₺ 7.151
WhatsApp	1359	3%	₺ 12.948
Unknown	799	2%	₺ 6.460
Social Med.	306	1%	₺ 10.439
Adv.TV/Mag./N.P	108	0%	₺ 11.442
Earthquake Victim	56	0%	₺ 9.334
Fair	20	0%	₺ 40.673
Overseas Branch	19	0%	₺ 17.299
Commission	8	0%	₺ 13.899
Total/Average	48870	100%	₺ 13.118

This section will present a general distribution table showing the patient acquisition channels for all patients and specific tables indicating the patient acquisition channels for each segment. To avoid a complex presentation, only the

highest-valued channels will be included, and channels with less than 1% will not be shown.

Table 6. Patient Acquisition Channels for Cluster 1

Pat. Acq. Channels	Num	%	Avr.M (₺)
Recommendation	1272	48%	46.219
Google Website	683	26%	47.332
Walk-in Patients	469	18%	45.943
WhatsApp	127	5%	47.331
Unknown	32	1%	42.799
Intt - Social Media	25	1%	47.619
Total/Average	**2630**	**100%**	**45.255**

Table 7. Patient Acquisition Channels for Cluster 2

Pat. Acq. Channels	Num.	%	Average M
Walk-in Patients	7	32%	₺ 293.800
Recommendation	7	32%	₺ 332.601
Google Website	5	23%	₺ 280.553
WhatsApp	3	14%	₺ 300.563
Total/Average	22	1	₺ 301.879

Table 8. Patient Acquisition Channels for Cluster 3

Pat. Acq. Channels	Num.	%	Average M
Recommendation	19744	43%	₺ 4.321
Google Website	14884	32%	₺ 3.889
Walk-in Patients	8818	19%	₺ 4.141
WhatsApp	1196	3%	₺ 4.972
Unknown	764	2%	₺ 4.324
Social Media	278	1%	₺ 5.513
Total/Average	**45867**	**100%**	**₺ 5.742**

Table 9. Patient Acquisition Channels for Cluster 4

Pat. Acq. Channels	Num.	%	Average M
Recommendation	175	50%	₺ 138.363
Google Website	82	23%	₺ 143.934
Walk-in Patients	49	14%	₺ 136.537
WhatsApp	33	9%	₺ 143.563
Unknown	3	1%	₺ 162.793
Fair	3	1%	₺ 145.962
Social Media	3	1%	₺ 157.094
Adv.TV/Mag./N.P	2	1%	₺ 183.675
Overseas Branch	1	0%	₺ 120.771
Total/Average	**351**	**100%**	**₺ 148.077**

V. DISCUSSION

The third group, which has the most significant number of members (45,867 people), is the group with the least number of members. In contrast, the second group has the highest average monetary return (304,057 TL). As expected, members in the third group have the lowest average monetary return (4,176 TL). In addition, patients in the second group have the lowest R (recency) and L (loyalty or age) values. This situation can be interpreted as being caused by a subjective factor (high payment amounts). The highest R-value is found in the third group.

Given that the third group represents 94% of the patients, it is normal for this group to have the most up-to-date patients. The highest L value is found in the fourth group, meaning the most loyal patients belong to this group. The second most loyal patient group is the first group. On the other hand, the R values of the first and second groups are the same, placing them in second place regarding recency. In other words, the most loyal and up-to-date patients are found in the fourth and first groups.

When examining which channel patients most commonly came from, it was found that the highest percentage of patients (43%) came via recommendations. This is a critical piece of information. Patients who came through recommendations, either from other patients or from staff, make up the highest proportion in every group. Particularly in the fourth group (most loyal) and the first group (most up-to-date), the percentage of patients coming via recommendation is even higher, at 50% and 48%, respectively. This suggests that the recommendation channel is the most effective regarding patient loyalty and profitability. As expected, the second most common source of patients is the Internet, notably Google and the website. The channel marked as WhatsApp is, in fact, a channel operating on the company's website, so the high effectiveness of Google and the website becomes clearer. The impact of social media appears to be lower. The third group of patients who come through the door (i.e., walk-ins) do not have a channel attribution so that no comments can be made about that. The patients with the highest average return come from fairs (international patients). While the second group consists of patients from overseas branches, the patients from advertisements are also the third group with the highest return. However, the high costs associated with fairs mean whether this channel is the most profitable is a separate issue.

After these evaluations for all patients, conducting separate assessments for each of the four groups would be helpful.

Group 1:

Patients in this group constitute 5% of the total patient population. Their recency and longevity values are higher than average. On the other hand, their financial returns are significantly above average (46,488 TL, compared to the average of 7,568 TL). Although the patients in this group represent 5% of the total patient population, they contribute 33% of the total revenue. The revenue-per-patient ratio (33/5) is 6.6. The distribution of channels through which patients in this group arrived is similar to that of the total patient population, with those arriving via recommendations making up the most significant portion of the group.

Group 2:

This group contains only 22 patients with the highest average revenue (304,057 TL). However, due to the very small number of patients, making meaningful evaluations seems complicated. Indeed, the distribution of arrival channels shows that the percentage of patients coming through the door is equal to that of patients arriving through recommendations in the entire population. This could be considered a random occurrence. Additionally, the patient ratio is extremely small (0.00045), and the revenue-per-patient ratio (0.00045/301,879) is disproportionately high, leading to an unrealistic value (670,842,222). The recency (R) and longevity (L) values for patients in this group are below average. This means these 22 patients are very young and have yet to visit in

the most recent period, but their last visit occurred much earlier than average.

Group 3:

This is the largest group, with 45,867 patients (94%). However, the average revenue per patient is only 4,176 TL, roughly one-third of the overall average. The revenue-per-patient ratio is 0.55 (52/94). The distribution of arrival channels for this group is as expected and similar to that of the total patient population. The recency (R) and longevity (L) values are very close to the average, indicating that patients in this group are relatively balanced regarding how recently they have visited and how long they have been patients.

Group 4:

This group contains 351 patients (1% of the total), with an average revenue per patient of 140,540 TL. This is more than ten times the average revenue (13,118 TL). Thus, the revenue-per-patient ratio (13/1) is 13. The channel distribution for patients in this group is similar to the others. These patients represent the most loyal group, with the highest recency values.

Determining the Target Segment (Group)

Based on the above comments and evaluations, the research's purpose will be achieved by identifying the target segment (group).

To identify the hospital's most ideal (profitable and loyal) patient segment, it would be useful to create a separate table that shows the values of L, R, and M, along with the Revenue per Patient/Patient Ratio. This will make it easier to visualize the data. These values are presented in Table 10.

Table 10. Comparison Table for Identifying the Target Segment

Cls.	Number	L	R	M	Rev Pat. Ratio / Patient Ratio.
1	**2630**	**5**	**314**	**46.488**	**6,6**
2	22	0	244	304.057	670.842.222
3	45.867	3	324	4.176	0,55
4	351	**7**	**314**	**140.540**	**13**
Av.		**2,86**	**323**	**7.568**	

The third group, which has the lowest revenue, is also huge, so targeting it would not be appropriate. Similarly, targeting the second group, which is very small, does not make much sense either. After all, the recency and loyalty of the patients in this group are lower than those of the other groups. Therefore, for a more focused approach, targeting patients from Group 4 and Group 1, who have high revenue, are recent, and loyal, would be the most rational approach.

While Group 4 has much higher revenue and loyalty, the small number of patients means that defining their profile would not be statistically significant, and the total revenue potential may be lower. For this reason, Group 1 should also be considered the target segment. Furthermore, the Revenue per Patient/Patient ratio for both groups is significantly higher than that of the largest group.

VI. CONCLUSION

This study serves as an example of the segmentation and targeting stages—a critical phase in strategic marketing—and focuses on segmenting the patients of a healthcare institution and targeting specific segments.

Following the research, as discussed in the literature section, several studies [29], [19] have demonstrated that the K-Means method, which is considered the most suitable method for the healthcare sector, was used for segmentation according to LRFM criteria. The model was revised according to the hospital's dataset and sector-specific dynamics. The model excluded the F-value, and the most suitable segment cluster was identified. Some of the aspects of this study align with and differentiate from other research encountered in the literature search.

Firstly, as many studies have expressed, it is most appropriate for segmentation studies to be done subjectively due to the unique data each brand possesses, which was also confirmed in this study. In fact, the F-value in the LRFM analysis was excluded, making the model more tailored to the firm's specific characteristics.

One of the significant differences between this research and similar studies is the dataset size. For example, studies like Wei et al. [33] and Wuu et al. [36] in the same sector had datasets of 2,258 and 1,462, while this research utilized a dataset of 48,870, significantly increasing the accuracy of the segmentation. Additionally, the number of segments identified in the mentioned studies was much larger, but limiting the segmentation to just four groups in this study made it easier to develop a marketing strategy focused on the target audience.

Another distinction of this research is that an analysis based on patient arrival channels was also conducted. While no similar study exists in the sector, Nakano and Kondo [22] have conducted a comparable analysis. Even though patient arrival channels did not create significant differences across the clusters, this analysis has helped identify essential marketing activities. It was found that referral channels and internet searches are crucial factors for patient acquisition.

Another unique feature of this research is that it includes the revenue per patient/patient ratio as a criterion for selecting the target segment. While this ratio is not the most critical factor, it serves as a comprehensive indicator for evaluation. Accordingly, the clusters with the highest revenue and the most loyal and up-to-date patients were selected as the target segments. Marketing activities directed at these segments are predicted to increase the hospital's long-term profits. However, one of the most necessary activities, patient profiling (persona) studies, could not be conducted due to data limitations. Nevertheless, based on the data obtained in this study, several recommendations can be made for hospital managers.

Recommendations Regarding Patients:

Targeting all patients in Group 1 and Group 4: These patients should be thanked and offered small gifts or discounts.

To improve the referral channel, "Bring a Friend" campaigns should be organized for staff and customers, with promotions offered to those who refer others.

Collect as much customer data as possible (demographic, psychological, sociological) to gain deeper insights.

Detailed analyses of patients' actions should be conducted, and records should be kept up-to-date to improve segmentation.

Patient characteristics in Groups 1 and 4 should be extracted to carry out future "look-alike sales" activities, and marketing budgets should be allocated to target potential customers with similar profiles**.**

Recommendations Regarding Patient Arrival Channels:

From a digital marketing perspective, the company's resources should be spent more on search engine (Google) searches than social media. The website should continually be updated, and the WhatsApp line should be prioritized.

When the proportion of walk-in patients is high, it is essential to ask them how they heard about the hospital and how they arrived, then define these channels in the system.

Profitability analysis for fairs and international patients should be conducted. While revenue per patient looks pretty high for these channels, especially considering the small number of international patients, ROI (Return on Investment) analysis should be included to evaluate their profitability.

Limitations of the Study and Future Research:

The data collected on patients could be more extensive, meaning the analyses conducted in this study are also constrained. In similar studies, sufficient customer data is assumed to be available, and the task of extracting patient profiles for target segments must be added to such research. Future research could also investigate the costs associated with patient arrival channels. Moreover, analyses of the procedures conducted could be broken down by segment, allowing identification of the most profitable and high-potential procedures.

Authors' Contributions

The authors' contributions to the paper are equal.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics

REFERENCES

[1] M. Alves Gomes, & T. Meisen, (2023). "A review of customer segmentation methods for personalized customer targeting in e-commerce use cases", Information Systems and e-Business Management, 21(3), 527-570.M.

[2] H.H., Channg., S.F., Tsay., '' Integrating SOM and K-man in Data Mining Clustering: An Empirical Study of CRM and Profitability Evaluation''. *Journal of Information Management*, 11(4), 161–203 (2004).

[3] Chen X, Fang Y, Yang M, Nie F, Zhao Z, Huang J.Z. Purtreeclust: a clustering algorithm for customer segmentation from massive customer transaction data. IEEE Trans Knowl Data Eng 30(3):559–572. https://doi.org/10.1109/TKDE.2017.2763620 (2018)

[4] Y.S., Chen, C.H., Cheng, C.J., Lai, , C.Y., Hsu, & Syu, H.J. Identifying patients in target customer segments using a two-stage clustering-classification approach: A hospital-based assessment. Computers in biology and medicine, 42(2), 213-221 (2012).

[5] A.J., Christy, A., Umamakeswari, L., Priyatharsini, & A., Neyaa. ''RFM ranking–An effective approach to customer segmentation.'' *Journal of King Saud University-Computer and Information Sciences*, 33(10), 1251–1257, (2021).

[6] S.,Dolnicar, B. Grün, & F., Leisch. Market segmentation analysis: Understanding, doing, and making it useful. Springer Nature.(2018).

[7] S.,Dolnicar, & K., Lazarevski,. Methodological reasons for the theory/practice divide in market segmentation. *Journal of Marketing Management*, 25(3-4), 357–373. https://doi.org/10.1362/026725709X429791, (2009).

[8] D., Dryglas, & M., Salamaga,. Segmentation by push motives in health tourism destinations: A case study of Polish spa resorts. *Journal of Destination Marketing & Management*, 9, 234–246. https://doi.org/10.1016/j.jdmm.2018.01.008, (2018).

[9] O., Ergun, ''Makine Öğrenmesi Algoritmaları İle Müşteri Segmentasyonu Ve Hepsiburada E-Ticaret Platformu Üzerine Bir Uygulama.'' Uludağ Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, (2023).

[10] Z.A., Hallab, Y., Yoon, & M., Uysal. Segmentation based on the healthy-living attitude: A market's travel behaviour. *Journal of Hospitality and Leisure Marketing*, 10(3/4), 185–198. https://doi.org/10.1300/J150v10n03_12 , (2003).

[11] S., Han, Y., Ye, X., Fu, & Z., Chen. Category role aided market segmentation approach to convenience store chain category management. Decision Support Systems, 57, 296–308. https://doi.org/10.1016/j.dss.2013.09.017 , (2014).

[12] A.M., Hughes, A*Strategic database marketing*. Chicago: Probus Publishing Company, (1994).

[13] S., Kanca, T., Özcan, & Y., Çelikbilek. Bir Tekstil Perakendecisinin Müşterileri İçin RFM Modeli ile Müşteri Segmentasyonu. *The Journal of International Scientific Researches*, *8*(3), 393-409, (2023).

[14] U., Kaymak. ''Fuzzy target selection using RFM variables, in Proceedings of the IFSA World Congress and 20th NAFIPS International Conference,'' vol. 2, 2001, pp. 1038–1043, (2001).

[15] U., Konuş, P.C, Verhoef, & S.A., Neslin. Multichannel shopper segments and their covariates. *Journal of Retailing*, 84(4), 398-413, (2008).

[16] P., Kotler & G., Armstrong. Principles of Marketing, 7Th ed., Englewood Cliffs, NJ: Prentice-Hall, (1996).

[17] J.B., MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1967, pp. 281–297, (1967).

[18] F., Marisa, S.S.S., Ahmad, Z.I.M., Yusof, F., Hunaini and T.M.A., Aziz. Segmentation model of customer lifetime value in small and medium enterprise (SMEs) using K-means clustering and LRFM model. *International Journal of Integrated Engineering*, *11*(3), (2019).

[19] G., McKernan. Customer Segmentatıon Approaches:'' A Comparıson Of Methods Wıth Data From The Medıcare Health Outcomes Survey '' (Doctoral Dissertation, University Of Pittsburgh), (2018).

[20] J.A., McCarty, & M., Hastak. Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of Business Research*, 60(6), 656–662, (2007).

[21] A., Moghaddam, & Harandi. ''An RFMV Model and Customer Segmentation Based on Variety of Products,'' 155-161, (2017).

[22] S., Nakano, & F.N., Kondo. Customer segmentation with purchase channels and media touchpoints using single source panel data. *Journal of Retailing and Consumer Services*, 41, 142-152, (2018).

[23] U.O., Nnaji, L.B., Benjamin, N.L., Eyo-Udo, & E.A., Etukudoh. A review of strategic decision-making in marketing through big data and analytics. Magna Scientia Advanced Research and Reviews, 11(1), 084-091, (2024).

[24] N.T., Nwosu, S.O., Babatunde, & T., Ijomah. Enhancing customer experience and market penetration through advanced data analytics in the health industry. *World Journal of Advanced Research and Reviews*, *22*(3), 1157-1170, (2024).

[25] P., Özkan, & İ.D., Kocakoç. Sağlık sektöründe LRFM analizi ile pazar bölümlendirme. *Kuşadası: Türkiye*, (2019).

[26] E., Sarıoğlu, & M., İnel. Müşteri Segmentasyon Modellerinin Karşılaştırılması Üzerine Ampirik Bir Araştırma. *Öneri Dergisi*, *19*(62), 130-145, (2024).

[27] J.M.C., Schijns, G.J., Schroder. Segment selection by relationship strength, *J. Direct Mark*. 10 (3) (1996) 69–79.), (1996).

[28] T.T., Shi, X.R., Liu, & J.J., Li. Market segmentation by travel motivations under a transforming economy: Evidence from the Monte Carlo of the Orient. Sustainability, 10(10), 3395. https://doi.org/10.3390/su10103395 , (2018).

[29] L., Torkzadeh, H., Jalilian, M., Zolfagharian, H., Torkzadeh, M., Bakhshi, & R., Khodayari-Zarnaq. Market segmentation in the health tourism industry: ''A systematic review of approach and criteria.'' *Journal of Policy Research in Tourism, Leisure and Events*, 16(1), 69-88, (2024).

[30] Y.C., Tsao, P.V.R.P., Raj, & V., Yu. Product substitution in different weights and brands considering customer segmentation and panic buying behaviour. *Industrial Marketing Management*, *77*, 209–220, (2019).

[31] M., Wedel, W., Kamakura.: Market Segmentation: Conceptual and Methodological Foundations. Kluwer Academic Publishers, Boston, 2nd edn. (2000)

[32] J.T., Wei, S.Y., Lin, & H.H., Wu. A review of the application of RFM model. *African journal of business management*, *4*(19), 4199, (2010).

[33] J.T., Wei, S.Y., Lin, C.C., Weng ve H.H., Wu, : "A Case Study of Applying LRFM Model in Market Segmentation of A Children's Dental Clinic,". Expert Systems With Applications, 39(5), 5529-5533, (2012).

[34] J.-T.,Wei, M.-C., Lee, H.-K., Chen, ve H.-H., Wu. ''Customer Relationship Management in the Hairdressing Industry: An Application of Data Mining Techniques. Expert Systems with Applications,'' 2013, 40(7), 7513- 7518, (2013).

[35] H.F., Witschel, S., Loo, & K., Riesen. How to support customer segmentation with useful cluster descriptions. In Advances in Data

Mining: Applications and Theoretical Aspects: 15th Industrial Conference, ICDM 2015, Hamburg, Germany, July 11-24, 2015, Proceedings 15 (pp. 17-31). Springer International Publishing, (2015).

[36] H.-H., Wu, S.-Y., Lin, & Liu, C.-W., Liu. Analyzing patients' values by applying cluster analysis and LRFM model in a pediatric dental clinic in Taiwan. *The Scientific World Journal*, 2014. https://doi.org/10.1155/2014/685495 , (2014).

[37] A.X., Yang. How to develop new approaches to RFM segmentation. *Journal of Targeting, Measurement and Analysis for Marketing, 13*, 50–60, (2004).

[38] I.C., Yeh, K.J., Yang ve T.M., Ting: "Knowledge Discovery on RFM Model Using Bernoulli Sequence". Expert Systems with Applications, 36:5866–5871, (2008).

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 118 - 126
DOI: 10.36287/ijmsit.8.2.11
Received: November 3, 2024; Accepted: December 11, 2024

Machine Learning and Vision Transformer for CT Scanners' Calibration and Quality Assessment

Khanh Quoc Man[1*], Majeed Soufian[2], Amani Mansour Alsaeedi[3], Jon Fulford [4] and Hairil Abdul Razak [5]

[1*] *Department of Computer Science, University of Exeter, Exeter, UK (km827@exeter.ac.uk) (ORCID: 0009-0003-0565-787X)*
[2] *Department of Computer Science, University of Exeter, Exeter, UK (m.soufian@exeter.ac.uk/magid@ieee.org) (ORCID: 0000-0002-8976-9187)*
[3] *Medical School, University of Exeter, Exeter, UK (amda201@exeter.ac.uk) (ORCID: 0000-0000-0000-0000)*
[4] *Medical School, University of Exeter, Exeter, UK (J.Fulford@exeter.ac.uk) (ORCID: 0000-0002-5945-1688)*
[5] *Medical School, University of Exeter, Exeter, UK (H.Abdul-Razak@exeter.ac.uk) (ORCID: 0000-0002-8266-0381)*

Abstract – In this study, we present the process and research for finding the best machine learning methodology and innovative approach to evaluate the image quality in Computed Tomography (CT) scanners by predicting Signal-to-Noise Ratio (SNR) and Contrast-to-Noise Ratio (CNR) from low-resolution CT images of a series of phantoms. Traditional methods of Image Quality Assessment (IQA), reliant on subjective evaluation by radiologists, often suffer from variability and inefficiency. To address these limitations, we explored both interpretable models like the Adaptive Neuro-Fuzzy Inference System (ANFIS) and other advanced deep learning architectures. Initially, ANFIS combined with Gray Level Co-occurrence Matrix (GLCM) features yielded suboptimal results, with an R-squared value of 0.634. Experimenting with various deep learning methodologies for improving the performance, directed us to develop a hybrid model integrating DenseNet, Vision Transformers, and reparameterization techniques, which showed that can achieve superior results with an R-squared value of 0.8892. This research paper focuses on searching for the optimal machine learning model and lays the groundwork for an automated tool that can optimize imaging protocols by providing a comprehensive quality assessment of CT images in CT calibration.
Keywords *– Machine learning, Deep learning, Vision Transformer, CT calibration, IQA.*

Citation: Man, K., Soufian, M., Alsaeedi, A.M., Fulford, J., Razak, H.A. (2024). Machine Learning and Vision Transformer for CT Scanners' Calibration and Quality Assessment. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 118 - 126.

I. INTRODUCTION

Machine learning and artificial intelligence methodologies have been applied increasingly in various medical fields such as medical imaging and pathogen identifications in recent years, started in 3 decades ago [1 and 2], when very few methodologies existed. The computed tomography machines provide pictorial anatomical information about the physiological state of internal organs by using X- rays and gives sensitive discrimination between healthy and diseased tissue. Ensuring the quality of CT images is essential for accurate medical diagnosis. Naturally calibration is a critical step in this process. In CT imaging, calibration is the first and most crucial step to ensure the reliability and accuracy of images used for diagnosis using an object called "Phantom" to simulate the organ. This includes adjusting the CT scanning equipment to correct any errors that might negatively affect image quality. Such calibration should be conducted regularly to maintain accuracy without distorting the images or reducing their value. Any distortion or lack of proper contrast in CT images can lead to diagnostic errors. To support accurate image analysis and the gathering of diagnostic information, producing high-quality CT images is essential.

There are two main methods in IQA, subjective and objective evaluation [3]. While subjective assessment is conducted by experts, such as diagnostic radiologists, objective assessment, is based on using various logical and mathematical algorithms. The subjective evaluation which has been formed by manual qualitative assessment of CT images by radiologists, usually involves identifying and measuring phantom image features. This process is often considered the gold standard but is limited by poor inter-observer agreement and the risk of fatigue and perceptual biases. At the same time, manual assessment by radiologists, is indeed time-consuming and prone to inconsistencies despite it requires significant expertise. These factors can lead to variations in diagnosis and inefficiencies in the workflow. Especially in CT calibration, to evaluate the quality of a phantom image, radiologists are traditionally required to manually indicate the location of the holes in each square in the phantom image [4]. Such challenges underscore the need for automated methods that can consistently and accurately assess CT image quality, particularly during the calibration process. They particularly highlight the need for developing automated methods based on machine learning and artificial intelligence that can reliably evaluate image quality metrics like SNR and CNR, improving efficiency and reducing variability in CT imaging.

This research focuses on the process of finding an innovative optimal machine learning methodology, which can evaluate SNR and CNR in CT images in the most efficient manner and at the same time can be transparent and interpretable. In contrary to 3 decades ago, a vast number of various machine/deep learning methodologies are available,

which makes it difficult to find an optimal model by using each individual one or by combining them together. We tried to cover a wide range of them to solve the problem of automated measurement of SNR and CNR values in the phantoms' hole images. The data consists of 45,500 holes images cropped from phantom CT images, with labels representing the SNR and CNR values of the images, which were manually assigned. We started our research with a simple model called ANFIS, a learning model commonly used in medical imaging due to its ability to integrate fuzzy systems with neural networks and its transparency and interpretability. Evaluating and recognizing the limitations of traditional methods such as ANFIS, we developed many robust solutions and transitioned to more complex wider deep learning models including ResNet, RNN, SE-ResNet, Fast-ViT, SE-ResNet, Unet-NILM and SqueezeNet in order to assess their performance in predicting SNR and CNR values from CT images. These directed us to introduce a hybrid architecture called SynQ-ViT (Synthetic Quality assessment for computed tomography calibration with Vision Transformer), which leverages a hybrid architecture combining DenseNet, Vision Transformers, and reparameterization techniques. This modification enables the model to effectively learn both local and global features. We reported the success of SynQ-ViT for this application with a view from medical imaging separately [5]. Here in the rest of this paper, after portraying related works, searching for an optimal machine learning model, which fit this application best, will be highlighted by presenting above methodologies in some details with their evaluations and experimental results.

II. Related works

Valdes et al. [6] developed a Virtual IMRT QA framework using a machine learning algorithm that accurately predicted gamma passing rates within 3% across different institutions and measurement techniques. In the studies on using ANFIS in medical imaging, Sharma and Mukharjee [7] utilized ANFIS to classify MR images. The integration of ANFIS using GLCM fuzzy rules in medical imaging provided superior classification accuracy when compared to traditional methods like Fuzzy C-Means (FCM) and K-Nearest Neighbor (K-NN). In early detection of COVID-19 through CT image analysis, with the ANFIS-based model achieving superior performance with an accuracy of 98.63% and rapid testing time [8]. In the study of Bahonar et al. [9] ANFIS model significantly outperforms multiple linear regression (MLR) in predicting breast dose during chest CT scans, with a correlation coefficient R of 0.93 and a Root Mean Square Error (RMSE) of 0.172. These findings suggest that ANFIS offers an accurate and efficient approach to medical imaging, especially in CT images.

In recent advancements within CT imaging quality assurance, a deep learning approach using convolutional neural networks (CNN) has been explored to predict whether CT scans meet the minimal diagnostic image quality threshold. Lee et al. [10] introduce a pre-trained VGG19 network was fine-tuned to analyze a dataset consisting of 74 high-resolution axial CT scans, with image quality rated by a radiologist. The network achieved an accuracy of 0.76 and an AUC of 0.78, highlighting the potential of deep learning methods in assessing and ensuring diagnostic quality in CT imaging, despite challenges posed by the relatively small number of cases. Study of a novel Blind Image Quality Assessment (BIQA) method for low-dose CT images [11], utilized a Denoising Diffusion Probabilistic Model (DDPM) and a transformer-based evaluator. The DDPM is employed to generate high-quality primary content from distorted images, mimicking the human visual system's active inference process, while a transformer-based evaluator predicts image quality by integrating this content with a dissimilarity map. Jensen et al. [12] evaluated the performance of a Deep Learning Image Reconstruction (DLIR) algorithm in contrast-enhanced oncologic abdominal CT, comparing it to the standard 30% Adaptive Statistical Iterative Reconstruction V (ASIR-V). The results demonstrated that DLIR significantly improved image quality, particularly at higher strengths, with a notable 4% reduction in noise and a 92-94% increase in contrast-to-noise ratio compared to 30% ASIR-V.

These studies emphasize the role of machine learning models, particularly deep learning models, in medical imaging and CT quality assurance. However, these studies focus on image evaluation during diagnosis rather than during the CT calibration process. In this paper, we address that gap by using a dataset collected during CT calibration. We aim to develop an objective method based on an optimal machine learning methodology to evaluate the quality of CT images during the preparation phase of a CT machine before it is put into use.

III. AI and Machine Learning methodologies

A. *In search of the optimal model*

The optimal model, which will be used as the most effective method for automated CT image quality assessment and calibration, not only should produce optimal accuracy among all other candidate models but also must have optimal number of parameters, easy and quick to train and implement in real medical working environment. The primary means to evaluate the accuracy of each model, is the Mean Squared Error (MSE) for the error of SNR and CNR, and the coefficient of determination, R^2 (R-squared) as an additional performance metric, which are defined in the next section. The optimal model must also be transparent and interpretable while capturing both local and global image features effectively. In search of such model, first ANFIS was considered.

B. *Basic model*

The main advantage of ANFIS [13] is in its transparent and interpretable architecture (Fig. 1) in the form of fuzzy "if-then" rules, which made it our first choice for this application.

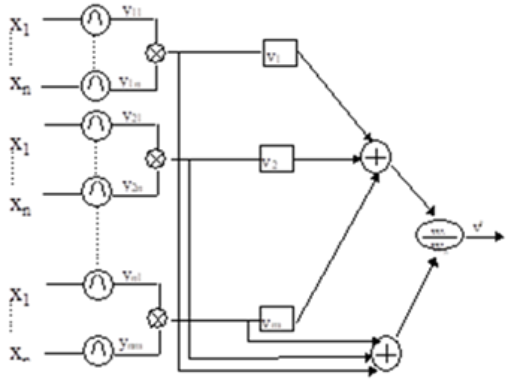


Fig 1. A typical structure of Adaptive Neuro-Fuzzy Inference System with n inputs (X_1 ... X_n) and one output (v') as bock diagram.

ANFIS is a hybrid intelligent system that integrates the benefits of both artificial neural networks from machine learning and fuzzy logic from artificial intelligence, allowing it to model complex, nonlinear relationships effectively. In this study, ANFIS was employed to predict SNR and CNR from phantom hole images, following steps below:

1) Feature Extraction: Before applying the ANFIS model, we first extract features from the CT images using GLCM. GLCM is a statistical method that analyzes the spatial relationships of pixels in an image. It is particularly useful for capturing texture information and is widely used in medical imaging, especially in CT [14]. The GLCM computes how frequently pairs of pixels with specific values and in a specified spatial relationship occur in an image, generating a matrix from which various texture features can be derived. The formula for GCLM feature extraction [15] included:

- Contrast: Measures the intensity contrast between a pixel and its neighbor over the whole image.

$$Contrast = \sum_{i,j} |i-j|^2 \, P(i,j) \tag{1}$$

- Dissimilarity: Similar to contrast but provides a more direct measure of the difference between pairs of pixels.

$$Dissimilarity = \sum_{i,j} |i-j| \, P(i,j) \tag{2}$$

- Homogeneity: Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$Contrast = \sum_{i,j} \frac{P(i,j)}{1+|i-j|} \tag{3}$$

- Energy: Provides the sum of squared elements in the GLCM, reflecting image uniformity.

$$Contrast = \sum_{i,j} P(i,j)^2 \tag{4}$$

- Correlation: Measures how correlated a pixel is to its neighbor over the whole image.

$$Contrast = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i \sigma_j} \tag{5}$$

In above formulations, $P(i,j)$ represents the probability of the co-occurrence of pixel pairs separated by a specific distance and angle, μ_i, μ_j and σ_i, σ_j are the means and standard deviations of the marginal distributions of i and j respectively.

2) Rule Generation and Fuzzy Inference: Once the features are extracted, the ANFIS model processes these features as input data. Each input variable is associated with a fuzzy membership function. ANFIS generates a set of fuzzy if-then rules based on all possible combinations of membership functions across the input variables. For instance, a typical rule might state that "if the contrast is high and the homogeneity is low, then the SNR will be high". The firing strength of each rule is calculated as the product of the membership values for the input variables involved in the rule:

$$w_j = \prod_{i=1}^{n} \mu A_{i,j} \, (x_i \tag{6}$$

where $\mu A_{i,j}$ is the membership value of input x_i in a fuzzy set $A_{i,j}$.

3) Normalization and output: The final output of the ANFIS model is a weighted sum of the normalized firing strengths and the corresponding linear functions of the inputs:

$$y = \sum_{j=1}^{M} \overline{w}_j f_j \, (x) \tag{7}$$

4) Evaluation: Let $\hat{y}_i = (\hat{y}_{i1}, \hat{y}_{i2})$ denote the predicted values for SNR and CNR, and $y_i = (y_{i1}, y_{i2})$ denote the true values. The overall MSE loss function measures the average squared difference between the predicted and true values:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} [(y_{i1} - \hat{y}_{i1})^2 + (y_{i2} - \hat{y}_{i2})^2 \tag{8}$$

where n is the number of samples. The overall R^2 metric assesses the proportion of variance in the dependent variables that is predictable from the independent variables, i.e.:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} [(y_{i1} - \hat{y}_{i1})^2 + (y_{i2} - \hat{y}_{i2})^2)]}{\sum_{i=1}^{n} [(y_{i1} - \bar{y}_1)^2 + (y_{i2} - \bar{y}_2)^2)]} \tag{9}$$

where $\bar{y}_1$ and $\bar{y}_2$ are the mean values of the true SNR and CNR, respectively.

5) Result: After training, ANFIS achieved an MSE of 39.9, indicating a large deviation between the predicted and actual values of the image quality metrics. Additionally, obtained R^2 of 0.634, suggests that the model could only explain 63.4% of the variance in the overall SNR and CNR values, leaving a considerable portion of the variability unaccounted for. While ANFIS provides a valuable framework for clear understanding and transparent modeling relationships in data, its application to the prediction of SNR and CNR in this study has not yielded satisfactory results.

C. Other Advanced Machine and Deep Learning Models

It's possible to improve the ANFIS performance by some clustering methods [16] however, it is anticipated that trying more complex and sophisticated models from a wide range of machine and deep learning approaches would increase the likelihood of achieving a better predictive performance and accuracy in terms of MSE and R^2 metrics, guiding us toward obtaining an optimal model as necessary condition. Although other metrics such as having minimum number of parameters, the ease and speed of training and implementation in real

medical working environments are important and will also be considered for finding the optimal model, the main drawbacks of machine and deep learning models are their lack of transparency and interpretability. To address these issues, a method is employed to visualize and interpret a model's decision-making process without changing its parameters and will be discussed later. A model with optimal MSE, R^2 and other metrics, which fails to highlight important areas such as the Region Of Interest (ROI) in the CT images, will not be considered as the optimal model.

Apart from ANFIS, for the same dataset, seven other models from a wide range of machine and deep learning methodologies were developed, which are explained in next subsections. For successful training and to optimize the performance of our models, we conducted an extensive hyper-parameter optimisation for each model training. The tuning process involved using a Random Search strategy, where 20 trials were executed to explore the hyperparameter space. Early stopping was implemented to monitor the validation loss, with a patience of 10 epochs. If the validation loss did not improve for 5 consecutive epochs, a callback was employed to decrease the Adam optimiser learning rate logarithmically. Further details are presented in the experimental section.

D. ResNet [17]

Residual Networks (ResNet), is a highly influential deep learning architecture that uses skip connections, allowing for the training of very deep networks without the issues of vanishing or exploding gradients. The architecture consists of a series of stacked residual blocks, where each block includes identity mappings and convolutional layers. It is widely used in various medical image tasks [18]. Figure 2 is the idea of ResNet with 1 residual result at each layer. After training, ResNet achieved an overall MSE of 22.01, indicating a much smaller deviation between the predicted and actual values of the image quality metrics with a good overall R^2 of 0.85.

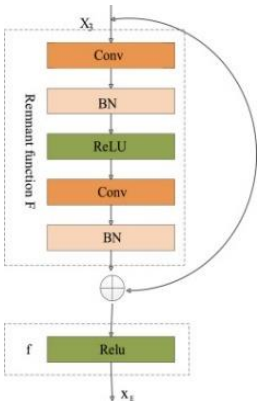


Figure 2. ResNet with 1 residual result at each layer [18]

E. RNN [19]

Recurrent Neural Networks (RNN) are a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence. The ability of RNN to maintain a memory of previous inputs is due to their feedback loops, which allow information to persist over time, thus enabling the network to capture patterns and dependencies that unfold across long sequences. In tasks related to medical imaging, particularly CT scanners, the RNN have been used as benchmarks for both GANs and deep learning networks [20]. After training, RNN achieved an MSE of 28.40 and a R^2 of 0.81. Figure 3 presents, the results of RNN performance during training for predicting SNR and CNR values from CT images of phantom holes in terms of overall R^2 for both training and validation datasets.

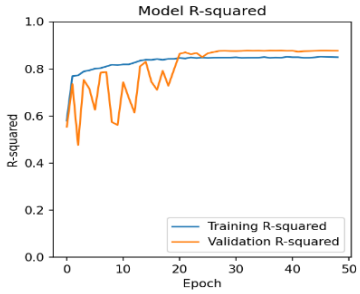


Figure 3. Performance of RNN during training

F. SE-ResNet [21]

This model enhances the traditional ResNet model by incorporating Squeeze-and-Excitation (SE) blocks, hence it is called SE-ResNet. The SE block operates by first applying global average pooling to squeeze global spatial information into a channel descriptor. This descriptor is then passed through a pair of fully connected layers to capture channel-wise dependencies, followed by a sigmoid activation to generate channel weights. It was used in various task of diagnosis from CT images [22]. After training, it achieved an MSE of 17.32 and a R^2 of 0.88. The results of SE-ResNet performance during training are presents in Figure 4 for both training and validation datasets.

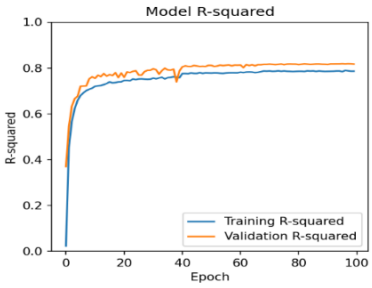


Figure 4. Performance of SE-ResNet during training

G. UNET-NILM [23]

This model leverages a one-dimensional U-NET-based Convolution Neural Networks (CNN) architecture for Non-Intrusive Load Monitoring (NILM), hence called UNET-NILM. It enables simultaneous appliance state detection and power consumption estimation [23]. By combining down-sampling and up-sampling blocks, it captures both local and global features of power signals effectively [24]. Figure 5 illustrating the structure of UNET-NILM with the idea of utilizing a U-Net architecture, originally designed for image segmentation, to effectively separate and identify individual electrical appliances' power usage from aggregated energy consumption data. By leveraging the encoder-decoder structure of U-Net, the model learns both local and global features, making it well-suited for accurately disaggregating energy signals at various levels of granularity. After training, UNET-NILM achieved an MSE of 20.24 and a R^2 of 0.86.

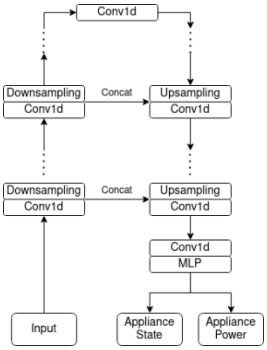


Figure 5. Architecture of UNetNILM [24]

H. SqueezeNet [25]

The model designed to achieve AlexNet-level accuracy on ImageNet with 50 times fewer parameters, reducing the model size to less than 0.5MB. The architecture accomplishes this through the use of "Fire modules," which combine 1x1 and 3x3 filters and late down-sampling to maximize accuracy while minimizing parameter count. This makes processing low-resolution images efficient. The efficiency of SqueezeNet for low-resolution medical images proved by Zhang et al. [26]. After training, SqueezeNet achieved an MSE of 17.55 and a R^2 of 0.88. Figure 6 presents, the results of SqueezeNet performance during training for both training and validation datasets.

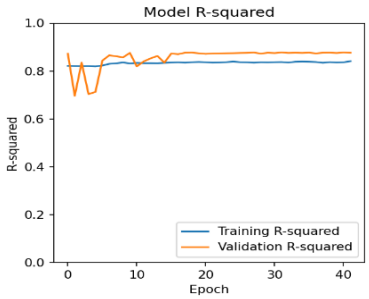


Figure 6. Performance of SqueeNet during training.

I. FastViT [27]

Fast Vision Transformer (FastViT) is a hybrid vision transformer architecture that combines the efficiency of CNN with the global context modeling capabilities of transformers. The key innovation of FastViT lies in its use of the RepMixer block, a novel token-mixing operator that employs structural parameterization and capacity to learn complex patterns. After training, FastViT achieved an MSE of 18.51 and a R^2 of 0.87.

J. The optimal model

The above developments for ample number of the CT input images with small sizes (6 to 9 pixels), imposed a model architecture that performs well with a low number of input parameters. It means the model must be able to capture both global and local features in large volume of datasets, which was also suggested by Talab et al. [28] for low-resolution images. As a result, we proposed SynQ-ViT, as optimal model that focuses on learning both local and global features linearly. For local features, we employed dense blocks from DenseNet [29] in it due to their ability to effectively learn through feature reuse. Adding attention mechanism from Vision Transformers (ViT) aids the model to capture long-range dependencies and contextual information across the entire images [30] or global features. RepMixer [27] was also introduced in our model for achieving efficient token mixing, reducing computational overhead through structural reparameterization, and enhancing the model's capacity to learn complex patterns. These are illustrated in Figure 7 showing SynQ-ViT model in train-inference phases with the dense block (Fig. 7a), transition block (Fig. 7b), RepMixer block (Fig 7c.1 in training mode and Fig 7c.2 in inference mode) and attention block (Fig 7d). These have also been explained in our other study [5] in some details. The model learns its parameters from data by passing outputs sequentially through its layers during training. During inference, by structural reparameterization the Batch Norm and skip connections are removed for simplifying RepMixer structure. This reduces computational overhead and memory access costs as shown by Weng et al. [31] that removing skip connections can improve computational efficiency and reduce resource requirements without significantly compromising accuracy. These are important for real-time applications and hardware deployments, especially when processing large volumes of CT images. The output layer is designed to predict SNR and CNR values by using a dense layer to transform the final feature representations into them. After training, it achieved an MSE of 16.03 and a R^2 of 0.89. Figure 8 presents, the results of SynQ-ViT performance during training for both training and validation datasets.

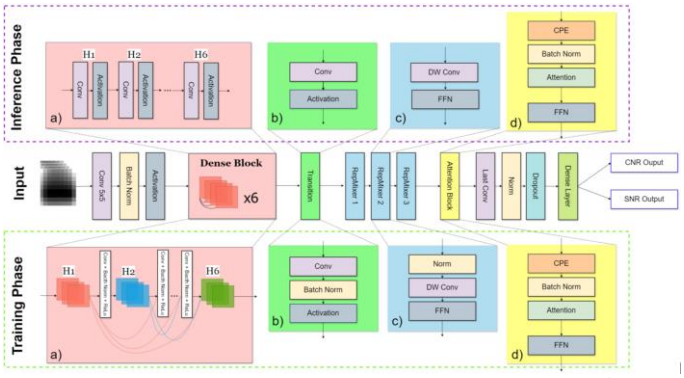


Fig 7. The optimal model architecture SynQ-ViT in training and inference phases: a) dense block, designed to maximize feature reuse by directly connecting each layer using H functions, which improves information flow and supports efficient learning with fewer parameters. b) transition block, reduces dimensionality, aiming to cut down computational overhead while retaining essential features for further processing. c) RepMixer block, plays a crucial role in optimizing the model's structure; during training, it incorporates skip connections for performance, but in the inference phase it removes both batch normalization and skip connections to reduce memory and computational costs. This structural reparameterization makes the model more efficient in real-time applications. d) attention block is used for token mixing, focusing on the most relevant parts of the input datasets and ensuring that the model captures key dependencies and patterns across the entire image.

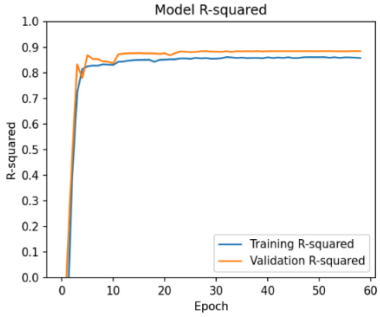


Fig 8. Performance of SynQ-ViT during training [5]

IV. Evaluation and Experimental Results

A. The Datasets

As stated earlier, the data acquired in this study consists of 45,500 holes images cropped from specifically designed Perspex phantom 500 CT images, with labels representing the SNR and CNR values that were manually calculated. The phantom was injected with AuNPs (0.005mg/ml) and scanned using a CT scanner (Biograph Vision 600 Siemens Definition Edge 128) under a variety of exposure settings to rigorously evaluate image quality metrics [5]. Figure 9 shows three randomly selected images from the dataset, which resized to 9x9 pixels to standardize the input dimensions. The presence of negative values in the SNR and CNR during data acquisition, led to implementing a filtering process to ensure their removal and consequence normalization to preserve the accuracy and reliability of the datasets.

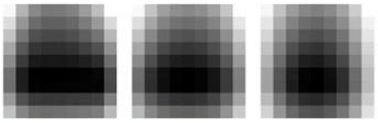


Fig 9. Three random sample images from the dataset in TIFF format. Each of these images represents a hole in phantom images [5].

B. The experiment design

After data acquisition and obtaining accurate and reliable datasets and exploring them, many models from a wide range of machine and deep learning methodologies were exploited to cover all possible models that satisfy the success criteria established in the sections III.A and III.C and to guide us toward an optimal model. The training and hyperparameter optimisation performed as mentioned in the section III.C If required in some models, the number of blocks within each stage was tuned between 1 and 4 with a growth rate between 12 and 48, the layer scale parameter was also adjusted logarithmically between 10^{-6} to10^{-4}, and both dropout and drop connect rates were varied between 0.0 and 0.5. After training, the performance of each model for predicting SNR and CNR values from CT images of phantom holes in terms of overall MSE and R^2 metrics were calculated and discussed in sections III.D to III.J, which are summarized in figure 10.

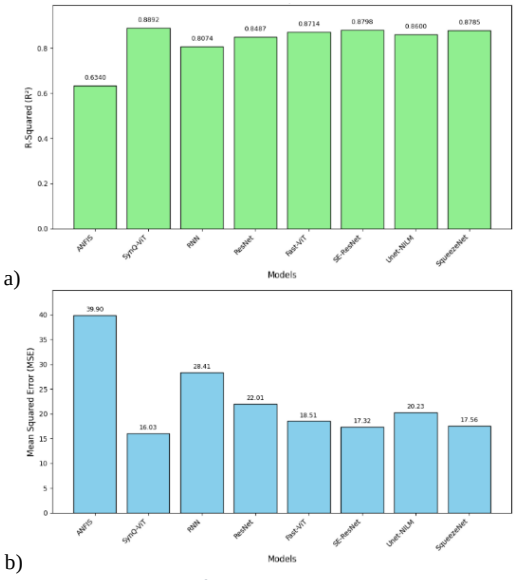


Fig 10. The comparison of a) R^2 and b) MSE performance across all models.

Considering only overall MSE and R^2 metrics, SynQ-ViT was suggested as the optimal model in section III.J. However, other success criteria and metrics are also important for evaluating each model in other to confirm the optimal one for this application as presented in the next section.

C. Models Evaluation

This section discusses also other success criteria and optimality conditions such as having highest speed, minimum number of parameters, the ease in development and implementation in real medical working environments. These can collectively be considered as working experience with each model and are shown in Table 1, including metrics such as epochs, and training time. Please note that one can roughly state that the number of parameters is inversely proportional to the speed of model in inference phase, i.e., a smaller number of parameters is a metric of desired implementation and ease of use in real medical working environments.

Table 1. Experience of working with each model.

Architecture	Epochs	number of Parameters	Training time
ANFIS	10	-	1 hour
SynQ-ViT	61	145,902	~ 1 hours
RNN	64	334,082	~ 1 hour
Resnet	87	118,002	~ 20 mins
Fast-ViT	36	3,210,530	~ 8 hours
SE-ResNet	44	8,672,624	~ 1 day
Unet-NILM	25	2,032,578	~ 6 hours
SqueezeNet	32	113,026	20 mins

- Comparative Analysis for Predicative Performance:

To validate that SynQ-ViT offers an optimal solution among the other AI and machine/deep learning models developed in this study, we conducted a thorough comparative analysis. For this purpose, each model is considered to serve as a benchmark to evaluate the effectiveness of SynQ-ViT in predicting image quality metrics, ensuring it delivers superior and reliable results. Figures 10 already showed SynQ-ViT achieved an impressive R^2 value of 0.89 and an MSE of 16.03 after 61 epochs on the validation set, with a convergence occurring after the 5th epoch (Figure 8). While complex models like SE-ResNet, FastViT and Unet-NILM also demonstrated robust performance, they did not surpass SynQ-ViT in terms of R^2 and MSE. On the other hand, among simpler models only SqueezeNet showed a close predictive power to SynQ-ViT while Resnet and RNN were found to be less suitable for this task, exhibiting notably lower performance metrics.

- Comparative Analysis for Working Experience:

From working experience with each model point of view, while models like FastViT and SE-ResNet performed well indicate strong predictive performance, these models require significant computational resources, with largest parameter counts of 3,210,530 and 8,672,624, and longest training time of approximately 8 hours and one day respectively. This could be inefficient, challenging and disadvantageous when developing and implementing a system that needs to handle large input volumes and operate in real time in medical working environments. On the other hand, models with smaller parameter counts, such as ResNet and RNN (with

118,002 and 334,082 parameters and training time of approximately 20 minutes and 1hour respectively), did not perform well. UNET-NILM demonstrated average performance, with neither parameter count nor efficiency standing out significantly.

Among the models compared, only SqueezeNet approached the performance of SynQ-ViT, with much lower parameter size (113,026) and 3 times faster training time (approximately 20 minutes) than SynQ-ViT, could claim the optimal model title. This required further detailed analysis of the performance during training and validation results. Notably, SynQ-ViT, converged around the 5th epoch (blue line in Figure 8) producing a stable validation result quickly (orange line in Figure 8) while that is not the case with SqueezeNet (orange line in Figure 6). Indeed, compared to the training results of all models, SynQ-ViT has achieved the highest stability and convergence, offering top working experience performance among the best predictive models, yet it requires to pass the last evaluation below before establishing itself as the optimal model for this application.

- Transparency and Interpretability:

As mentioned earlier, any candidate model that fails transparency and interpretability criteria, will not be considered as the optimal model. The transparency and interpretability will be evaluated by the model's ability to highlight important areas such as the ROI for the CT images in this application. For this purpose, a method called Gradient-weighted Class Activation Mapping (Grad-CAM) by Selvaraju et al. [32] has been utilized to visualize and interpret a model's decision-making process without changing its parameters. The Grad-CAM addresses the lack of transparency and interpretability by producing heatmaps and highlighting the regions in the heatmaps that the model relied on for its predictions. In order to do so, Grad-CAM uses randomly selected images and the corresponding final layers of the model.

We applied Grad-CAM to the final convolutional layers of SynQ-ViT as shown in Figure 11, to verify and visualize the areas of focus when SynQ-ViT making its decision. Figure 12 illustrates the heatmaps generated by Grad-CAM from randomly selected images and the corresponding final layers of SynQ-ViT, highlighting the regions that SynQ-ViT relied on for its predictions. Analysis of these heatmaps determine whether SynQ-ViT has been focusing on important areas such as the ROI or not. The red areas in the heatmap correspond to the ROI in the CT images. These visualizations confirm that this model accurately focuses on the critical areas of the input images, thereby validating its effectiveness and reliability in predicting SNR and CNR values as the optimal model.

V. Conclusion, Discussion and Further Work

In this paper we presented our research journey for finding an innovative optimal machine/deep learning methodology, which can evaluate SNR and CNR in CT images in the most efficient manner and at the same time can be transparent and interpretable. After acquiring the required datasets and defining the domain challenge in some details, main success criteria and metrics for achieving optimal predictive performance and working experience were defined. Considering the importance of acquired datasets in the training and evaluation phases, a rigorous preprocessing phase was implemented for ensuring the uniformity, fidelity, accuracy and reliability of the datasets by applying appropriate filtering, normalization, standardization and other procedures for image data in TIFF format, along with the associated SNR and CNR target values.

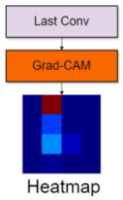


Fig 11. Showing how Grad-CAM was applied to the final convolutional layers in Figure 7 just before the model makes its prediction.

Based on the acquired datasets and the nature of the domain dynamics, eight models from a wide range of AI, machine and deep learning methodologies were consider as optimal candidate models. For successful training and to optimize the performance of our models, we conducted extensive hyperparameter optimisation experiments for each model training. The first choice in choosing an optimal model was the ANFIS model due to its inherent transparency interpretability and explainability properties. The ANFIS model also proved the optimal choice for working experience metrics because of its fast training (10 epochs in an hour with a rapider convergence) and smaller parameter set but underperformed predictively with an R-squared value of just 0.63, indicating its inadequacy for this application. These experiments revealed that each model has its own advantages and limitations when applied to CT image quality assessment. For instance, models like SE-ResNet and FastViT, despite their competitive R-squared values, require high computational resources, making them less feasible for real-time applications. On the other hand, simpler models such as RNN showed limited predicative performance, as reflected by their lower R-squared values. While models with smaller parameter counts like SqueezeNet and Unet-NILM came close to SynQ-ViT's predictive performance, they did not achieve the same level of stability and efficiency. This highlights the importance of balancing model complexity and efficiency, and SynQ-ViT demonstrates an optimal combination, achieving highest accuracy and stability with lower computational demands. Please note that although the model converged quite early at the 5th epoch, it was able to further reduce the error and that is why training continued until epoch 61. Finally, analyzing heatmaps created by Grad-CAM validated the interpretability of the SynQ-ViT's predications.

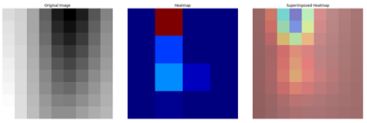


Fig 12. Example of Grad-CAM created heatmaps from a randomly selected original image (left). The middle heatmap indicating the red areas that the model focuses on to predict SNR and CNR, and the right image is the heatmap superimposed on the original image. highlighting the regions the model relied on for its predictions.

There is a discussion about how the optimal model obtained its predictive power and if it is possible to achieve higher

predictive performance while keeping other metrics at optimal. We introduced SynQ-ViT, as a model that combines DenseNet, attention mechanisms, and reparameterization techniques to efficiently learn both local and global features from CT images of phantom holes. The predictive performance and early convergence with low-resolution input images can be attributed to SynQ-ViT's architecture, which effectively reuses learned features from previous layers, allowing for effective feature extraction and model optimization. However, the question on the possibility of increasing its predictive performance, say a R^2 value of above 0.95, is an open research challenge. When compared to other advanced models, SynQ-ViT consistently achieved superior accuracy while maintaining a lower parameter count, demonstrating its efficiency, particularly in real-time applications involving large datasets. Its rapid convergence and ability to handle resource constraints make it an ideal candidate for clinical deployment, where timely processing is critical.

Predicting SNR and CNR for each hole in the phantom image is a crucial first step toward creating a robust quality assessment tool for CT calibration. Moving forward, our goal is to expand this model into a comprehensive system that aggregates these predictions to provide a holistic quality evaluation of the entire phantom image. This tool would help automate imaging protocol optimization in clinical settings, advancing medical imaging and improving patient care.

ACKNOWLEDGMENT

The first author, Mr Khanh Quoc Man, would like to acknowledge Dr Mohammed M. Abdelsamea from Department of Computer Science at University of Exeter, for introducing him to the team.

Authors' Contributions

The authors' contributions to the paper are equal.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics

REFERENCES

[1] Soufian, M., Robinson F.V.P., and Soufian M.,1996, Fuzzy Logic Controller for Whole Body NMR imaging. IEE Colloquium on Fuzzy Logic Controllers in Practice (Digest No. 96/200). London, U.K, DOI: 10.1049/ic:19961125.

[2] Bright J. J., Claydon M. A., Soufian M., and Gordon D. B., 2002, Rapid typing of bacteria using Matrix-Assisted Laser Desorption Ionisation Time-of-Flight Mass Spectrometry and Pattern Recognition Software, Journal of Microbiological Methods, Vol. 48, Issue 2-3, pp 127-138, https://doi.org/10.1016/S0167-7012(01)00317-7 . PMID:11777563

[3] Kumar, R. and Rattan, M., 2012. Analysis of various quality metrics for medical image processing. *International Journal of Advanced Research in Computer Science and Software Engineering*, *2*(11), pp.137-144.

[4] Wang, C.L., Wang, C.M., Chan, Y.K. and Chen, R.T., 2012. Image-quality figure evaluator based on contrast-detail phantom in radiography. *The International Journal of Medical Robotics and Computer Assisted Surgery*, *8*(2), pp.169-177.

[5] Khanh, Q. M., Alsaeedi, A. M., Soufian, M., Fulford, J. and Razak, A. H., 2024. SynQ-ViT: Synthetic Image Quality Assessment for CT Calibration with Vision Transformer, Submitted to IEEE-Engineering in Medicine & Biology Society Conference on Biomedical Engineering and Science (IECBES2024), Penang, Malaysia.

[6] Valdes, G., Scheuermann, R., Hung, C.Y., Olszanski, A., Bellerive, M. and Solberg, T.D., 2016. A mathematical framework for virtual IMRT QA using machine learning. Medical physics, 43(7), pp.4323-4334.

[7] Sharma, M. and Mukharjee, S., 2012. Artificial neural network fuzzy inference system (ANFIS) for brain tumor detection. arXiv preprint arXiv:1212.0059, pp.1-5.R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.

[8] Hossam, A., Fawzy, A., Elnaghi, B.E. and Magdy, A., 2022. An intelligent model for rapid diagnosis of patients with COVID-19 based on ANFIS. In Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2021 (pp. 338-355). Springer International Publishing.

[9] Bahonar, B.M., Changizi, V., Ebrahiminia, A. and Baradaran, S., 2023. Prediction of breast dose in chest CT examinations using adaptive neuro-fuzzy inference system (ANFIS). *Physical and Engineering Sciences in Medicine*, *46*(3), pp.1071-1080.

[10] Lee, J.H., Grant, B.R., Chung, J.H., Reiser, I. and Giger, M., 2018, March. Assessment of diagnostic image quality of computed tomography (CT) images of the lung using deep learning. In *Medical Imaging 2018: Physics of Medical Imaging* (Vol. 10573, pp. 399-405). SPIE.

[11] Shi, Y., Xia, W., Wang, G. and Mou, X., 2024. Blind ct image quality assessment using ddpm-derived content and transformer-based evaluator. *IEEE Transactions on Medical Imaging*.

[12] Jensen, C.T., Liu, X., Tamm, E.P., Chandler, A.G., Sun, J., Morani, A.C., Javadi, S. and Wagner-Bartak, N.A., 2020. Image quality assessment of abdominal CT by use of new deep learning image reconstruction: initial experience. *American Journal of Roentgenology*, *215*(1), pp.50-57.

[13] Jang, J.S., 1993. ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, *23*(3), pp.665-685.

[14] Korchiyne, R., Farssi, S.M., Sbihi, A., Touahni, R. and Alaoui, M.T., 2014. A combined method of fractal and GLCM features for MRI and CT scan images classification. *arXiv preprint arXiv:1409.4559*.

[15] Ramamurthy, B. and Chandran, K.R., 2012. Content based medical image retrieval with texture content using gray level co-occurrence matrix and k-means clustering algorithms. *Journal of Computer Science*, *8*(7), p.1070.

[16] Soufian M., Molaei M., and Nefti S., 2017, Adaptive clustering based inclusion and computational intelligence for fed-batch fermentation process control. In IEEE Development in eSystem Engineering (DeSE), Paris, France. DOI: 10.1109/DeSE.2017.45.

[17] Targ, S., Almeida, D. and Lyman, K., 2016. Resnet in resnet: Generalizing residual architectures. *arXiv preprint arXiv:1603.08029*.

[18] Xu, W., Fu, Y.L. and Zhu, D., 2023. ResNet and its application to medical image processing: Research progress and challenges. *Computer Methods and Programs in Biomedicine*, *240*, p.107660.

[19] Grossberg, S., 2013. Recurrent neural networks. *Scholarpedia*, *8*(2), p.1888.

[20] Zhang, H. and Qie, Y., 2023. Applying deep learning to medical imaging: a review. *Applied Sciences*, *13*(18), p.10521.

[21] Thiruppathi, K., Selvakumar, K. and Shenbagavel, V., 2023. SE-RESNET: Monkeypox Detection Model. *International Journal of Advanced Computer Science and Applications*, *14*(9).

[22] Abdelrahman, A. and Viriri, S., 2023. FPN-SE-ResNet model for accurate diagnosis of kidney tumors using CT images. *Applied Sciences*, *13*(17), p.9802.

[23] Faustine, A., Pereira, L., Bousbiat, H. and Kulkarni, S., 2020, November. UNet-NILM: A deep neural network for multi-tasks appliances state detection and power estimation in NILM. In *Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring* (pp. 84-88).

[24] Virtsionis Gkalinikis, N., Nalmpantis, C. and Vrakas, D., 2023. Variational regression for multi-target energy disaggregation. *Sensors*, *23*(4), p.2051.

[25] Iandola, F.N., 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. *arXiv preprint arXiv:1602.07360*.

[26] Zhang, W., Li, J. and Qiu, X., 2019, December. SAR image super-resolution using deep residual SqueezeNet. In *Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing* (pp. 1-5).

[27] Vasu, P.K.A., Gabriel, J., Zhu, J., Tuzel, O. and Ranjan, A., 2023. FastViT: A fast hybrid vision transformer using structural

reparameterization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 5785-5795).

[28] Talab, M.A., Awang, S. and Ansari, M.D., 2020. A Novel Statistical Feature Analysis-Based Global and Local Method for Face Recognition. *International Journal of Optics*, *2020*(1), p.4967034.

[29] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[30] Li, Y., Wang, J., Dai, X., Wang, L., Yeh, C.C.M., Zheng, Y., Zhang, W. and Ma, K.L., 2023. How does attention work in vision transformers? A visual analytics attempt. *IEEE transactions on visualization and computer graphics*, *29*(6), pp.2888-2900.

[31] Weng, O., Marcano, G., Loncar, V., Khodamoradi, A., Sheybani, N., Meza, A., Koushanfar, F., Denolf, K., Duarte, J.M. and Kastner, R., 2024. Tailor: Altering skip connections for resource-efficient inference. *ACM Transactions on Reconfigurable Technology and Systems*, *17*(1), pp.1-23.

[32] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 127-132
DOI: 10.36287/ijmsit.8.2.12
Received: November 6, 2024; Accepted: December 16, 2024

Cloud Computing Based Smart Irrigation System for Big Farms

Haider Abdulamer kamel[1*], Laith Ali Abdul Rahim[2]

[1*]*Department of Electrical Engineering, Faculty of Engineering, Kirkuk University, Kirkuk, Iraq (haideramer_hwj@ntu.edu.iq) (ORCID: 0000-0002-6916-4463)*
[2]*1Department of Electrical Engineering, Faculty of Engineering, Babylon University, Babylon, Iraq (drlaithanzy@uobabylon.edu.iq) (ORCID: 0000-0001-8064-4401)*

Abstract – The main purpose of this study is to determine the optimum consumption of water and the required water consumption in agricultural irrigation systems. A smart irrigation system includes utilizing innovation to optimize and computerize the watering of plants. It regularly utilizes sensors, climate data, and automation systems to deliver the appropriate amount of water at the right time, thus increasing efficiency and conserving resources. The increase in water needs compared to the increase in population growth requires management of water sources and saving consumption. With progression in innovated technologies, we will set up a system that controlled the irrigation such that there's productive usage of water and make an ease of work for the farmers. By using internet of things and embedded technology, a cloud based smart irrigation system have been implemented in this work.

In this system, the required amount of water is accurately supplied to the plants by obtaining the required information regarding moisture of soil, temperature levels and changes in lighting intensity, which is provided via sensors. These sensors are connected through peripheral devices deployed in the work field to collect information about the weather and soil condition and send this data to the nearest wireless server that will store it on the cloud. Field workers will be able to monitor changes in parameters through dashboards on a website integrated with cloud storage. IoT utilization enables the workers in the field to make an estimation of the required amount of water within the upcoming days. These technological means enable us to study, compare and analyze data for different times during the year and find different ways to reduce and conserve water consumption.

Keywords *– Smart irrigation, Internet of Things (IoT), Sensors, Cloud computing, Water conservation*

Citation: Kamel, H., AL-anzy, L. (2024). Cloud Computing Based Smart Irrigation System for Big Farms. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 127-132.

I. INTRODUCTION

Water resources are considered as one of the most important elements for all living organisms on Earth. Failure to manage water properly can result in the loss of these vital resources. Depletion and unnecessary and excessive energy consumption, as well It causes effects such as decreased product efficiency [1]. That is why many countries are developing new strategies to increase sustainability. Food production and emphasizing the require for successful management of water resources. The greater efficacy of automated irrigation planning may be related to its capacity to self-tune the watering requirements based on the specific characteristics of each plot and its quicker responsiveness to changes in soil, plant, and weather conditions [2]. Creating an ideal irrigation strategy requires experience as well as making the necessary measurements. So, Irrigation is planned in many countries In line with the experiences of farmers and technical advisors [3]. In this context, achieving precise and sustainable irrigation enables farmers to save time and reduce the need for expertise, it reveals the fact that they need to implement irrigation using digital and technological infrastructure.

Beside technological developments, smart systems play an important role in agricultural irrigation planning and irrigation systems. This depends on the system's ability to meet the requirements of the self-irrigation system according to each agricultural plot of land, in addition to responding quickly to changes in soil, plant and air conditions. Mathematical methods such as control models help determine and implement optimal irrigation management [4]. One of the water planning methods is the Food and Agriculture Organization Penman-Monteith Method (FAO-PM) recommended by the Food and Agriculture Organization of the United Nations (FAO). Automated water system planning based on the FAO water balance model and humidity sensors has been the subject of significant investigate over the past few decades [5]. The FAO Soil-Water Balance approach is often used to calculate irrigation requirements by comparing inputs to outputs in the soil and plant system.

With new generation advanced technology, simple and detailed methods have been used to improve smart irrigation systems. One of the simplest approaches has been automatic systems that turn the irrigation system on or off when soil moisture is above or below pre-set thresholds. In slightly more complex systems, water amounts are determined using feedback from plant or soil sensors. These and similar irrigation systems are insufficient to determine the water needs of the plant, cause more water to be consumed than necessary, and cannot optimize the energy and cost spent [6]. Energy costs make up about 40%-60% of the water costs spent in irrigation systems. Today's shortage of water resources and

increased energy demand in irrigation systems make optimal irrigation planning mandatory.

In this project report, a smart irrigation system will be proposed that relies on technologies used in Internet of Things systems, by using data obtained from soil moisture, temperature, and light intensity sensors to calculate the optimal parameters for watering cultivated soil, by using a cloud computing environment to monitor, analyze, and store data, and control data. On irrigation processes [7].

The aim of this project is to guide farmers and urge them to use effective methods of crop production and to help them in times of drought due to lack of rainfall. The manual strategy of irrigation the crops leaves the farms uncultivated because of shortage in water during dry seasons which comes about in decrease of crop retainability [8].

In this proposed model, the information collected through sensors will play the important roll utilizing these data to gives an estimated amount of required water for crops according to environment situation through collecting data from temperature, soil moisture and light intensity sensors. Moreover, valves have an essential role in providing a path for water to enter and exit the field. The valve opens/closes according to the instructions sent by the client.leads to large amounts of water can be spared and over-watering of plants can be avoided utilizing this technique [9]. t

The information collected from the sensors, such as soil moisture, temperature, and lighting intensity which deployed around the agricultural crops, will be sent to the nearest gateway. The data is then transferred to the cloud through the portal, where it'll store the recorded data sequentially after linking it to the database system. Then the threshold value is decided agreeing to the normal climate condition of the surrounding area, and its value is set during installation. The user will be alerted by the system through sending a message or notification, when the soil moisture threshold has been exceeded by sensed value, then then desired action done by the user according to the situation of land. Ones the owner performs the required action it instantly sends response to the database which in turn provides request to the controller and the motor turned on and the valves gets opened up outlet of water into the fields until the moisture level reach the threshold value. Thus, the surplus outlet water can be stored to be used during dry season. Hence, gathering of water can be done with effective usage of water. The database can be also useful in giving accurate agriculture information. A detailed study of the sensor values can also predict the yields percent. This can also prove beneficial in suggest the farmer to put extra manure or fertilizer for meeting its demand.

II. Materials and Method

The Internet of Things is the physical connection between different devices over the Internet, where data is collected, exchanged, and controlled. While cloud computing provides on-demand online access to computer resources and services. This system complies by collecting data from sensors in real-time with the aim of automating irrigation by comparing threshold values. Obtained information from the sensors can be stored the Thing Speak platform, as this platform provides sufficient space to store, monitor and analyze the data and make decisions based on the information received from the sensors and the impact of each value on decision-making after performing the calculations to obtain the optimal result through which the appropriate action is carried out.

A. Proposed System Architecture and Description

The proposed system enables the farmer to irrigate the field from wherever he is, which reduces his effort. different soil parameters coming from the soil moisture sensors will be sent to the Arduino Uno. As the soil values change, they will be compared with the threshold values via the Arduino, and commands will be sent to the relay to start or stop the motor to begin the watering process for the crops. All information will be uploaded to the cloud platform (Thing Speak) via the Esp8266 Wi-Fi module to be stored and analyzed later to obtain the ideal irrigation method and managing the amount of water needed for each specific period of time. The motor can be controlled via the Android application by the user, and different soil values are obtained through the cloud platform, which will help farmers in the irrigation process. Figure (1) Overview of the proposed system architecture.

Farmers can access the data from anywhere, and can choose a different threshold value for irrigation based on seasons and crops. They can also plan and schedule the optimal utilizing of water resources. With out mechanisms to track amount of water in the soil, farmers must monitor plants and examine the soil manually, and this is a cumbersome process and takes a lot of time and effort. This can be done by means of a smart system that alerts the farmer when the water level drops below the limits set by the farmer [10].

For measure the moisture level in the soil, a soil moisture sensor is used for measuring. After that the data is sent to IOT Gateway. Then this information will be uploaded to the cloud platform by IOT gateway using ESP8266 Wi-Fi Module. The cloud platform in the proposed system includes a web server, a database, and decision logic. The data coming from the Internet of Things portal is stored in the database. In addition, the decision on the necessity of watering plants is made through decision logic. For example, in the developed system, the temperature threshold is maintained at 25°C. the database will trigger decision logic when the temperature exceeded the threshold. It will then send a notification to the farmer via the application on the phone. Depending on the farmer's action, whether to turn on/off the irrigation, a signal will be sent to the cloud and from the cloud to the gateway, which will then send a signal to turn on the relay and turn on the water pump [11].

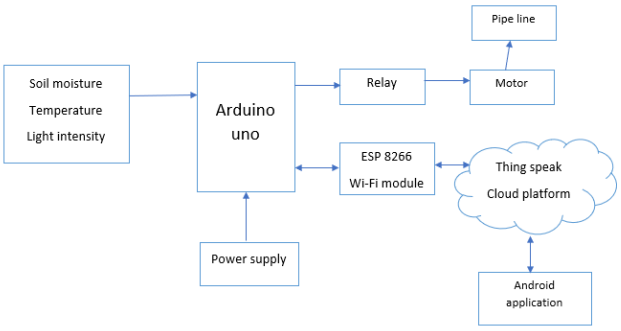


Fig. 1. Proposed system architecture

An IoT-based smart irrigation system uses real-time data to make irrigation decisions. The farmer needs his own data, such as a user name and a special password, to enter the application.

Then he is allowed to choose the irrigation method and choose the crop for that season. The proposed system is implemented in three parts: 1- Sensing 2- Processing 3- Distributing information

Measurement of physical variables is done in the sensing stage, as this stage includes measuring temperature, soil moisture, and lighting intensity. The sensors are collected on the Arduino Uno microcontroller board. This panel acts as a gateway to send data to the cloud [12]. After sending this data by the ESP8266 Wi-Fi module, the data is processed in the cloud. The sensor data is saved in the database in the cloud, in addition to that decisions are made in the decision logic unit based on the sensed data. In the final stage in terms of information distribution, the outputs of the decision logic are sent from the cloud to the Android application and then to the Internet of Things portal [13]. Below is the overall algorithm for the smart irrigation system. Figure (2) shows the complete flow chart of the proposed algorithm.

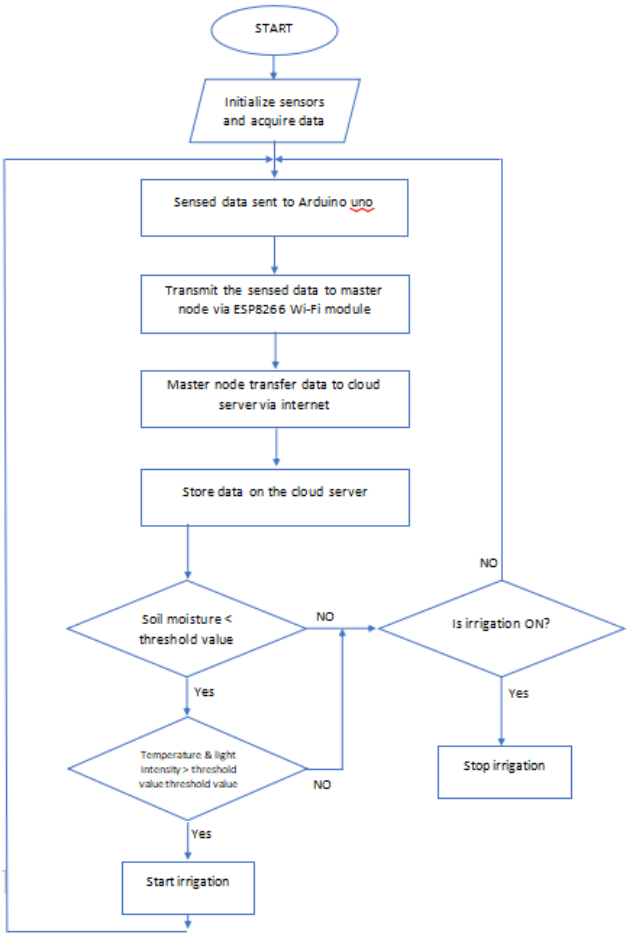


Fig.2. Flow chart of proposed irrigation algorithm

B. System design and implementation

With the system designed and implemented in the study, water resources and irrigation are used efficiently, so that the parameters of the area can be monitored at any time and place via a computer or smart device with internet access. To make an irrigation decision, the soil moisture value has crucial importance. Although rain is an influencing factor for humidity parameters, but it cannot immediately reflect its effect on the parameter values. for this reason, the system must be monitored and intervened at any time. In such systems, the threshold value is specified once and it is defined in the system.

The used sensors in this design are composed mainly of three types (Soil moisture, DHT22 and LDR).

The DHT22 temperature and humidity sensor is an advanced sensor that outputs a calibrated digital signal. It is highly reliable and stable in long-term studies. It Contains 8-bit microprocessor and responds quickly. HL-69 contains sensors for measuring humidity. The sensor is also factory calibrated and hence easy to interface with other microcontrollers. With an accuracy of ±1°C and ±1%, the sensor can measure temperature from 0°C to 70°C and humidity from 10% to 90%. Thus, this sensor can be the best option to measure in this range.[15]

The MH soil moisture sensors are made to calculate the volumetric water content of the soil by using the soil's dielectric constant, also known as its bulk permittivity. You might think of the dielectric constant as the electrical transmission capacity of the soil. As soil's water content rises, so does the soil's dielectric constant. The reason for this reaction is that water has a far higher dielectric constant than the other constituents of soil, including air. Consequently, a reliable estimate of the water content can be obtained by measuring the dielectric constant. When these probes are immersed in soil or liquid, resistance occurs, creating a potential difference between the probe terminals. Depending on the size of this potential difference, the amount of moisture is measured.

LDR (light dependant resistor) is a unique kind of resistor that operates on the principle of photoconductivity, which asserts that resistance varies with light intensity. The more intense the light, the lower its resistance becomes. It is frequently utilized as an automatic public light, a brightness meter, a light sensor, and in other places where light sensitivity is required. Another name for LDR is a light sensor. Typically, LDR are offered in 5 millimeters 8millimeters, 12 millimeters and 25 millimeters sizes

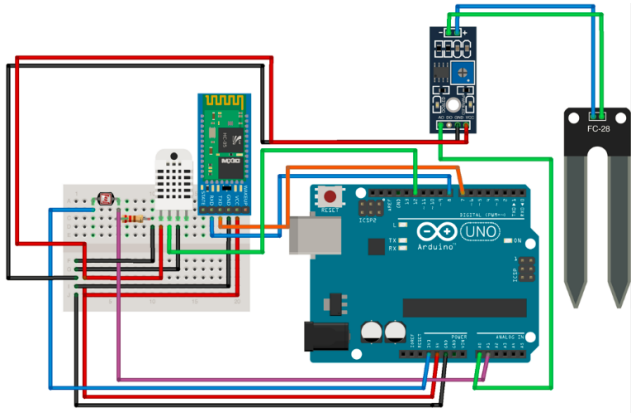


Fig. 3. Wiring diagram of control components

However, in the system designed in the study, the threshold value of the Arduino Uno is not specified in the controller and is used in the cloud. It is determined by Thing Speak, which creates the system. According to this specified limit value in the channel, all incoming data is scanned, and if it is below a

certain threshold value, a Thing HTTP request is triggered by the React application and sent by the microcontroller card. A command is added to the Talkback controlled queue for execution. According to the command read by the talk back queue, the motor is turned on/off and irrigation management is performed. In a designed system, a partial control system that contains hardware components the IoT service that stores and visualizes it as it appears on the Internet is called a cloud system [10]. The control system, including wiring diagram in figure(3) and the hardware components of the system, is shown in Figure (4).

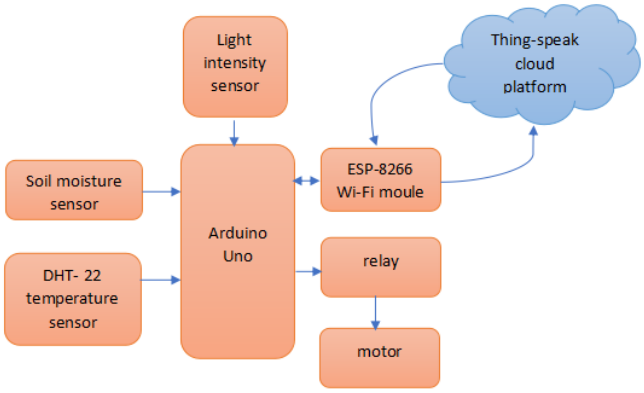


Fig. 4. Control components

Environmental variables that will affect The sensors used for measure the soil moisture coefficient, which has important value in making irrigation decisions. Sensor data is read by the microcontroller. This data is then serially sent to the cloud through the ESP8266 Wi-Fi module connected to the microcontroller. Figure (5) shows the recorded sensor values, which were visualized, and the command queue created in the cloud that was read by the microcontroller.

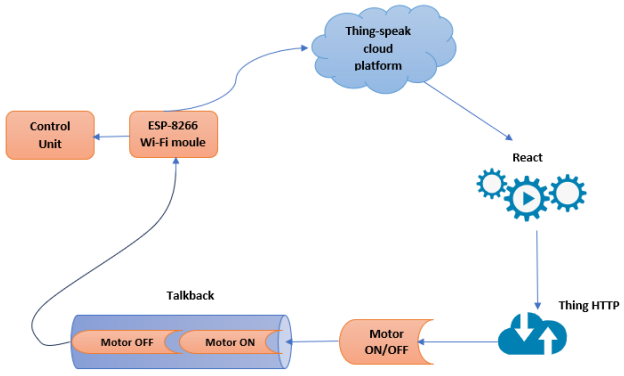


Fig. 5. Cloud System Components.

After connecting the control components and ensure remote monitoring of the system and to store the sensor data obtained from the field in the cloud, then registration is done in Thing Speak, which is an IoT service. A channel is opened here to save data [14]. The channel structure in Thing-Speak is similar to the table structure in a database. The areas where we will record on the channel and their names are specified. An image of the Thing-Speak channel used in this study is shown In Figure (6).

Each time the data reaches the Thing-Speak channel area, the react application performs the comparison operation with the specified threshold value in the cloud. In the React application The pattern of incoming data is tested in the form of frequency, condition type (numeric value, state, geographic location), conditional control and action to be taken when the specified condition is met. In addition, the Activate channel option will only be set the first time or every time the specified condition is met. React application that defines the soil moisture threshold value, test frequency, condition type, and application execution option It has been given in Figure (7).

Thing-HTTP requests (motor On and motor Off) appear in the React application If triggered on, it writes a command string to the Talkback queue on the channel. The Thing-HTTP request body consists of the key string and commands specifically defined for the talkback application. The Talkback queue has a FIFO (First in First Out) structure. TCP communicates with the ESP8266 module through the +IPD command the response queue is suitable for reading and interpretation by the microcontroller. The first read command from the queue is executed by giving a parameter to the function written on the microcontroller side. On the side of the device Irrigation system management is provided.

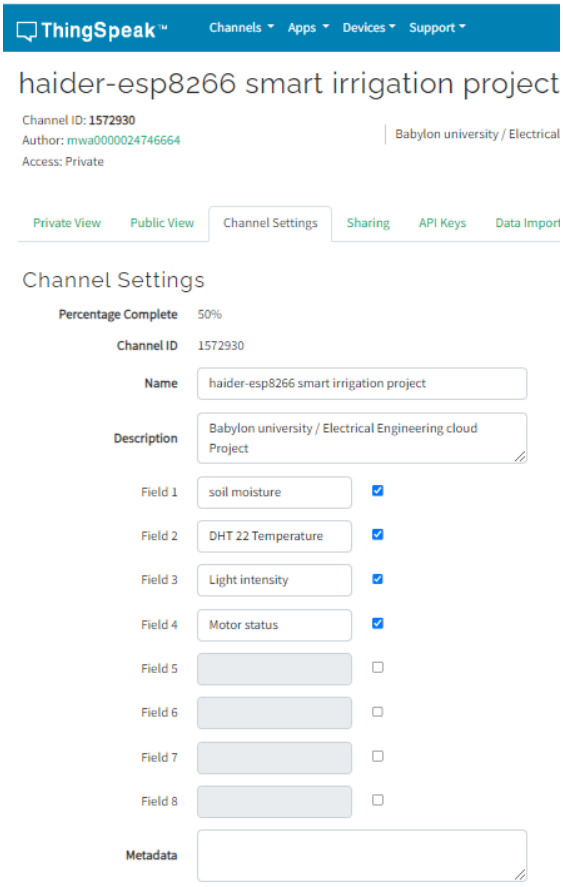


Fig. 6. Thing-Speak channel setting

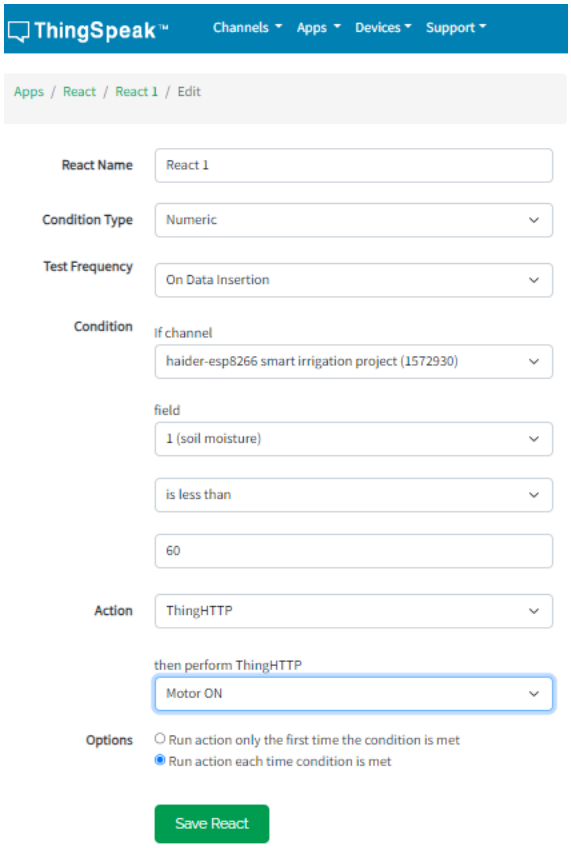
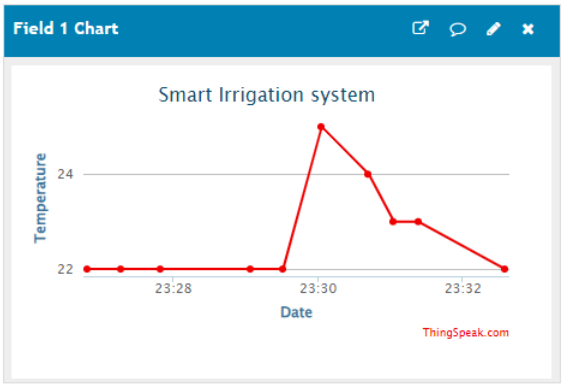


Fig. 7. React Application

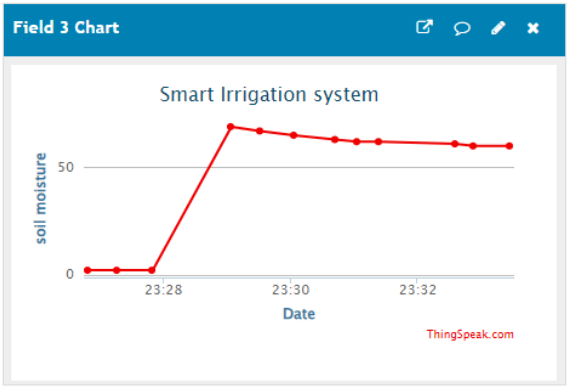
III.RESULTS

As aresult , Soil moisture is a critical parameter for developing a smart irrigation system. By enabling the soil moisture threshold value to be changed, unnecessary water usage was switched online. Soil moisture value, which is an important parameter in irrigation, is measured with sensors that monitor environmental conditions. It moves inversely proportional to the light value and temperature value, and directly proportional to the rain sensor. It has been observed that there is a tendency for change. Created based on data collected in the future determining the irrigation requirements for the pilot area with the model and growing it under current conditions It is planned to estimate the products that may be deemed appropriate. In addition, soil moisture status In case of wireless detection, the energy requirement Alternative sources will be used to meet the needs.

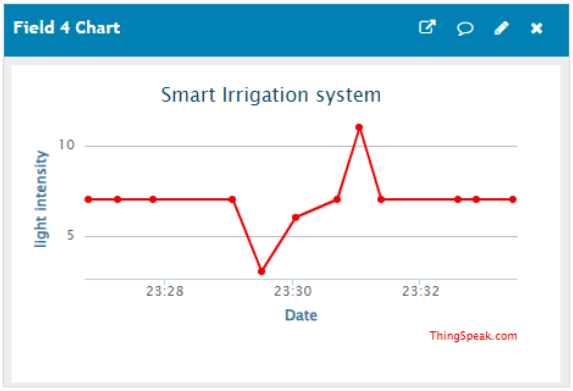
After the obtaining of the results from the cloud platform, the following results were shown below in figures. The fig (8) shows the results of temperature information changing over time, beside that in fig (9) the results of soil moisture variations, which is the fundamental data for understanding the state of the soil that the system depends on to triggered the actuators, in addition fig (10) showing the outcomes of the light intensity change in the area where the sensor is situated. Using the cloud platform's integrated apps, including React, ThingHTTP, and Talkback, a decision is made based on the information gathered.



Fig(8)



Fig(9)



Fig(10)

IV.DISCUSSION

As presented in the results that any change in the level of soil moisture such that the soil moisture value is less than what is required for a particular crop requires the operation of the irrigation system, which operates based on the commands coming from the cloud to the Arduino control unit through wireless communication, where the system works by giving the signal to the relay to operate the irrigation pump. The effect of information coming from temperature sensors and lighting intensity has an additional role in determining the irrigation period each time the pump is operated, depending on the weather condition in terms of the intensity of sunshine or in the event of rain, high air humidity, and low temperature.

V. CONCLUSION

Within the framework of preserving water sources and reducing the inappropriate use of irrigation on farms. We conclude from this study that by using technological means we can conserve water by reducing the amount of water wasted using irrigation, as farmers can monitor farm data on the cloud platform through available Internet applications and mobile devices. In addition, accurate statistics can be generated from data stored for long periods to calculate the least amount of water needed during the agricultural period, which enables farmers to challenge the obstacles of water scarcity during summer and drought periods.

In addition, it possible to conclude that by using this digital and technological structure we can successfully manage water resources by implementing precise and sustainable irrigation. Automating irrigation and reducing the amount of water thus saves the electrical energy consumed as well. Which requires optimal planning for irrigation in a mandatory manner. This will contribute effectively to increasing cultivated lands and rationing water consumption in times of scarcity and summer.

REFERENCES

[1] García, I.F., Montesinos, P., Poyato, E.C., and Díaz, J.R.,, "Optimal design of pressurized irrigation networks to minimize the operational cost under different management scenarios," *Water resources management,* p. 31(6), 1995.

[2] G. M. P. P. M. a. L. D. Cáceres, "Smart Farm Irrigation Model Predictive Control for Economic Optimal Irrigation," *Agriculture, Agronomy,* vol. 11, no. 9, p. 1810, 2021.

[3] J. O.-M. J. G. J. a. C. J. Domínguez-Niño, "Differential irrigation scheduling by an automated algorithm of water balance tuned by capacitance-type soil moisture sensors," *Agricultural Water Management,* vol. 228, no. 105880, 2020.

[4] A. H. N. a. R. S. McCarthy, "Advanced process control of irrigation," *the current state and an analysis to aid future development, Irrigation Science,* vol. 31(3), pp. 183 - 192, 2013.

[5] J. M. M. M. J. a. G. J. Casadesús, "A general algorithm for automated scheduling of drip irrigation in tree crops,," *Computers and Electronics in Agriculture,* vol. 83, pp. 11-20, 2012.

[6] R. P. E. a. D. J. Perea, "Forecasting of applied irrigation depths at farm level for energy tariff periods using Coactive neuro-genetic fuzzy system," *Agricultural Water Management,* no. 107068., p. 256, 2021.

[7] A. R. a. J. P. M. Bhattacharya, "Smart Irrigation System Using Internet of Things," *Applications of Internet of Things,* pp. 119-129, 2021.

[8] A. S. Chandra Prakash Meher, "IoT based Irrigation and Water Logging monitoring system using Arduino and Cloud Computing," in *International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN),* 2019.

[9] S. &. P. B. &. P. S. &. M. D. &. N. D. &. D. B. Laha, "An IOT-Based Soil Moisture Management System for Precision Agriculture Real-Time Monitoring and Automated Irrigation Control," in *International Conference on Smart Electronics and Communication (ICOSEC-2023),* 2023.

[10] G. D. M. G. Y. L. A. S. Ahmed Z, "An Overview of Smart Irrigation Management for Improving Water Productivity under Climate Change in Drylands," *Agronomy,* vol. (8), no. 2113, p. 13, 2023.

[11] S. S. G. Sushanth, "IOT Based Smart Agriculture System," in *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET),* 2018.

[12] K. Y. B. A. A. A. M. N. T. Y. C. M. M. J. H. &. R. M. Obaideen, "An overview of smart irrigation systems using IoT.," *Energy Nexus,* no. 100124, p. 7, 2021.

[13] V. &. H. S. &. S. G. Reddy, "IoT and Cloud Based Sustainable Smart Irrigation System," in *E3S Web of Conferences 472. 10.1051/e3sconf/202447201026.*, 2024.

[14] S. &. A. S. &. I. M. &. A. A. &. A. A. &. A. A. Habib, "Design and Implementation: An IoT-Framework-Based Automated Wastewater Irrigation System," *Sensors,* vol. 23, no. 2358, p. 4, 2023.

[15] Riyam Salah Yaseen and Mariwan Ridha Faris, "Water Scarcity Effect on Kirkuk irrigation project " 2024 *IOP Conf.*

[16] *Ser.: Earth Environ. Sci.* **1374** 012065.
M. E. Seno, A. A. Abed, Y. A. Hamad, U. M. Bhatt, B. Ravindra Babu. and S. Bansal, "Cloud Based Smart Kitchen Automation and Monitoring," 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 1544-1550.

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 133-137
DOI: 10.36287/ijmsit.8.2.13
Received: December 4, 2024; Accepted: December 11, 2024

Analysis of Coronary Heart Diseases by Kinetic Features: Applying Variational Mode Decomposition to ECG Signals and Classification Using Machine Learning Algorithms

Fırat Orhan Bulucu[1], Fatma Latifoğlu[2,3*], Ayşegül Güven[4], Semra İçer[5] and Aigul Zhusupova[6]

[1] *Department of Biomedical Engineering, Faculty of Engineering, Inonu University, Türkiye (firat.orhanbulucu@inonu.edu.tr), (ORCID: 0000-0003-4558-9667)*
[2*] *Neural Information Technologies Inc., Kayseri, Türkiye*
[3] *Department of Biomedical Engineering, Faculty of Engineering, Erciyes University, Türkiye (flatifoglu@erciyes.edu.tr), (ORCID: 0000-0003-2018-9616)*
[4] *Department of Biomedical Engineering, Faculty of Engineering, Erciyes University, Türkiye (aguven@erciyes.edu.tr), (ORCID: 0000-0001-8517-3530)*
[5] *Department of Biomedical Engineering, Faculty of Engineering, Erciyes University, Türkiye (ksemra@erciyes.edu.tr), (ORCID: 0000-0002-3323-9953)*
[6] *Department of Cardiology, Faculty of Medicine, Erciyes University, Türkiye (ag.jusupova@gmail.com), (ORCID: 0009-0003-6002-9171)*

Abstract **–** This study presents an approach for the diagnosis of myocardial infarction (MI) and other coronary heart diseases using 12-lead electrocardiogram (ECG) signals. In the presented approach, 12-lead ECG signals recordings of MI types (STEMI-NSTEMI), other heart diseases (OHD) and healthy control (HC) participants, who presented to the Emergency Department of Erciyes University Hospital for heart disease, were used. In the first stage, the noise-cleaned ECG signals were decomposed into subbands by applying the Variational Mode Decomposition (VMD) method and kinetic features were obtained, and the ones that would positively affect the performance of the classifiers were determined by Chi-square test. In the classification stage, these features were evaluated by Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN) algorithms, and AUC, Accuracy, and Negative Predictive Value ratios were obtained. Classification procedures were performed for HC-OHD, HC-MI (NSTEMI+STEMI), and STEMI-NSTEMI-OHD groups. When evaluated in terms of AUC, rates that can be considered successful (80% and above) were obtained. The findings of this research may contribute to the systems that can be developed for the rapid and accurate diagnosis of coronary heart diseases from ECG signals, which can be difficult to interpret manually.

Keywords *– Coronary heart disease, 12-lead electrocardiogram (ECG) signal, Kinetic features, Variational mode decomposition, Machine learning algorithms.*

Citation: Bulucu et.al., (2024). Analysis of Coronary Heart Diseases by Kinetic Features: Applying Variational Mode Decomposition to ECG Signals and Classification Using Machine Learning Algorithms. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 133-137.

I. INTRODUCTION

Coronary heart disease is the leading cause of death in the world [1]. The diagnostic device used for the diagnosis of such heart diseases is the electrocardiogram (ECG), which is a non-invasive measurement [2, 3]. Fast and accurate diagnosis of Myocardial Infarction (MI), a type of coronary heart disease, is important for the patient's life. MI is divided into ST-elevation MI (STEMI) or non-ST-elevation MI (NSTEMI), which occur in the waves in the ECG signal. These heart diseases can be difficult and complex to analyze with standard 12-lead ECG signals [4]. For this reason, thanks to the analyses made with devices that provide computer-aided automatic diagnosis of 12-lead ECG signals developed with recent research, patient-specific diagnosis of heart diseases is faster and can help cardiologists [3-5]. Artificial intelligence techniques have an important role in the development of these devices.

Most of the studies for predicting heart diseases are focused on obtaining distinctive diagnostic features from ECG signals and classifying them with machine learning algorithms [6]. In these studies, signal processing methods such as Fourier transform and wavelet analysis are usually applied to extract time-dependent or morphological features from ECG signals [6-8]. These features are applied as input to machine learning algorithms such as SVM [9], k-Nearest Neighbors (k-NN) [10], ANN [11], and many other classifiers to predict diseases [12]. Using these machine learning methods, studies are reporting high accuracy rates in the development of decision support systems that can assist healthcare professionals in the diagnosis of MI. These studies generally consist of signal

preprocessing, feature extraction, or selection from signals and classification [11]. Sahu et al. proposed a new technique for MI detection and localization [12]. In their proposed technique, ECG signals are decomposed into components using the variational mode decomposition (VMD) method, and the significant features from these components are selected by the regularized neigh bourhood component analysis method and evaluated in the k-NN classifier. In their study on MI detection, Anwar et al. applied the discrete wavelet transform (DWT) method in the preprocessing step and deep auto encoder method in the feature extraction step to ECG signals and classified them with k-NN algorithm [13]. In another study on MI detection, three different techniques, namely DWT, Empirical Mode Decomposition (EMD), and Discrete Cosine Transform (DCT), were applied to ECG signals, and the obtained features were analysed with k-NN classifier [14]. Similarly, in the study by Zeng et al. on MI prediction [15], a model based on tunable quality wavelet transform (TQWT) and VMD methods was proposed for feature extraction from ECG signals, and classification was performed with neural network-based algorithms. Similar to the previously described research, features are retrieved using methods as Random Forest (RF), Naive Bayes, SVM, Decision Tree, EMD, VMD, wavelet transform, etc. [15, 16]. There are many studies on the classification of the prediction of MI with machine learning algorithms [17-19]. The results obtained in these studies are analyzed in detail in the discussion section.

In this study, 12-lead ECG signals obtained from our recordings of HC, STEMI, NSTEMI, and (OHD) other heart disease (non-NSTEMI and STEMI) groups were analyzed. In the analysis, ECG signals were decomposed into VMD sub-bands, and various kinetic features representing the dynamic and statistical properties of cardiac activity were obtained from each sub-band. From these features, the ones that will positively affect the classifier performance were determined by the Chi-square test (χ^2), and the classification process was performed using machine learning algorithms ANN, SVM, and RF. With the classification process, HC-OHD, STEMI-NSTEMI-OHD, and HC-MI (NSTEMI+STEMI) groups were predicted. The flow diagram summarising the study is given in Fig.1 and detailed explanations of the procedures are given in Section 2. In Section 3 of the study, the findings obtained and the comparison of these findings with the results obtained in similar studies are discussed. In the last section, the results of the study are summarised and future studies are mentioned.

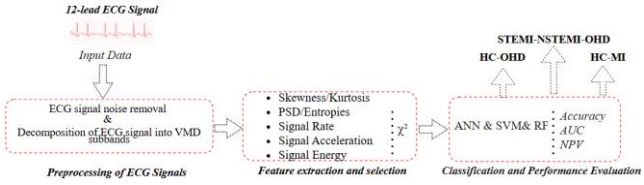


Fig. 1. Flow diagram of the study

II. METHOD

A. Data acquisition

The ECG signals analyzed in this study were taken from people who came to Erciyes University Hospital Emergency Department with chest pain between 2018-2023 (Ethics decision no: 2022/536). The recordings consist of 10-second 12-lead ECG signals with a sampling frequency of 500 Hz. The healthy group consisted of participants aged between 18 and 80 years without myocardial infarction or other heart disease. The NSTEMI, STEMI, and OHD groups consisted of participants aged 18-80 years with clinical findings diagnosed by at least two cardiologists. In the study, 159 records were analyzed for each group.

B. Preprocessing of ECG Signals

At this stage of the study, some methods were applied for pre-processing and decomposition of ECG signals. This stage is necessary for better analysis of ECG signals and accuracy of the signal. In the first stage, after the ECG signals are collected, filtering is applied to remove low/high frequency noise and fundamental errors. A low-pass filter with a cut-off frequency of 45 Hz was applied to remove high-frequency noise, and a high-pass filter with a cut-off frequency of 0.5 Hz was applied to correct baseline fluctuations (Fig.2). After filtering, each ECG lead was decomposed into sub-bands using the VMD method (Fig.3). The VMD method proposed by Dragomiretskiy and Zosso decomposes the original input signal into intrinsic mode components of different frequency and amplitudes [20, 21]. Thanks to this method, the input signal can be analyzed more easily and important information in the time and frequency domain is stored [15]. In this study, various kinetic and statistical features representing the dynamic and statistical properties of cardiac activity were extracted from each sub-band of ECG signals decomposed into intrinsic mode components by the VMD method.

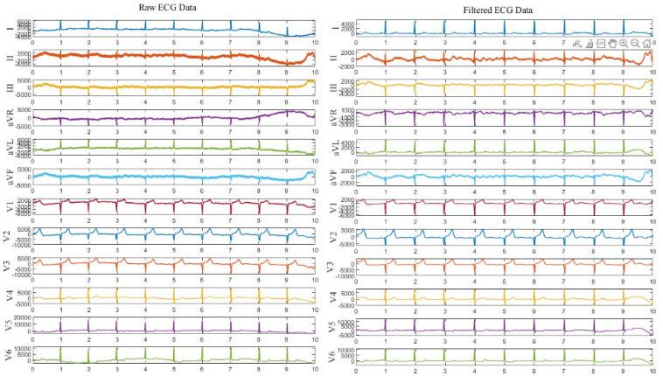


Fig.2. Original and filtered 12-lead ECG signals

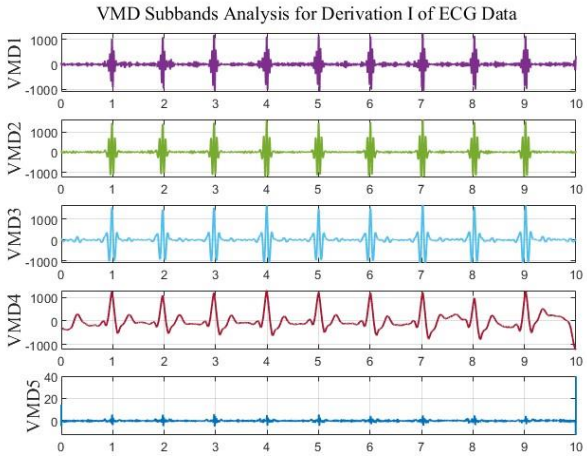


Fig. 3. Sample ECG signal decomposed into subbands by VMD method

C. Feature extraction and selection

Various kinetic and statistical features representing the dynamic properties of cardiac activity were extracted from each sub-band of ECG signals decomposed into sub-bands by applying the VMD method. Kinetic features are parameters

such as the average rate of change of the signal, instantaneous velocity/acceleration, maximum and minimum values for instantaneous velocity/acceleration indicating the highest change points of the signal, average velocity, and average acceleration indicating the overall activity of the signal, and energy of velocity/acceleration signals. These features were used to determine the rate of change, acceleration, and energy density of the signal over time. The statistical features obtained in the study include standard deviation, which measures the amount of variation or spread in the ECG signal, skewness, which shows the asymmetry of the amplitude distribution of the signal, and kurtosis, which measures the degree of tailedness of the amplitude distribution of the signal. In the frequency domain, the power spectral density of each subband in the range of 0.5-45 Hz was calculated. In addition, peak-to-peak amplitude and entropy-based features of each subband were calculated to provide information about the amplitude distribution and complexity of the signal. It has been reported that such features are effective in the analysis of ECG signals [22]. In the feature selection process, the Chi-square test was applied to reduce the size of the dataset and to determine the features that will positively affect the performance of the classifiers [23], and prediction was performed between groups in the classification processes.

III. Experimental Results and Discussion

In this study, 12-lead ECG signals obtained from patients with different coronary heart diseases and healthy groups in the emergency department were analyzed by applying the VMD method and three different machine-learning algorithms. As a result of the analysis, HC-OHD, MI (NSTEMI+STEMI)-OHD, and HC-MI groups were classified and heart diseases were predicted. All operations in the study were performed using MATLAB R2023b program on a computer with Windows 10 operating system, Intel i7 processor, and 16 GB RAM. In the classification process, 5-fold cross-validation was applied and Area Under Curve (AUC), Accuracy (ACC), and Negative Predictive Value (NPV) parameters were obtained. Accuracy is the parameter that measures the accuracy of the classification process [24, 25]. AUC is a parameter that measures the balance between specificity and sensitivity and is a measure of the accuracy of classification [25]. NPV is used in diagnostic test performances and is the probability of the disease occurring [24]. The classification results obtained by applying SVM, ANN, and RF algorithms are given in Table 1.

Table 1. Classification Results

Classifier (%)	HC-MI	HC-OHD	MI-OHD
	AUC-ACC.-NPV	AUC-ACC. NPV	AUC-ACC.-NPV
SVM	81.70-74.76-74.54	80.37-77.11-76.10	70.55-66.45-67.10
ANN	80.81-73.52-75.50	80.92-74.60-73.15	68.79-65.50-66.44
RF	82.22-71.34-70.59	85.46-76.80-74.72	73.73-68.10-67.26

According to the results in Table 1, it is seen that the classification of coronary heart diseases with HC groups is more successful than the classification of heart diseases among themselves. All three machine learning algorithms were close to each other and successful results were obtained in terms of AUC ratio. When the AUC ratios are analysed, the RF algorithm showed the most successful performance. It is normal that the classification results of STEMI-NSTEMI and other heart diseases, which have differences that can be difficult to see in the ECG signal, show slightly lower performance. The algorithm could not fully reveal the differences between these signals. With the techniques and models to be developed in the future, more successful results can be obtained in the classifications of these diseases. In the studies conducted in the literature, mostly MI and HC groups were analyzed and classified with artificial intelligence techniques [16-19, 25, 26]. It is seen that there are fewer studies on the analysis of MI types and other heart diseases [27]. When we look at the multiple classification studies analysing the types of cardiovascular diseases as in this study [28-30], it is seen that the data are unevenly distributed. In the classification results obtained with this situation, an accuracy rate of 90% and above was obtained. The common point of these studies is that they applied deep learning models in the classification phase [27, 28, 30]. However, we do not think that the samples in our study are sufficient for deep learning. In order to train deep learning models, the sample size should be sufficient [31]. For this reason, in this study, the analysis of heart diseases was analysed with machine learning methods. In future studies, the number of samples can be increased and analyzed with deep learning models and we can improve our results. In our results in this study, it is important that our AUC ratios, which provide medically important information [25], are 80% and above in the classification of HC and heart diseases. In the classification of heart diseases among themselves, not-bad AUC ratios were also achieved. We acknowledge that our results show lower performance compared to the results of similar studies [14-16]. However, we also think that our study has contributions and advantages to the literature. The contributions and advantages of this study can be explained as follows:

- It is important that the ECG signals used in the study are not from a ready-made dataset but from our own recordings and that these signals are 12-lead and provide more information for clinical evaluations than single-lead signals [32].
- While most of the studies in the literature focus only on the classification of MI-HC groups [27-30], in this study, in addition to the MI-HC group, heart diseases were classified and analysed among themselves and with HC groups.
- Machine learning algorithms SVM, ANN and RF were used for classification and the performance of different models were compared and the algorithm with the best performance was determined.
- Another point that we consider important is the application of the VMD method, which captures the signal information at a significant rate, in such a study with original data. Because the VMD method provides easier analysis of the ECG signal and can reveal the modes containing the dominant energy in the cardiac vector signal in ECG [15].

This research can be considered as a preliminary study. For this reason, we think that it has some shortcomings. The deficiencies and disadvantages of the research can be explained as follows:

- Examining only the VMD method in the study can be seen as a deficiency. In future studies, in

addition to the VMD method, different techniques and features with similar characteristics can be tested and their performances can be compared.

- In addition to the data consisting of our records, the results of the study can be strengthened by testing the method applied with a similar data set.

IV. CONCLUSION

This research has presented an effective approach to diagnose MI and other cardiac disorders from coronary heart diseases using 12-lead ECG signals. Diagnosis of coronary heart disease by manually analyzing ECG signals can be complex and difficult. For this reason, there is a need to develop systems that can provide effective and fast diagnosis. In this study, which is not only analyzed as HC-heart diseases but also analyzed within heart diseases, we think that the findings obtained can contribute to the systems that can be developed. With future research, the deficiencies of the study can be improved and models can be developed by obtaining more successful performance parameters.

ACKNOWLEDGMENT

This study was supported by the Presidency of TÜRKİYE Health Institutes (TÜSEB) with project number: 20116, TURKIYE. We thank TÜSEB for their support.

Authors' Contributions

Methodology and analysis, F.L.; validation, F.L., F.O.B, A.G, and S.İ.; writing, original draft preparation, F.O.B; review and editing, F.L, F.O.B, A.G, S.İ. and A.Z.; data collection, A.Z.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

Ethical approval for the conduct of the study was obtained from Erciyes University Clinical Research Ethics Committee (decision no: 2022/536).

REFERENCES

[1] World Health Organization. (2023). World Health Statistics 2023 Monitoring health for the SDGs Sustainable Development Goals HEALTH FOR ALL.

[2] Ansari, S., Farzaneh, N., Duda, M., Horan, K., Andersson, H. B., Goldberger, Z. D., ... & Najarian, K. (2017). A review of automated methods for detection of myocardial ischemia and infarction using electrocardiogram and electronic health records. IEEE reviews in biomedical engineering, 10, 264-298.

[3] Ribeiro, A. H., Ribeiro, M. H., Paixão, G. M., Oliveira, D. M., Gomes, P. R., Canazart, J. A., ... & Ribeiro, A. L. P. (2020). Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature communications, 11(1), 1760.

[4] Chauhan, C., Tripathy, R. K., & Agrawal, M. (2024). Third-order tensor-based cardiac disease detection from 12-lead ECG signals using deep convolutional neural network. In Signal Processing Driven Machine Learning Techniques for Cardiovascular Data Processing (pp. 19-34). Academic Press.

[5] Schläpfer, J., & Wellens, H. J. (2017). Computer-interpreted electrocardiograms: benefits and limitations. Journal of the American College of Cardiology, 70(9), 1183-1192.

[6] Sun, Q., Wang, L., Li, J., Liang, C., Yang, J., Chen, Y., & Wang, C. (2024). Multi-phase ECG dynamic features for detecting myocardial ischemia and identifying its etiology using deterministic learning. Biomedical Signal Processing and Control, 88, 105498.

[7] Sadhukhan, D., Pal, S., & Mitra, M. (2018). Automated identification of myocardial infarction using harmonic phase distribution pattern of ECG data. IEEE Transactions on Instrumentation and Measurement, 67(10), 2303-2313.

[8] Zhang, J., Liu, M., Xiong, P., Du, H., Zhang, H., Lin, F., ... & Liu, X. (2021). A multi-dimensional association information analysis approach to automated detection and localization of myocardial infarction. Engineering Applications of Artificial Intelligence, 97, 104092.

[9] Dohare, A. K., Kumar, V., & Kumar, R. (2018). Detection of myocardial infarction in 12 lead ECG using support vector machine. Applied Soft Computing, 64, 138-147.

[10] Arif, M., Malagore, I. A., & Afsar, F. A. (2012). Detection and localization of myocardial infarction using k-nearest neighbor classifier. Journal of medical systems, 36, 279-289.

[11] Muminov, B., Nasimov, R., Mirzahalilov, S., Sayfullaeva, N., & Gadoyboyeva, N. (2020, May). Localization and classification of myocardial infarction based on artificial neural network. In 2020 Information Communication Technologies Conference (ICTC) (pp. 245-249). IEEE.

[12] Sahu, G., & Ray, K. C. (2021). An efficient method for detection and localization of myocardial infarction. IEEE Transactions on Instrumentation and Measurement, 71, 1-12.

[13] Anwar, S. M. S., Pal, D., Mukhopadhyay, S., & Gupta, R. (2024). A Lightweight Method of Myocardial Infarction Detection and Localization from Single Lead ECG Features Using Machine Learning Approach. IEEE Sensors Letters.

[14] Acharya, U. R., Fujita, H., Adam, M., Lih, O. S., Sudarshan, V. K., Hong, T. J., ... & San, T. R. (2017). Automated characterization and classification of coronary artery disease and myocardial infarction by decomposition of ECG signals: A comparative study. Information Sciences, 377, 17-29.

[15] Zeng, W., Yuan, J., Yuan, C., Wang, Q., Liu, F., & Wang, Y. (2020). Classification of myocardial infarction based on hybrid feature extraction and artificial intelligence tools by adopting tunable-Q wavelet transform (TQWT), variational mode decomposition (VMD) and neural networks. Artificial Intelligence in Medicine, 106, 101848.

[16] Sharma, L. D., & Sunkaria, R. K. (2018). Inferior myocardial infarction detection using stationary wavelet transform and machine learning approach. Signal, Image and Video Processing, 12(2), 199-206.

[17] Chakraborty, A., Chatterjee, S., Majumder, K., Shaw, R. N., & Ghosh, A. (2022). A comparative study of myocardial infarction detection from ECG data using machine learning. In Advanced Computing and Intelligent Technologies: Proceedings of ICACIT 2021 (pp. 257-267). Springer Singapore.

[18] Satty, A., Salih, M. M., Hassaballa, A. A., Gumma, E. A., Abdallah, A., & Khamis, G. S. M. (2024). Comparative Analysis of Machine Learning Algorithms for Investigating Myocardial Infarction Complications. Engineering, Technology & Applied Science Research, 14(1), 12775-12779.

[19] Maindarkar, P., & Reka, S. S. (2022, April). Machine Learning-Based Approach for Myocardial Infarction. In International Conference on Artificial Intelligence and Sustainable Engineering: Select Proceedings of AISE 2020, Volume 1 (pp. 17-27). Singapore: Springer Nature Singapore.

[20] Dragomiretskiy, K., & Zosso, D. (2013). Variational mode decomposition. IEEE transactions on signal processing, 62(3), 531-544.

[21] Maji, U., & Pal, S. (2016, September). Empirical mode decomposition vs. variational mode decomposition on ECG signal processing: A comparative study. In 2016 international conference on advances in computing, communications and informatics (ICACCI) (pp. 1129-1134). IEEE.

[22] Xie, L., Li, Z., Zhou, Y., He, Y., & Zhu, J. (2020). Computational diagnostic techniques for electrocardiogram signal analysis. Sensors, 20(21), 6318.

[23] Memiş, G., & Sert, M. (2019, April). Classification of Obstructive Sleep Apnea using Multimodal and Sigma-based Feature Representation. In 2019 27th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

[24] Azar, A. T., & El-Said, S. A. (2014). Performance analysis of support vector machines classifiers in breast cancer mammography recognition. Neural Computing and Applications, 24, 1163-1177.

[25] Xiong, P., Lee, S. M. Y., & Chan, G. (2022). Deep learning for detecting and locating myocardial infarction by electrocardiogram: A literature review. Frontiers in cardiovascular medicine, 9, 860032.

[26] Sraitih, M., Jabrane, Y., & Hajjam El Hassani, A. (2022). A robustness evaluation of machine learning algorithms for ECG myocardial infarction detection. Journal of Clinical Medicine, 11(17), 4935.
[27] Latifoğlu, F., Zhusupova, A., İnce, M., Ertürk, N. A., Özdet, B., İçer, S., ... & Kalay, N. (2024). Preliminary Study Based on Myocardial Infarction Classification of 12-Lead Electrocardiography Images with Deep Learning Methods. The European Journal of Research and Development, 4(1), 42-54.
[28] Jahmunah, V., Ng, E. Y. K., San, T. R., & Acharya, U. R. (2021). Automated detection of coronary artery disease, myocardial infarction and congestive heart failure using GaborCNN model with ECG signals. Computers in biology and medicine, 134, 104457.
[29] Acharya, U. R., Fujita, H., Sudarshan, V. K., Oh, S. L., Adam, M., Tan, J. H., ... & Chua, K. C. (2017). Automated characterization of coronary artery disease, myocardial infarction, and congestive heart failure using contourlet and shearlet transforms of electrocardiogram signal. Knowledge-Based Systems, 132, 156-166.
[30] Baloglu, U. B., Talo, M., Yildirim, O., San Tan, R., & Acharya, U. R. (2019). Classification of myocardial infarction with multi-lead ECG signals and deep CNN. Pattern recognition letters, 122, 23-30.
[31] Han, C., & Shi, L. (2020). ML–ResNet: A novel network to detect and locate myocardial infarction using 12 leads ECG. Computer methods and programs in biomedicine, 185, 105138.
[32] Barandas, M., Famiglini, L., Campagner, A., Folgado, D., Simão, R., Cabitza, F., & Gamboa, H. (2024). Evaluation of uncertainty quantification methods in multi-label classification: A case study with automatic diagnosis of electrocardiogram. Information Fusion, 101, 101978.

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 138-143
DOI: 10.36287/ijmsit.8.2.14
Received: December 5, 2024; Accepted: December 16, 2024

CoCrW Alaşımının Yüzey Özelliklerinin İyileştirilmesi için ZIF-8 Sentezi ve Elektroforetik Biriktirme ile Kaplanması

Yakup UZUN[1], Ayşenur ALPTEKİN[2], Şükran Merve TÜZEMEN[3]*, Burak ATİK[4], Yusuf Burak BOZKURT[5], Ayhan ÇELİK[6]

[1]*Atatürk Üniversitesi, Mühendislik Fakültesi, Maküne Mühendisliği Bölümü, Erzurum, Türkiye (yuzun@atauni.edu.tr) (ORCID: 0000-0002-5134-7640)*
[2]*Atatürk Üniversitesi, Mühendislik Fakültesi, Maküne Mühendisliği Bölümü, Erzurum, Türkiye (aysenr.alptekn98@gmail.com) (ORCID: 0009-0007-1362-5711)*
[3*]*Atatürk Üniversitesi, Mühendislik Fakültesi, Maküne Mühendisliği Bölümü, Erzurum, Türkiye (sukrantuzemen@atauni.edu.tr) (ORCID: 0000-0003-0400-5602)*
[4]*Atatürk Üniversitesi, Mühendislik Fakültesi, Maküne Mühendisliği Bölümü, Erzurum, Türkiye (burak.atik@atauni.edu.tr) (ORCID: 0000-0003-2117-9284)*
[5]*Atatürk Üniversitesi, Mühendislik Fakültesi, Maküne Mühendisliği Bölümü, Erzurum, Türkiye (yusufbozkurt@atauni.edu.tr) (ORCID: 0000-0003-3859-9322)*
[6]*Atatürk Üniversitesi, Mühendislik Fakültesi, Maküne Mühendisliği Bölümü, Erzurum, Türkiye (ayhcelik@atauni.edu.tr) (ORCID: 0000-0002-8096-0794)*

Türkçe Özet – Aşınma ve yüksek korozyon direncinin yanı sıra biyouyumluluk açısından da mükemmel özellikler sergileyen kobalt bazlı alaşımlar, diğer metalik implant malzemeleri arasında öne çıkmaktadır. Ancak döküm, dövme ve talaşlı imalat gibi geleneksel üretim yöntemleri, CoCr alaşımları gibi işlenmesi zor metalik özel parçaların üretimi için birçok endüstriyel uygulamada yetersiz kalmaktadır. Bu nedenle gelişen teknoloji ile birlikte toz metalurjisi ve eklemeli imalat gibi modern üretim yöntemlerine olan ilgi artmıştır. Lazer toz yatağı füzyonu (L-PBF) ile seçici lazer eritme (SLM), geleneksel imalat yöntemlerinin destekleyemediği karmaşık iç yapılara sahip malzemelerin üretimi için en çok tercih edilen eklemeli imalat yöntemlerinden biridir. Ayrıca diğer tüm metalik malzemeler gibi CoCr alaşımları da vücut içerisinde biyouyumluluk açısından inert özellikler sergilediğinden, bu alaşımlardan üretilen implantlarda osseointegrasyon, biyoaktivite, antibakteriyellik gibi davranışların olmaması implantın başarısını olumsuz yönde etkilemektedir. Bu açıdan bu malzemelerin yüzeylerine çeşitli yüzey işlemleri uygulanarak yüzey özellikleri iyileştirilebilmektedir. Bu anlamda, altıgen geometrisi ve gözenekli yapısıyla hidrofobik karakter sergileyen ZIF-8 metal organik çerçeve (Zn-Zeolitik imidazolat çerçeve) kaplama malzemesi, antibakteriyel ve biyouyumluluğunun yanı sıra kimyasal ve termal kararlılığıyla da dikkat çekmektedir. Bu çalışmada, ZIF-8 kaplama malzemesi ilk olarak $Zn(NO_3)_2$ ve 2-metilimidazol ile 50 °C'de metanol çözücüsü kullanılarak sentezlenmiştir. Ayrıca SLM yöntemi ile 10 x 10 x 2 mm^3 boyutlarında CoCrW alaşım numuneleri üretilmiştir. Sentezlenen ZIF-8 malzemesi elektroforetik biriktirme (EPD) yöntemi ile CoCrW numunelerinin yüzeyine kaplanmıştır. Sentez ve kaplama işlemlerinin ardından, hem ZIF-8 malzemesinin başarılı sentezini hem de ZIF-8 malzemesinin EPD yöntemiyle CoCrW numune yüzeylerine başarılı bir şekilde kaplandığını doğrulamak için yapısal (XRD) ve morfolojik (SEM) karakterizasyonlar gerçekleştirilmiştir.

Anahtar Kelimeler *– CoCrW, biyomalzeme, antibakteriyel, ZIF-8, metal organik çerçeveler*

Atıf: Uzun, Y. et al. (2024). CoCrW Alaşımının Yüzey Özelliklerinin İyileştirilmesi için ZIF-8 Sentezi ve Elektroforetik Biriktirme ile Kaplanması. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 138-143.

Synthesis and Coating with Electrophoretic Deposition of ZIF-8 for the Improvement of Surface Properties of CoCrW Alloy

Extended Abstract

Cobalt-based alloys, which exhibit excellent properties in terms of wear and high corrosion resistance as well as biocompatibility, stand out among other metallic implant materials. However, traditional production methods such as casting, forging and machining are insufficient in many industrial applications for the production of difficult-to-machine metallic special parts such as CoCr alloys. Therefore, with the developing technology, interest in modern production methods such as powder metallurgy and additive manufacturing has increased. Selective laser melting (SLM) with laser powder bed fusion (L-PBF) is one of the most preferred additive manufacturing methods for the production of materials with complex internal structures that traditional

manufacturing methods cannot support. In addition, since CoCr alloys, like all other metallic materials, exhibit inert properties in terms of biocompatibility within the body, the lack of behaviors such as osseointegration, bioactivity, antibacteriality in implants produced from these alloys negatively affects the success of the implant. In this respect, surface properties can be improved by applying various surface treatments to the surfaces of these materials. In this sense, ZIF-8 metal organic framework (Zn-Zeolitic imidazolate framework) coating material, which exhibits hydrophobic character with its hexagonal geometry and porous structure, is remarkable for its chemical and thermal stability as well as its antibacterial and biocompatibility. In this study, ZIF-8 coating material was first synthesized with $Zn(NO_3)_2$ and 2-methylimidazole using methanol solvent at 50 °C. In addition, CoCrW alloy samples with dimensions of 10 x 10 x 2 mm3 were produced by SLM method. The synthesized ZIF-8 material was coated on the surface of CoCrW samples by electrophoretic deposition (EPD) method. Following the synthesis and coating processes, structural (XRD) and morphological (SEM) characterizations were performed to confirm both the successful synthesis of ZIF-8 material and the successful coating of ZIF-8 material on CoCrW sample surfaces by EPD method.

Keywords *– CoCrW, biomaterial, antibacterial, ZIF-8, metal organic frameworks*

Citation: Uzun, Y. et al. (2024). Synthesis and Coating with Electrophoretic Deposition of ZIF-8 for the Improvement of Surface Properties of CoCrW Alloy. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 138-143.

I. INTRODUCTION

CoCrW alloys are frequently preferred in biomedical and aerospace applications due to their superior properties such as high temperature, wear and corrosion resistance [1]. In addition to these superior properties, other important reasons why CoCrW alloy is preferred as an implant material in biomedical applications in hip, knee and dental implants are its non-magnetic, radiopaque and MRI compatible properties [2]. However, some studies have shown that the release of Co^{2+}, Cr^{3+} and Cr^{6+} ions into the body from implants made of CoCr alloys may cause inflammatory effect as a result of hypersensitivity at the implant site [3]. In addition, since these alloys are frequently used in in-body implants, they are constantly exposed to corrosion and wear damage. On the other hand, studies show that most of the failures of metallic implant materials during application are due to bacterial invasion. Therefore, in addition to its superior properties, there are many problems to be solved, such as structural, chemical and biological incompatibilities that may lead to the rejection and failure of implants and prostheses made of CoCrW alloy by the body [4]. To overcome all these problems, the structural, chemical and biological surface properties of these structures have been investigated to improve their biocompatibility and antibacterial properties. Metal organic frameworks (MOFs) are three-dimensional porous polymeric materials in which multivalent organic ligands and metal ions come together to form a densely packed, periodic network structure [5]. MOFs are known for their high specific surface areas and porosity, which enable them to effectively store and release substances. Additionally, some MOFs can intelligently respond to pH changes, facilitating the controlled release of corrosion ions and drugs. This unique property makes them particularly promising for applications in antibacterial fields, where targeted delivery and release can enhance therapeutic efficacy [6], [7]. Zinc metal organic framework (ZIF-8), an important MOF, possesses chemical, thermal stability and hydrophobic characteristic and exhibits easy synthesis [8]. The size of ZIF-8 particles can be effectively controlled by adjusting several factors, including the molar ratio of Zn^{2+} to 2-methylimidazole, the choice of reaction solvent, the temperature of the reaction, and various other conditions. This flexibility allows for the production of ZIF-8 particles ranging from dozens of nanometers to hundreds of micrometers in size [9-11].

Furthermore, the rapid synthesis of ZIF-8 nanocrystals in aqueous system at room temperature has been developed, which can easily and environmentally friendly prepare thermally and chemically stable nanoscale ZIF-8. It is necessary to improve it by developing it with good biocompatibility [12]. Over the years, various coating techniques have been developed to enhance surface properties for biomedical applications. These techniques include immersion coating, thermal spraying, plasma phase coatings, layer-by-layer deposition, and electrochemical coatings. Each method offers distinct advantages for improving biocompatibility, corrosion resistance, and other surface characteristics crucial for medical devices and implants [13-15]. Electrophoretic deposition (EPD) offers several advantages, including low equipment cost, the ability to operate at room temperature, fast deposition times, the capability to coat porous or three-dimensional substrates, easy control over coating morphology and thickness, and the flexibility to use both aqueous and organic suspensions for coating substrates [16], [17]. In addition, the EPD method is a versatile coating technique that allows the control of process parameters such as the amount of material deposited on the base material, voltage, time, distance between electrodes or suspension properties such as concentration, pH, stability. With the EPD method, it is possible to coat a wide range of materials, including polymers, bioceramics, metals and their composites, on a suitable base material [16]. In this sense, ZIF-8, a metal organic polymer that has proven its antibacterial and osteogenic properties, has recently attracted attention [17]. Due to its excellent biocompatibility and relatively simple synthesis process, ZIF-8 has emerged as one of the most widely used MOFs in biomedical applications. Additionally, the release of Zn^{2+} ions from ZIF-8 has been shown to possess antimicrobial properties, providing long-lasting antibacterial effects [18], [19].

In a study by Ling et al. on AZ31 Mg alloy, Cu-doped ZIF-8 particles were solvothermally coated on the alloy surface and as a result of the study, they obtained a bioactive surface in terms of antibacterial, osteogenic and corrosion resistance [18]. Wen et al. coated titanium with ZIF-90, a type of Zn imidazolate framework, and as a result of the study, they found

that the coating increased the antibacterial and osteogenic effect on Ti implants in orthopedic fields [19]. In the study by Tao et al., the implant material pure titanium was first anodized and the titanium oxide nanotubes formed on the surface were coated with cobalt MOF EPD method supplemented with OGP material that supports bone formation. The results of the study show that bone formation is significantly improved compared to pure titanium and MOFs will be an important step in the development of drug-assisted implants [20].

In order to increase the surface area of ZIF-8 particles and thus their surface activity, studies on smaller sized ZIF-8 are very valuable in this sense. Venna et al., synthesized ZIF-8 at room temperature using methanol. In this synthesis, they saw that if nano-sized ZIF-8 is to be synthesized, high nucleation rate and low crystallization rate are required [21]. Ordonez et al., summarized the general synthesis methods and basic applications of MOFs in their study and proved that they can be used in many applications. In their study on the synthesis of ZIF-8, it is aimed to reduce the particle size, control the crystal size and produce it in nano size. They used DMF as a solution for the synthesis of ZIF-8 and synthesized ZIF-8 with a BET surface area of 1300 m^2/g with a size of 50-150 nm at 140 °C [22]. In another study, Pan et al. synthesized ZIF-8 using DMF as a solution for 5 min at room temperature. It was reported that the particle size of ZIF-8 decreased as the $C_4H_6N_2/Zn(NO_3)_2$ ratio increased during this synthesis [23].

In this study, ZIF-8 particles of approximately 50-100 nm in size were chemically synthesized at 50 °C and structural and morphological analyses were carried out after synthesis. The synthesized ZIF-8 particles were coated with EPD by applying different potentials of 20 V and 40 V to CoCrW alloy with dimensions of 10 x 10 x 2 mm^3 produced by selective laser melting, an additive manufacturing method. After the coating process, structural and morphological analyses were performed and it was confirmed that the coating was successfully formed.

II. Materials and Method

Within the scope of the study, synthesis of ZIF-8, production of CoCrW samples by selective laser melting (SLM) method and coating of ZIF-8 particles on CoCrW by electrophoretic deposition were carried out. The processes carried out in the sub-headings are explained in detail.

A. ZIF-8 synthesis

The chemicals used for the synthesis of ZIF-8 were methanol, $Zn(NO_3)_2$, 2-methyl imidazole and the equipment were magnetic stirrer, centrifuge and oven. Firstly, 1 g of 2-methyl imidazole was added to 50 ml of methanol, which was previously heated to 50 °C and continuously stirred at 400 rpm in a magnetic stirrer. After obtaining a homogeneous solution, 1.4 g of $Zn(NO_3)_2$ in 25 ml of methanol was added dropwise into this solution [23]. A white solution was obtained and the solution was heated and stirred for 1 hour. After 1 hour, the heating was stopped and the solution was stirred for 6 hours. This white solution was centrifuged at 4000 rpm for 10 minutes and the white precipitate was subjected to dehumidification in an oven at 60 °C overnight.

B. Figures and Tables Fabrication of CoCrW samples by selective laser melting (SLM)

CoCrW specimens with dimensions of 10 x 10 x 2 mm^3 were produced using CoCrW powder according to ASTM F75 standard by selective laser melting, which is an additive manufacturing method that allows the production of specific, complex, difficult-to-machine materials with high melting temperatures, using the parameters given in Table 1. The manufacturings were performed using the CONCEPT LASER MLab Cusing device. Also, CoCrW alloy contain 58.85 wt% Co, 26.30 wt% Cr, 12.62 wt% W, 1.13 wt% Si and 1.1 wt% C [25].

Table 1. Parameters of manufacturing of CoCrW samples by SLM method

Parameters	Values
Laser Power	95 W
Plane scan speed	1500 mm/s
Contour scan speed	1500 mm/s
Layer thickness	30 µm

C. Electrophoretic deposition of ZIF-8

The synthesized ZIF-8 particles were coated on the produced samples by electrophoretic deposition (EPD), an electrochemical coating method.

For the coating process, 1 g/L ZIF-8 colloidal suspension was first prepared. For the suspension, 99 ml of deionized water and 1 ml of acetic acid was used as electrolyte [24], [25]. After stirring the suspension for 1 hour, the coating step was started. The coating process was carried out for 15 minutes by applying 20 V and 40 V voltages to examine the effect of voltage on the coating morphology. Figure 1 shows a schematic representation of the EPD method.

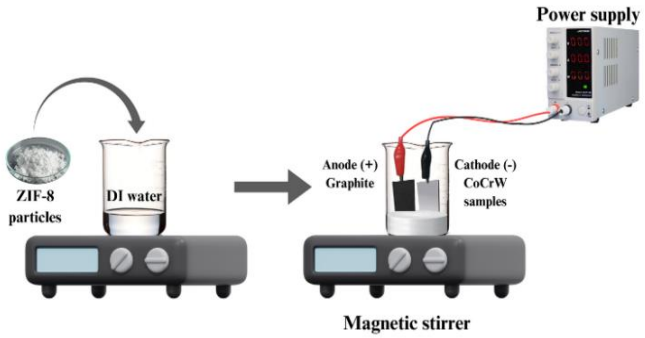


Fig. 1 Schematic representation of ZIF-8 coating by EPD method

After all procedures, structural and morphological analyses were performed.

III. Results

X-ray diffraction (XRD) and scanning electron microscopy (SEM)-energy dispersive X-ray spectroscopy (EDX) analyses were performed for structural and morphological investigations of the synthesized ZIF-8 particles, respectively. The counts-2θ graph obtained as a result of XRD analysis is given in Figure 2. According to the obtained XRD graph, the belrigin peaks confirm that the synthesis of ZIF-8 was successfully realized when compared with the literature [20-23].

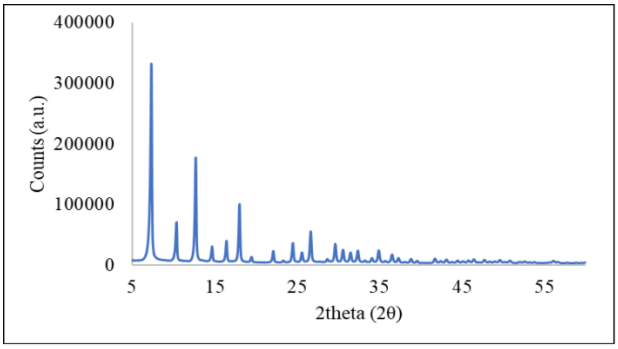


Fig. 2 XRD graph of synthesized ZIF-8 particles

Figure 3 shows the images of (a) SEM and (b) EDX analyses performed to investigate the morphological structure of ZIF-8 particles. The images confirm that porous ZIF-8 particles with a hexagonal crystal structure of approximately 50-100 nm in size were successfully synthesized. In addition, EDX results confirm that ZIF-8 particles contain Zn element densely in the desired direction.

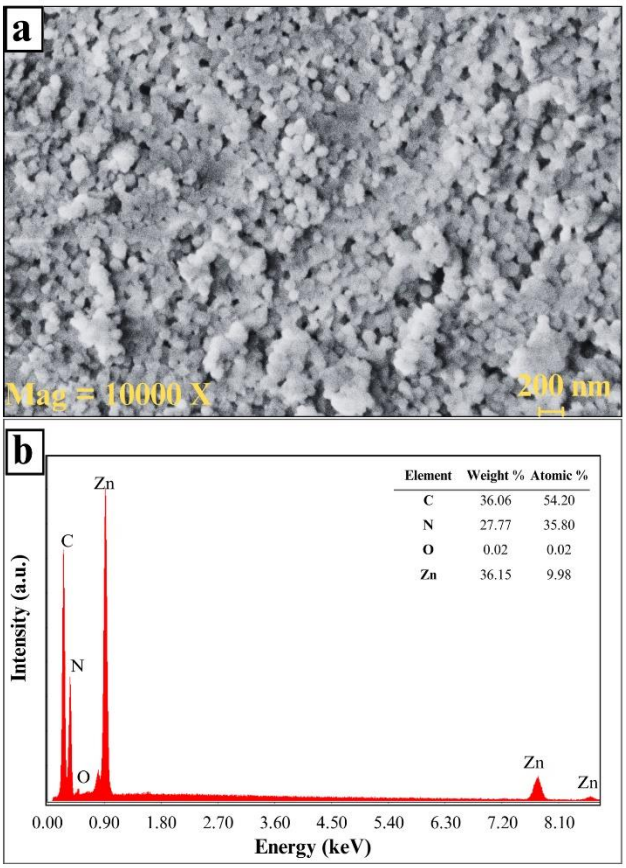


Fig. 3 (a) SEM image and (b) EDX analysis of the synthesized ZIF-8 particles

Figure 4 shows the XRD graph of CoCrW sample produced with SLM and it is confirmed by comparison with the literature that the prominent peaks belong to CoCrW alloy [24].

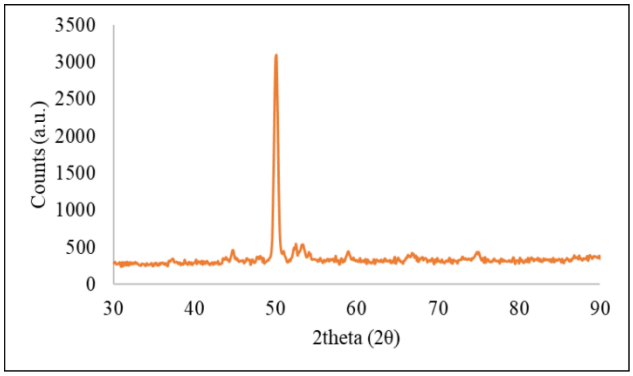


Fig. 4 XRD graph of CoCrW sample

Figure 5 shows SEM images of CoCrW specimens coated with ZIF-8 by EPD method at (a) untreated, (b) 20 V, (c) 40 V for 15 minutes. When the images are examined, it is confirmed that the coatings are realized compared to the untreated samples. When the images are examined, it is seen that the homogeneous deposition of the sample coated for 20 V - 15 minutes is more than the sample coated for 40 V - 15 minutes. This situation is considered to negatively affect the coating of ZIF-8 particles with EPD by applying more voltage.

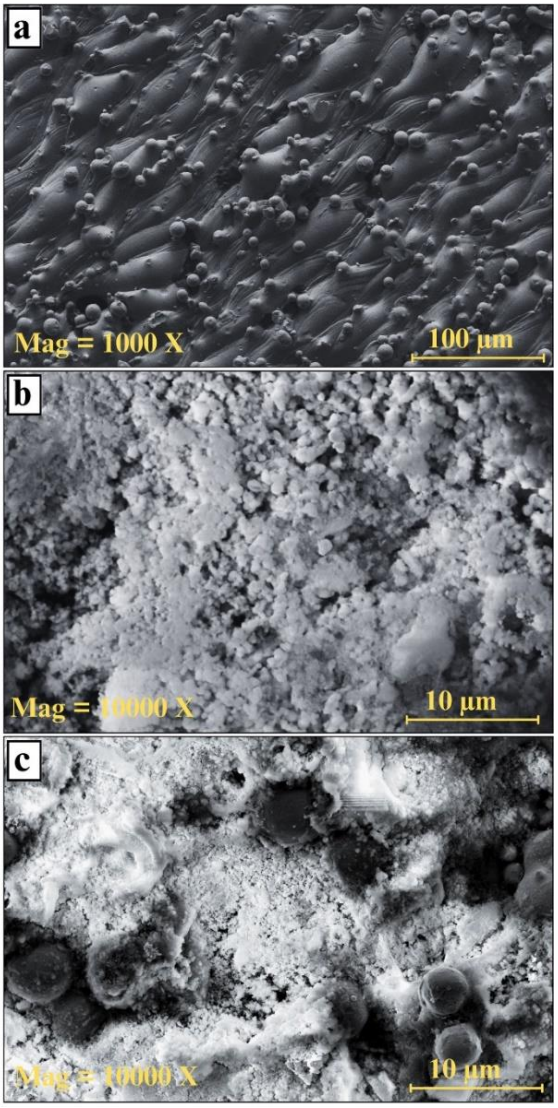


Fig. 5 SEM images of CoCrW samples: (a) untreated, ZIF-8 coated with EPD method applying for (b) 20 V-15 min, and (c) 40 V-15 min

Figure 6 shows SEM analysis and cross-sectional images of (a) 20 V and (b) 40 V coatings to evaluate the coating morphologies in terms of thickness. When the cross-section and surface images were evaluated, it was observed that the coating applied with a voltage of 20 V was both more homogeneous and thicker. Therefore, for optimum results, the coating condition of 20 V - 15 minutes is considered to be more suitable for the desired coating of ZIF-8 particles on CoCrW alloy.

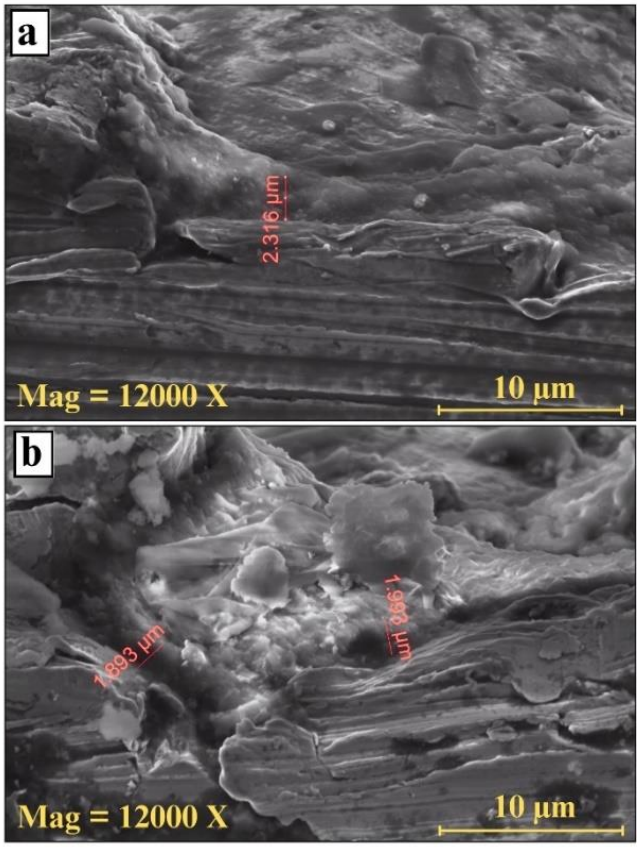


Fig. 6 SEM cross-section images of CoCrW samples: ZIF-8 coated with EPD method applying for (a) 20 V-15 min and (b) 40 V-15 min

Figure 7 shows the results of EDX analyzes of CoCrW samples (a) untreated, ZIF-8 coated with EPD method applying for (b) 20 V-15 min, and (c) 40 V-15 min. Compared to the untreated sample, Co, Cr and W elements as well as Zn, N, C elements due to ZIF-8 coating were found in the 20 V-15 min and 40 V-15 min coated samples. This confirms that the coating was successfully realized. In addition, when the analyzes are examined, the fact that the Zn, C and N ratios in the 20 V-15 min coating are higher than the 40 V-15 min coated sample indicates that the application of 20 V voltage, such as cross-sectional SEM images in terms of coating parameter, will provide a more homogeneous and thicker film.

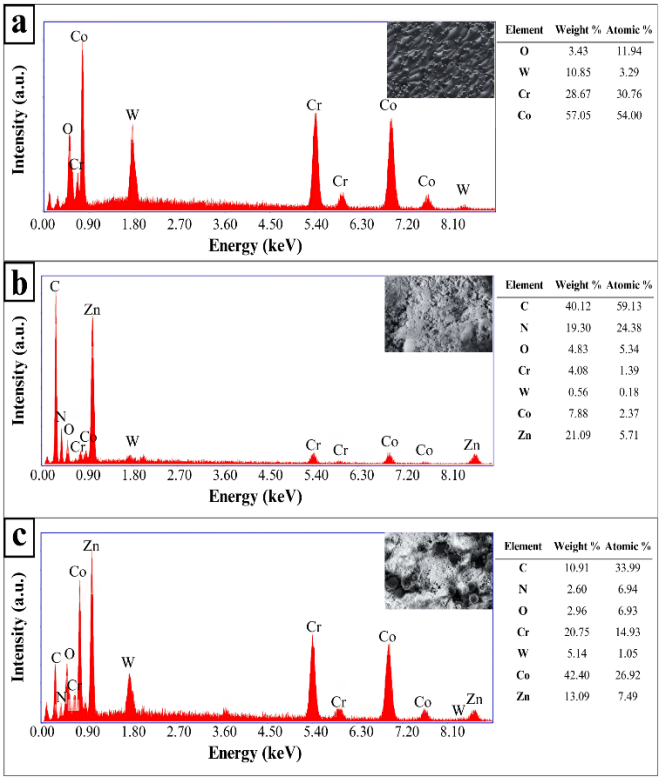


Fig. 7 EDX analyzes of CoCrW samples: (a) untreated, ZIF-8 coated with EPD method applying for (b) 20 V-15 min, and (c) 40 V-15 min

IV. DISCUSSION

When the XRD graph obtained as a result of ZIF-8 synthesis was evaluated, similar results were encountered with the literature. In addition, morphological analysis of ZIF-8 particles synthesized by SEM analysis was performed and compared with the literature. As a result of the study, XRD and SEM analyses confirm that the synthesis of ZIF-8 at 50 °C was successfully synthesized at a smaller grain size than the ZIF-8 particles synthesized at room temperature.

After synthesis, ZIF-8 particles were coated on CoCrW samples produced with SLM by EPD method at different stress parameters. The surface and cross-sectional morphological analysis after the coating process confirms that the coating was successful.

V. CONCLUSION

The results of the study are summarized as follows:

- ✓ ZIF-8 particles with a size of about 50 nm were homogeneously synthesized at 50 °C compared to ZIF-8 particles synthesized at room temperature.
- ✓ CoCrW samples can be easily produced in 10 x 10 x 2 m^3 dimensions with the additive manufacturing method SLM.
- ✓ The synthesized nano-sized ZIF-8 particles were successfully coated on CoCrW samples by electrophoretic deposition method by applying different voltage parameters.
- ✓ 20 V - 15 minutes coating parameter was evaluated as the optimum parameter as a result of SEM analysis.

This study is a preliminary study and will be subjected to corrosion and abrasion tests for in-vitro analysis after the coating parameters are made more suitable. As a result of this

study, the coating of ZIF-8 particles on CoCrW alloy by EPD method was confirmed.

ACKNOWLEDGMENT

This study was presented as an oral presentation at the "International Trend of Tech Symposium (ITTSCONF-2024)" conference.

Authors' Contributions

The authors' contributions to the paper are equal.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics

REFERENCES

[1] I. Milošev, CoCrMo Alloy for Biomedical Applications, Djokić, S. (Eds.), Biomedical Applications, Modern Aspects of Electrochemistry, Boston, MA: Springer, 55, 1-72, 2012.

[2] Q. Chen, G. A., Thouas, "Metallic implant biomaterials", Mater Sci Eng R Rep, 87, pp. 1-57, 2015.

[3] A. W. E. Hodgson, S. Kurz, S. Virtanen, V. Fervel, C. O. A.Olsson, S. Mischler, "Passive and transpassive behaviour of CoCrMo in simulated biological solutions", Electrochim Acta, 49, pp. 2167–2178, 2004.

[4] A. Igual Muñoz, S. Mischler, "Effect of the environment on wear ranking and corrosion of biomedical CoCrMo alloys", J Mater Sci: Mater Med, 22, pp. 437–450, 2011.

[5] C. Chu, M. Su, J. Zhu, D. Li, H. Cheng, X. Chen, G. Liu, "Metal-Organic Framework Nanoparticle-Based Biomineralization: A New Strategy toward Cancer Treatment", Theranostics,18, 9 (11), pp. 3134-3149, 2019.

[6] J. Liu, D. Wu, N. Zhu, Y. Wu, G. Li, "Antibacterial mechanisms and applications of metal-organic frameworks and their derived nanomaterials", Trends in Food Science & Technology, 109, pp. 413-434, 2021.

[7] S. Feng, Q. Tang, Z. Xu, K. Huang, H. Li, Z. Zou, "Development of novel Co-MOF loaded sodium alginate-based packaging films with antimicrobial and ammonia-sensitive functions for shrimp freshness monitoring", Food Hydrocolloids, 135, 108193, 2023.

[8] R. Yang, B. Liu, F. Yu, H. Li, Y. Zhuang, "Superhydrophobic cellulose paper with sustained antibacterial activity prepared by in-situ growth of carvacrol-loaded zinc-based metal organic framework nanorods for food packaging application", International Journal of Biological Macromolecules, 234, 123712, 2023.

[9] J. Matusiak, A. Przekora, W. Franus, "Zeolites and zeolite imidazolate frameworks on a quest to obtain the ideal biomaterial for biomedical applications: A review", Materials Today, 67, pp. 495-517, 2023.

[10] V. Hoseinpour and Z. Shariatinia, "Applications of zeolitic imidazolate framework-8 (ZIF-8) in bone tissue engineering: A review", Tissue and Cell, 72, 101588, 2021.

[11] S. Kouser, A. Hezam, M. J. N. Khadri et al., "A review on zeolite imidazole frameworks: synthesis, properties, and applications", J Porous Mater, 29, pp. 663–681, 2022.

[12] J. Haider, A. Shahzadi, M. U. Akbar, I. Hafeez, I. Shahzadi et al., "A review of synthesis, fabrication, and emerging biomedical applications of metal-organic frameworks", Biomaterials Advances, 140, 213049, 2022.

[13] E. Avcu, F. E. Baştan, H. Z. Abdullah, M. A. U. Rehman, Y. Y. Avcu, A. R. Boccaccini, "Electrophoretic deposition of chitosan-based composite coatings for biomedical applications: A review", Progress in Materials Science, 103, pp. 69-108, 2019.

[14] Z. Hadzhiev and A. R. Boccaccini, "Recent developments in electrophoretic deposition (EPD) of antibacterial coatings for biomedical applications - A review", Current Opinion in Biomedical Engineering, 21, 100367, 2022.

[15] S. Bakhshandeh, S. A. Yavari, "Electrophoretic deposition: a versatile tool against biomaterial associated infections", J. Mater. Chem. B, 6, pp. 1128-1148, 2018.

[16] C. Y. Sun, C. Qin, X. L. Wang, G. S. Yang, K. Z. Shao, Y. Q. Lan, Z. M. Su, P. Huang, C. G. Wang, E. B. Wang, "Zeolitic imidazolate framework-8 as efficient pH-sensitive drug delivery vehicle", Dalton Trans., 41 (23), 6906, 2012.

[17] J. Chen, X. Zhang, C. Huang, H. Cai, S. Hu, Q. Wan, X. Pei, J. Wang, "Osteogenic activity and antibacterial effect of porous titanium modified with metal-organic framework films", J Biomed Mater Res Part A, 105A, pp. 834–846, 2017.

[18] L. Ling, S. Cai, Y. Zuo, M. Tian, T. Meng, H. Tian, X. Bao, G. Xu, "Copper-doped zeolitic imidazolate frameworks-8/hydroxyapatite composite coating endows magnesium alloy with excellent corrosion resistance, antibacterial ability and biocompatibility", Colloids and Surfaces B: Biointerfaces, 219, 112810, 2022.

[19] X. Wen, J. Ma, D. Jiang, J. Ma, "Fabrication of indocyanine green-loaded zeolitic imidazole frameworks-90 coating on titanium implants to enhance antibacterial and osteogenic effects", Materials Letters, 351, 135064, 2023.

[20] B. Tao, W. Yi, X. Qin, J. Wu, K. Li, A. Guo, J. Hao, L. Chen, "Improvement of antibacterial, anti-inflammatory, and osteogenic properties of OGP loaded Co-MOF coating on titanium implants for advanced osseointegration", Journal of Materials Science & Technology, 146, pp. 131-144, 2023.

[21] S. R. Venna and M. A. Carreon, "Highly Permeable Zeolite Imidazolate Framework-8 Membranes for CO2/CH4 Separation", Journal of the American Chemical Society, 132(1), pp. 76-78, 2010.

[22] M. J. C. Ordoñez, K. J. Balkus, J. P. Ferraris, I. H. Musselman, "Molecular sieving realized with ZIF-8/Matrimid® mixed-matrix membranes", Journal of Membrane Science, 361, 1–2, pp. 28-37, 2010.

[23] Y. Pan, Y. Liu, G. Zeng, L. Zhao, Z. Lai, "Rapid synthesis of zeolitic imidazolate framework-8 (ZIF-8) nanocrystals in an aqueous system", Chemical Communications, 47(7), 2071, 2011.

[24] Ş. M. Tüzemen, Y. B. Bozkurt, B. Atik, Y. Uzun, and A. Çelik, "Investigation of the Effect of Bioactive Glass Coating on the Corrosion Behavior of Pre-treated Ti6Al4V Alloy", TJNS, no. 1, pp. 87–91, October 2024.

[25] Ş. M. Tüzemen, Y. B. Bozkurt, B. Atik, Y. Uzun, and A. Çelik, "Electrochemical Impedance Spectroscopy Analysis of 45S5 Bioglass Coating on After Oxidation of CoCrW Alloy", TJNS, no. 1, pp. 82–86, October 2024.

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 144-150
DOI: 10.36287/ijmsit.8.2.15
Received: December 1, 2024; Accepted: December 18, 2024

Potato Leaf Disease Detection Using Faster R-CNN and YOLO Models

Sara Medojević[1*]

[1*]*Faculty of Applied Sciences, University of Donja Gorica , Podgorica, Montenegro (sara.medojevic2@gmail.com) (ORCID: 0009-0002-9995-9416)*

Abstract – Potato is one of the most important food crops globally in terms of total food production, significantly impacting the global economy. Infected potato plants show visible symptoms on their leaves, which drastically simplifies the process of early detection, disease prevention, and minimizing the risk to uninfected plants. Smart farming and new advanced technologies incorporate different tools for real-time monitoring and analysis. Most of the models used for potato leaf disease detection are based on Deep Learning architectures, most commonly on Convolutional Neural Network (CNN) architecture, which is suitable for computer vision and image recognition. This paper depicts and compares the performances of the YOLOv11 Object Detection (Fast) model, YOLOv11s model, and Faster R-CNN X101-FPN model. These models were trained on a dataset developed for object detection in Roboflow. This dataset consists of 1200 images and 1500 annotations. A single object was labeled as one of the six classes: Pest, Bacteria, Fungi, Healthy, Phytophthora, and Nematode. Performance metrics show that these models achieve reputable results without excessive training time, making them suitable for real-time monitoring systems. YOLOv11 Object Detection (Fast), YOLOv11s, and Faster R-CNN X101-FPN achieved mAP50 scores of 95.1%, 97.6%, and 92.62%, respectively.

Keywords *– YOLO models, Faster R-CNN model, Roboflow, object detection, potato leaf disesase*

Citation: Medojević, S. (2024). Potato Leaf Disease Detection Using Faster R-CNN and YOLO Models. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 144-150.

I. INTRODUCTION

Potato is one of the most important food crop globally in terms of total food prodaction, standing alongside rice, wheat and corn [1]. The latest FAOSTAT data indicate that potato production surpassed 376 million metric tons globally in 2022 [2]. China, India, Ukraine, and Russia are the primary regions where potato markets and production are widely established. Potato production has significant impact on global economy. The potato market size is estimated at 115.74 billion dolars in 2024, and is expected to reach 137.46 billion dolars by 2029 [3]. Potato production is hampered by various potato diseases which contribute to the yield loss. Phytophthora infestans is the most widespread potato disease, accounting for up to 10 billion dolars in yield losses and management costs [4]. Early detection of diseases and damage play a crucial role in harvesting the maximal potato capasity.

The potato plant has a complex structure, with both underground and aboveground components as well as external and internal structures. Infected potato plants display visual symptoms that can be identified by inspecting the leaves. This greatly aids in prevention, early disease detection, and in stopping the further spread of infection, there by protecting healthy plants.

Categorization of potato leaf diseases can be based on the pathogens that cause them. These pathogens include bacteria, fungi, viruses, mycoplasma, nematodes, and adverse environmental conditions [5]. This paper focuses on the following categories of potato diseases: bacteria, fungi, Phytophthora, nematodes, and pests.

Diseases caused by bacteria primarily impact tubers and stems, with changes in the leaves being a byproduct of the bacterial infection. Bacterial infections prevent water absorption and impair the plant's ability to extract nutrients from the soil. One of the most common bacterial pathogens is Ralstonia solanacearum, a soil-borne bacterium that infects plants through the roots. Symptoms of bacterial infection include rapid wilting and curling of leaves, which can lead to the collapse of entire plants [6],[7].

Diseases caused by fungi can result from a broad group of organisms. Depending on the specific pathogen, potato leaves may exhibit various symptoms. One of the most common fungal diseases is Early Blight, caused by Alternaria solani. Symptoms of this infection include small, dark brown to black spots that appear in circular patterns. Additionally, leaves with slightly sunken spots may have yellow tissue surrounding the affected areas [6].

Phytophthora disease is caused by the oomycete plant pathogen Phytophthora infestans. One of the most common types is Late Blight. Symptoms of this infection include dark gray to brown water-soaked spots on leaf tissue, often surrounded by white, mold-like growth around the edges. Certain lesions may enlarge and develop into necrotic patches [4],[6].

Nematode infections can be divided into two main categories: root-knot nematodes and cyst nematodes. One

common symptom is chlorosis, which is caused by reduced chlorophyll content due to nutrient deficiencies, appearing as yellowing of the leaves [6],[8].

Potato pests can be divided into three categories: sucking pests, tuber and root damaging pests, and foliage feeders or defoliating pests. Symptoms of pest damage include distorted leaves with holes and/or leaves dotted with a silver coloration [6].

Healthy leaves appear as uniformly green leaves with no discoloration and a perfect leaf shape without any imperfections [6].

The development of advanced technology and the Internet of Things (IoT) has significantly transformed agriculture and improved sustainable agricultural practices. To maximize crop yields and improve resource management, traditional farming methods with limitations such as reliance on human labor, simple tools and machinery, and basic observation have needed to be replaced with smart farming methods. Smart farming incorporates IoT, Global Positioning Systems (GPS), sensors, robotics, drones, precision equipment, actuators, and data analytics for real-time monitoring of crops, soil, water, nutrients, and microclimate. These measurements help maintain soil quality, reduce soil degradation, conserve water resources, improve land biodiversity, and ensure a natural and healthy environment. Inspecting potato leaves for visible signs of infection aids in early identification, disease prevention, and minimizing the risk to uninfected plants, supporting sustainable agricultural practices. Real-time monitoring contributes to plant protection, product quality, fertilization, and disease detection. Leveraging available data and predictive models enables informed decision-making [9].

Artificial Intelligence (AI) is a term that encompasses a wide range of fields and techniques, some of which may overlap [10]. AI imitates human intelligence, with the ability to learn, recognize patterns, adapt, and create models based on previously acquired knowledge and data [11]. Deep learning is a subfield of machine learning that uses algorithms to analyze multi-layered representations of data, enabling the modeling of complex relationships within that data [12]. Various Deep Learning (DL) models are built upon Convolutional Neural Networks (CNN), with notable examples including the Region-Based Convolutional Neural Network (R-CNN), Mask Region-Based Convolutional Neural Network (Mask R-CNN), AlexNet, ResNet, Single Shot Multibox Detector (SSD), and YOLO (You Only Look Once) [13-15]. Such models have a wide range of applications in agriculture and can be used for detecting potato leaf diseases.

One of the first papers on this topic was published by Islam, Dinh, Wahid, and Bhowmik at the IEEE 30th Canadian Conference on Electrical and Computer Engineering in 2017 [16]. Significant advancements in this field have been made since 2020, marked by the publication of numerous scientific papers. The study by Ashikuzzaman, Roy, Lamon, and Abedin [17], compares the performances of nine Deep CNN models: Inception V3, VGG16, VGG19, InceptionResNetV2, NasNetMobile, NasNetLarge, ResNet50V2, ResNet101V2, and DenseNet201, with the DenseNet201 model achieving the highest validation accuracy of 96%. The main objective of the study by Zarrouk, Yandouzi, Grari, Bourhaleb, Rahmoune, and Hachami [18], focuses on detecting late blight disease using the following models: Faster-RCNN (RS50), Faster-RCNN (VGG19), Faster-RCNN (VGG16), YOLOv8, YOLOv7, and YOLOv6. The best-performing models are Faster-RCNN (RS50) with a precision of 93.92%, recall of 94.01%, and mAP of 95.32%, and Faster-RCNN (VGG16) with a precision of 91.96%, recall of 91.47%, and mAP of 93.22%. The paper by Kothari, Mishra, Gharat, Pandey, Gharat, and Thakur [19], focuses on comparing the performances of four models: CNN, GoogleNet, ResNet50, and VGG16. All of the models have classification accuracy around 97%. Papers written on this topic continue to improve. The majority of papers on this topic focus on image classification, with fewer studies addressing object detection and segmentation.

II. Materials and Method

A. An Overview of the Dataset

The data used in this paper were obtained from the dataset [20], originally developed by multiple teams from Multimedia Nusantara University and Gadjah Mada University. Images were collected from several potato farms, primarily located in Central Java, and in an uncontrolled environment. This method of data collection provided a wide range of image lighting, sharpness, angles, backgrounds, the number of leaves in an image, and types of diseases. The dataset consists of 3076 images, which were divided into seven classes: nematode, fungi, bacteria, pest, virus, Phytophthora, and healthy. The images are in JPEG color format with a resolution of 1500 x 1500 pixels.

In order to develop a new dataset, the data needed to be downloaded from the original dataset, cleaned, augmented, labeled, and annotated. The process of cleaning the data included selecting relevant data to be part of the new dataset, removing any anomalies that could negatively impact the training of models, and isolating relevant objects in the images to bypass excessive noise that can occur when there is an excessive number of objects in a single image. The original dataset was unbalanced. The Nematode class consists of 68 images, while the Fungi class consists of 748 images. To develop a balanced dataset, data needed to be augmented. The process of augmentation included data manipulation techniques such as rotation, manipulating the background of images, cropping, and blurring parts of images. The Roboflow platform was used to create the dataset. Preprocessing of the data included enabling the Auto-Orient option and resizing. All images were resized to 640x640 pixels and were prepared for annotation. Because this dataset is created for object detection, the annotation process involved drawing bounding boxes around objects and labeling them accordingly. Examples of infected and healthy leaves can be seen in Fig. 1.

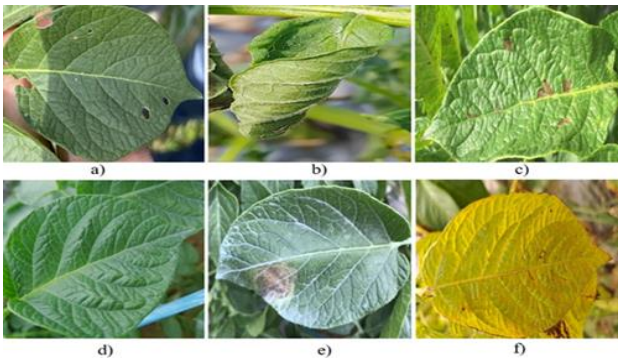


Fig. 1. Samples of leaves that fall into one of the six classes: a) Pest, b) Bacteria, c) Fungi, d) Healthy, e) Phytophthora, f) Nematode

A single object could be labeled as one of the six classes: Pest, Bacteria, Fungi, Healthy, Phytophthora, and Nematode. This process can be simplified by using Roboflow's new Auto Label feature.

This new dataset [21], consists of 1200 images and 1500 annotations. The dataset was divided into three sets, in the proportion of 70/20/10, that are used for training (840 images), validation (224 images), and testing (116 images). After finishing the dataset, dataset was exported in YOLOv11 format with TXT annotations and YAML configurations suitable for YOLOv11 model and COCO format with JSON annotations sutible for Efficient Det Pytorch and Detectron 2 models.

B. Selection of Deep Learning Models, Methods and Tools

A subset of Machine Learning, called Deep Learning, consists of deep neural networks. These neural networks are complex, multilayered structures made of interconnected nodes [22]. Deep neural network architecture has one input layer, hundreds or thousands of hidden layers, and one output layer. These layers enable the extraction of intricate features from the data.

The most commonly implemented deep learning architecture used for computer vision and image recognition tasks is Convolutional Neural Networks (CNN) [23]. A Convolutional Neural Network takes input data represented as a tensor. For an input image, three-dimensional tensors are commonly used, characterized by the image's height, width, and the number of channel layers. The number of channel layers corresponds to color channels (R, G, B) and is typically three for RGB images [24]. Convolutional Neural Network architecture consists of:

1) Convolution Layer,
2) Pooling Layer,
3) Fully-Connected Layer.

The Convolutional Layer is used to extract specific features by applying both linear and nonlinear operations, specifically convolutional operations and activation functions. Convolution is a particular linear operation specialized for feature extraction, where a kernel is applied across the input tensor. Both the input tensor and kernel are matrices. To calculate the value of a certain element, the element-wise product between the kernel and corresponding elements of the tensor needs to be summarized. The matrix that contains values calculated in this way is called a feature map. Stride is the distance between two consecutive positions of the kernel on the tensor. The process of building a feature map, where stride = 0, is shown in Fig. 2.

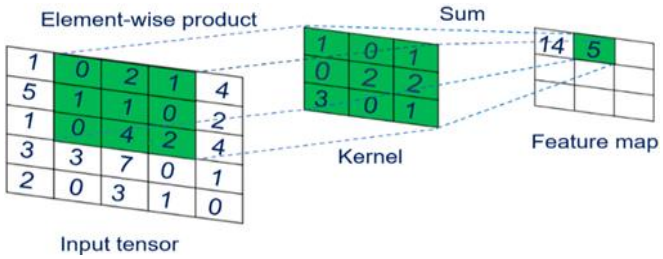


Fig. 2. Process of building the feature map

The dimension of the feature map depends on the dimensions of the tensor and the kernel. One way to increase the dimension of the feature map is by using zero-padding.

Zero-padding can be: valid padding, same padding, and full padding.

An activation function is applied after each convolutional layer. During the training of a model based on convolutional neural networks, the activation function introduces nonlinearity. This nonlinearity captures complex relationships among features within an image, enabling the model to identify hidden patterns and intricate connections between characteristics that linear operations alone would not be able to capture. Two of the most t commonly used activation functions are the Rectified Linear Unit (ReLU) and the Sigmoid function.

The Pooling Layer is a layer used for reducing the dimensions of the feature map. This layer is important because it decreases the number of learnable parameters, as well as the amount of data processed within the layer. By doing so, it reduces the memory requirements during training and helps mitigate the issue of overfitting. Two main types of pooling are: max pooling and average pooling.

Fully-Connected Layer is the final layer in a Convolutional Neural Network. As the name suggests, all nodes in this layer are fully connected to the nodes in the previous layer [24-26]. The process of transforming input data in a Convolutional Neural Network is shown in the Fig. 3.

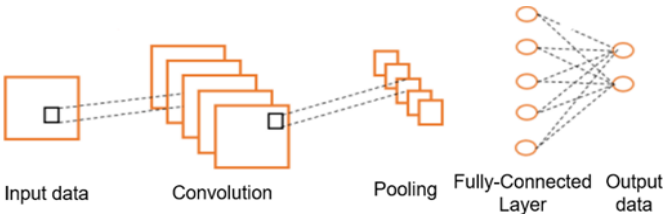


Fig. 3 Transforming input data in a Convolutional Neural Network

Convolutional neural networks (CNNs) can be used for classification, object detection, and image segmentation. This paper focuses on object detection.

Object detection involves locating an object within an image by marking the located object with a rectangular bounding box and classifying it into one of the predefined classes. Object detection methods can be divided into two groups: single-stage object detectors and two-stage object detectors. Single-Stage Object Detectors eliminate the need for a separate region of interest (RoI) extraction process, directly classifying and marking objects. Examples of single-stage object detectors include YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector). Two-Stage Object Detectors are network models that detect objects in two phases. In the first phase, regions of interest are identified, and in the second phase, the objects within these regions are classified [27]. The first part of the architecture is called the backbone. This part is responsible for extracting features from the input data. The extracted features are then passed to the second part, known as the neck. In the neck, the features from the backbone are aggregated and adjusted before being forwarded to the head for further processing. The head is the final part of the architecture, where the prediction is made [28]. The process of object detection in images can be represented by the architecture shown in Fig. 4.

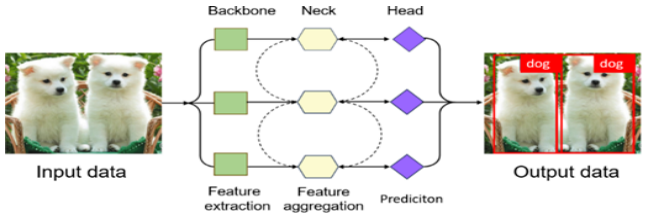


Fig. 4 Architecture of an object detection model

The Ultralytics YOLO model (You Only Look Once) is based on Convolutional Neural Networks. YOLO is a popular model for object detection and image segmentation. Although there are 11 versions of this model, this paper utilizes its latest version [29]. When using YOLOv11, it's possible to choose from several variants: YOLOv11n, YOLOv11s, YOLOv11m, YOLOv11l, and YOLOv11x. While YOLOv11x delivers high precision, its extended training time and large number of parameters demand substantial GPU resources, resulting in slower performance. In this paper, YOLOv11s was selected for its balanced precision, faster processing speed, and reduced parameter count, making it well-suited for rapid predictions. The YOLO model leverages the Ultralytics libraries, which are designed to work in Python. These libraries allow for easy configuration of training parameters, such as the number of epochs, image size, and task type (such as detection, segmentation, or classification). During model training, it is essential to set the number of epochs. An epoch represents one complete pass through the entire dataset. To ensure the model performs at a satisfactory level, it should be trained over a sufficiently large number of epochs.

The metrics used in this study are: recall, precision, mAP@0.50, and mAP@0.50-0.95. Recall is a metric used to calculate the rate of true positive instances. Precision is a metric used to calculate the model's ability to make positive predictions for attributes that are actually positive. The higher the precision, the more skilled the model is at identifying true positives and avoiding false positives. Mean Average Precision (mAP@0.50) represents the average precision calculated at an Intersection over Union (IoU) threshold of 0.50. It measures the model's accuracy while considering only "easier" detections. Mean Average Precision (mAP@0.50-0.95) is the average precision calculated across various IoU thresholds, ranging from 0.50 to 0.95. This metric provides a comprehensive view of model performance across different levels of detection difficulty [30].

The YOLOv11 Object Detection (Fast) model, developed by Roboflow, utilizes the COCO dataset as a checkpoint. This model offers faster training times, though with slightly lower accuracy compared to its counterpart, the Accurate model [31].

Region-based Convolutional Neural Network (R-CNN) is a Deep Learning framework used for object detection. R-CNN uses a Region Proposal Network (RPN) to suggest regions that potentially contain objects within an image. These regions are generated without annotated data. The algorithm employs a method called selective search, which is an approach that balances the number of proposals while maintaining high object recall, ensuring efficient object detection. The proposed regions are then processed by a CNN, which extracts features, and a binary Support Vector Machine (SVM), which helps identify objects in the regions. A bounding box regressor is used for refining the location and size of the bounding box to closely match the actual object, while the classifier predicts the category of each object.

R-CNN faces several limitations, including the rigid and non-learnable Selective Search algorithm, which can generate poor region proposals for object detection. Because of real-time applications and significantly increases disk memory usage, new variations of R-CNN have been introduced: Fast R-CNN, Faster R-CNN, Mask R-CNN, and Cascade R-CNN. Faster R-CNN improves the original R-CNN by integrating a Region Proposal Network (RPN), which generates region proposals directly from CNN feature maps, removing the need for selective search. It also shares convolutional features between the RPN and the detection network, which reduces computation time. As a result, Faster R-CNN achieves real-time processing speeds of approximately 0.1 seconds per image [32]. An iteration refers to a single update step during training. Detectron2 offers the tools and framework needed for developing and training a Faster R-CNN model [33].

III. Results

All of the models were developed in Google Colab using GPU T4. **YOLOv11 Object Detection (Fast) model** was trained for 300 epochs. Fig. 5 captures changes in mAP50 and mAP50:95 throughout 300 epochs.

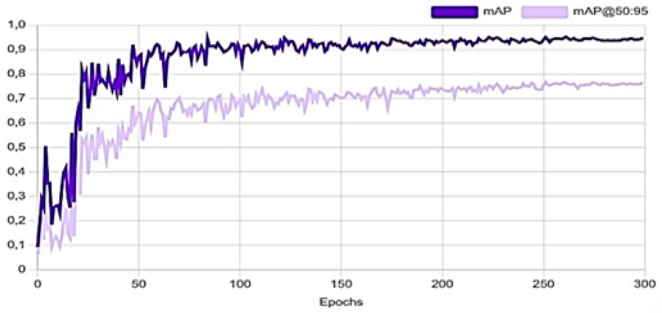


Fig. 5 Changes in mAP50 and mAP50:95 throughout 300 epochs for the YOLOv11 Object Detection (Fast) model

After 300 epochs, the model has an mAP50 of 95.1% and an mAP50:95 of 76.7%. This model has a precision of 96.4% and a recall of 90.2%. Fig. 6 shows how this model detects different classes.

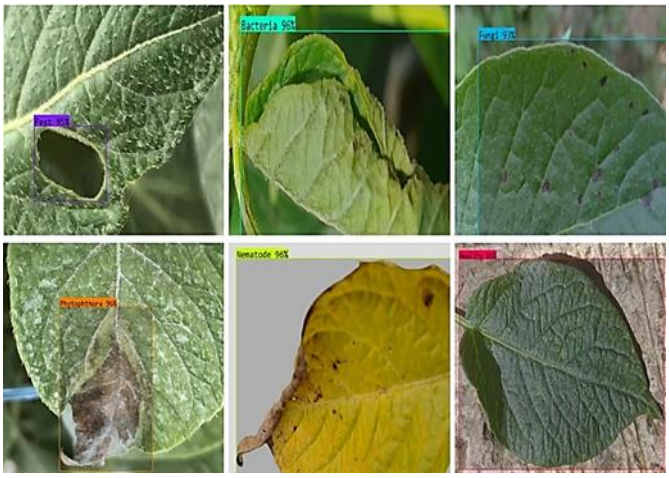


Fig. 6 Different results of the YOLOv11 Object Detection (Fast) model

YOLOv11s model was trained for 300 epochs.

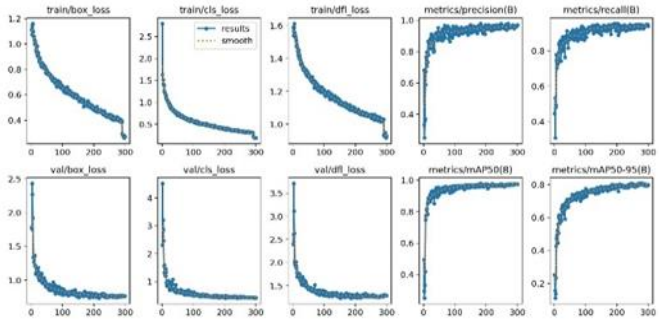


Fig. 7 Yolov11s key performance metrics, train and validation losses during model training process

Throughout the model training process, it is important to monitor training and validation losses as well as other performance metrics. These changes are represented in Fig. 7. This model has a precision of 96.1%, recall of 93.1%, mAP50 of 97.6%, and mAP50-95 of 80.6%. In-depth performance metrics are shown in Table 1.

Table 1. Performance metrics of the YOLOv11s model by class

Class	Precision	Recall	mAP50	mAP50-95
All	96.1%	93.1%	97.6%	80.6%
Bacteria	93.1%	91.8%	95%	77.1%
Fungi	97.9%	97.6%	98.7%	80%
Healthy	100%	91.1%	99.5%	95.3%
Nematode	90.4%	97.7%	97.9%	92.6%
Pest	96.8%	82.8%	94.9%	66.8%
Phytophthora	98.2%	97.7%	99.3%	71.8%

A normalized confusion matrix is one of the tools used in evaluating the performance of the model by comparing true and predicted detections. The normalized confusion matrix for this model has a prominent main diagonal with fewer values in the row and column used for showing true and predicted detections of the background. The normalized confusion matrix is represented in Fig. 8.

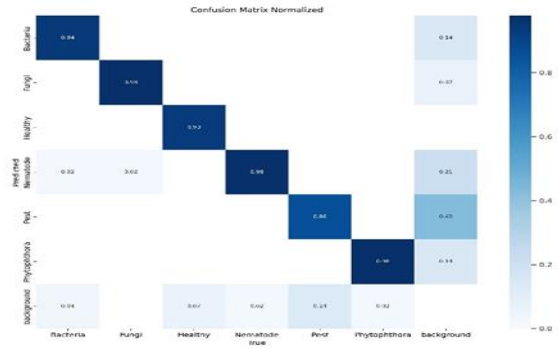


Fig. 8 Normalized confusion matrix of the YOLOv11s model

A graphical representation illustrating the variation in a model's F1 score across different thresholds is known as an F1-Confidence Curve, shown in Fig. 9. This graph shows that F1 score is 0.94 for all the classes when the threshold is set to 0.471.

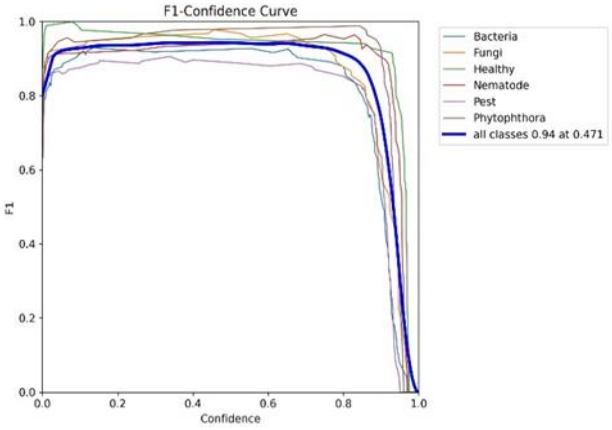


Fig. 9 F1-Confidence Curve of the YOLOv11s model

After developing the model, the results can be seen in Fig. 10.

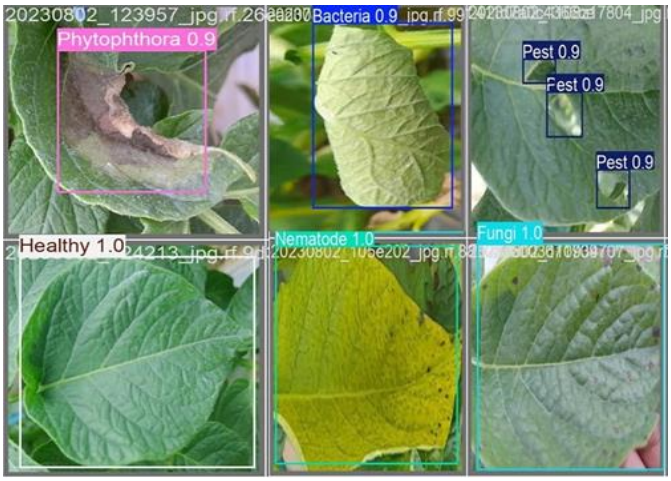


Fig. 10 Different results of the YOLOv11s model

Detectron2 Faster R-CNN X101-FPN model was trained for 2000 iterations, with a batch size of 4 images and a base learning rate of 0.001. After training, this model achieved an AP50 of 92.62%, APs of 12.08%, APm of 58.97%, and APl of 70.83%. The AP for classes: Bacteria, Fungi, Healthy, Nematode, Pest, and Phytophthora are 69.39%, 70.68%, 82.08%, 85.04%, 50.90%, and 63.81%, respectively. The results of this model are shown in Fig. 11.

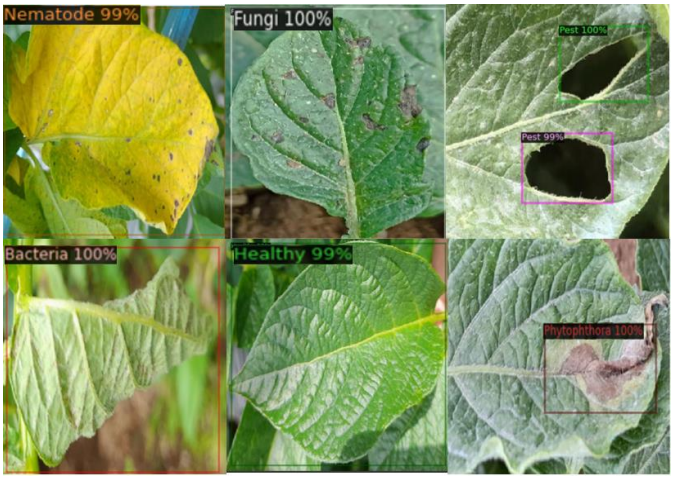


Fig. 11 Different results of the Faster RCNN X101-FPN model

Table 2 depicts mAP50, the number of epochs/iterations, and the time spent while training the previously shown three models.

Table 2. Performance metrics of the YOLOv11s model by class

Model	mAP50	Epochs/Iterations	Training time
YOLOv11 Object Detection (Fast)	95.1%	300 Epoches	1 hour
YOLOv11s	97.6%	300 Epoches	1.87 hours
Faster R CNN X101-FPN	92.62%	2000 Iteration	1.58 hours

IV.DISCUSSION

The YOLOv11 Object Detection (Fast) model prioritizes processing speed, which is often used to achieve real-time performance. Fig. 5 shows that the model achieves reputable results at the 200th epoch, after which the mAP is slightly refined. YOLOv11s has the longest training time but the best performance metrics out of these three models. The class Healthy has the best performance, while the class Pest has the worst performance out of all six classes. The reason for this is that different pests can cause numerous varying size holes on a single leaf. Depending on the background, some of these holes can be hard to detect. When a leaf is drastically damaged, the model may struggle with drawing bounding boxes because it is hard to identify where one damage ends and another pest damage begins. The confusion matrix showed that the model's most frequent errors are misidentifying background elements as objects or failing to detect actual objects, considering them part of the background. This was expected because the dataset contains numerous images with multiple leaves, some with more or less blurred backgrounds, making it difficult for the model to differentiate between leaves and background. The Faster R-CNN X101-FPN model, although showing good AP, has the longest training time and the worst AP out of these three models. Just like the other models, this model also struggles to identify the Pest class, which has the worst AP out of all six classes. The best performance is seen in the Nematode and Healthy classes. The AP values indicate that the model performs best at recognizing large objects and worst at recognizing small objects, which is a common trend for most models. Reason for this is that larger object occupy more pixels and contain more identifiable features, making it easier for the model to detect and classify them accurately, while smaller object have fewer distinctive features and often merge with the background.

V.CONCLUSION

All three models demonstrate strong performance and efficient training times. The YOLOv11s model achieves the best metric results, although it has a slightly longer training time compared to the other two models. The YOLOv11 Object Detection (Fast) model demonstrated strong performance metrics while having the shortest training time. These models can be deployed within surveillance systems to enable real-time monitoring and predictive analytics. Future improvements could include developing a larger dataset and utilizing more powerful GPUs capable of supporting models that achieve higher precision while maintaining rapid prediction speeds.

Statement of Research and Publication Ethics

The author declares that this study complies with Research and Publication Ethics

REFERENCES

[1] Y. P. S. Bajaj, *Potato*, vol 3. Springer Science & Business Media, 2013.

[2] (2024) Food and Agriculture Organization of the united Nations. [Online]. Available: https://www.fao.org/faostat/en/#search/potato

[3] (2024) Potato Market Size- ndustry Report on Share, Growth Trends & Forecasts Analysis (2024-2029) on Mordor Intelligence [Online]. Available:https://www.mordorintelligence.com/industry-reports/potato-market

[4] S. M. Dong, S.Q. Zhou,“ Potato late blight caused by Phytophthora infestans: From molecular interactions to integrated management strategies,” Journal of Integrative Agricultur, vol. 21, pp. 3456--3466, Dec. 2022.

[5] S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[6] W. J.Hooker., *Compendium of potato diseases. International Potato Center*, International Potato Center, 1981.

[7] A. Charkowski, K. Sharma, M. L. Parker, G. A. Secor and J. Elphinstone, “Bacterial Diseases of Potato” The potato crop: its agricultural, nutritional and social contribution to humankind, pp. 351--388, Dec. 2019.

[8] M. Sun, S. Chen and J. E. Kurle, “Interactive effects of soybean cyst nematode, arbuscular-mycorrhizal fungi, and soil pH on chlorophyll content and plant growth of soybean“ Phytobiomes Journal., vol. 6, pp. 95--105, Jan. 2022.

[9] M. Dhanaraju, P. Chenniappan, Poongodi, K. Ramalingam, S. Pazhanivelan, R. Kaliaperumal, “Smart farming: Internet of Things (IoT)-based sustainable agriculture” ,Agriculture, vol. 12, pp. 1745 , Sep. 2022.

[10] (2024) Subsets of Artificial Intelligence on Free Learning Platform for Better Future. [Online]. Available: https://tinyurl.com/2azettr3

[11] C. R. Arias, “An introduction to artificial” AI, Faith, and the Future: An Interdisciplinary Approach., pp. 12, Jun. 2022.

[12] L. Deng, Y. Dong, “Deep learning: methods and applications.” Foundations and trends® in signal processing., vol. 7, pp. 197--387 , 2014.

[13] (2024) R-CNN – Region-Based Convolutional Neural Networks on GeeksforGeeks. [Online]. Available: https://www.geeksforgeeks.org/r-cnn-region-based-cnns/

[14] (2024) Different types of CNN models on Opengenus. [Online]. Available: https://iq.opengenus.org/different-types-of-cnn-models/

[15] (2024) Ultralytics YOLO11. [Online]. Available: https://docs.ultralytics.com/models/yolo11/

[16] M. Islam, A. Dinh, K. Wahid, P. Bhowmik, Detection of potato diseases using image segmentation and multiclass support vector machine., 2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE), 2017 .

[17] M. Ashikuzzaman, K. Roy, A. Lamon, S. Abedin, Potato Leaf Disease Detection By Deep Learning: A Comparative Study., 2024 6th International Conference on Electrical Engineering and Information \& Communication Technology (ICEEICT), 2024 .

[18] Y. Zarrouk, M. Yandouzi, M. Grari, M. Bourhaleb, M. Rahmoune, K. Hachami., “Revolutionizing Potato Late Blight Surveillance: uav-driven Object Detection Innovations, Journal of Theoretical and Applied Information Technology., vol. 102, Apr. 2024.

[19] D. Kothari, H. Mishra, M. Gharat, V. Pandey, M. Gharat, R. Thakur, “Potato leaf disease detection using deep learning” Int. J. Eng. Res. Technol, vol. 1, pp. 569–571, Nov. 2022.

[20] (2023) Potato Leaf Disease Dataset in Uncontrolled Environment on mendeley data. [Online]. Available: https://data.mendeley.com/datasets/ptz377bwb8/1

[21] (2024) Potato_leaf_disease Computer Vision Project dataset on Roboflow. [Online]. Available: https://universe.roboflow.com/potato-leaf-diseases/potato_leaf_disease

[22] (2024) What is deep learning? on IBM. [Online]. Available: https://www.ibm.com/topics/deep-learning

[23] (2024) Difference between Shallow and Deep Neural Networkson on GeeksforGeeks [Online]. Available: https://www.geeksforgeeks.org/difference-between-shallow-and-deep-neural-networks/

[24] (2024) Convolutional Neural Network (CNN) on TensorFlow [Online]. Available: https://www.tensorflow.org/tutorials/images/cnn

[25] R. Yamashita, M. Nishio, R. K. G. Do, K. Togashi, "Convolutional neural networks: an overview and application in radiology" Insights into imaging, vol. 9, pp. 611--629, Jun. 2018.
[26] (2024) What are convolutional neural networks? on IBM. [Online]. Available:https://www.ibm.com/topics/convolutional-neural-networks
[27] (2019) Object Detection: Architectures, Models, and Use Cases on The Random Walk Blog. [Online]. Available: https://randomwalk.ai/ blog/object-detection-architectures-models-and-use-cases/
[28] (2024) What is object detection? on IBM. [Online]. Available: https://www.ibm.com/topics/object-detection
[29] G. Jocher, J. Qiu (2024), Ultralytics YOLO11, version = 11.0.0, year = {2024}, [Online].Available:https://github.com/ultralytics/ ultralytics
[30] H. Vedoveli (2013) Metrics Matter: A Deep Dive into Object Detection Evaluation.[Online].Available:https://medium.com/@henriquevedoveli/metrics-matter-a-deep-dive-into-object-detection-evaluationef01385ec62
[31] J. Solawetz, P. Guerrie.(2022). What to Think About When Choosing Model Sizes. Roboflow Blog: https://blog.roboflow.com/ computer-vision-model-tradeoff/
[32] (2024) R-CNN – Region-Based Convolutional Neural NetworksNetworkson on GeeksforGeeks [Online]. Available: https://www.geeksforgeeks.org/r-cnn-region-based-cnns/
[33] (2019) Y. Wu, A. Kirillov, F. Massa, et al. Detectron2.[Online]. Available: https://github.com/facebookresearch/detectron2

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 151-158
DOI: 10.36287/ijmsit.8.2.16
Received: November 28, 2024; Accepted: December 25, 2024

AI-Powered Classification of Oral Lesions: Improving Early Detection and Diagnosis

Hakan Yılmaz[1*], Mehmet Özdem[2]

[1*] *Medical Engineering Dept., Karabuk University, Karabuk, Türkiye (hakanyilmaz@karabuk.edu.tr) (ORCID: 0000-0002-8553-388X)*
[2] *Türk Telekom, Ankara, Türkiye (mehmet.ozdem@turktelekom.com.tr) (ORCID: 0000-0002-2901-2342)*

Abstract **–** Oral malignancies pose significant global health challenges, with oral squamous cell carcinoma (OSCC) being the most prevalent form. Early detection of potentially malignant oral disorders (OPMDs) such as leukoplakia and oral submucous fibrosis is crucial for improving patient prognosis. Traditional diagnostic approaches often face limitations like subjective interpretation and potential delays. This study aimed to develop and evaluate a deep learning-based model for the classification of oral lesions as benign or malignant using publicly available image datasets. Utilizing a modified VGG16 architecture and optimized preprocessing techniques, the model was trained on 330 annotated intraoral images and achieved an overall accuracy of 94.79%. Key performance metrics included a precision of 95.11%, sensitivity and specificity of 94.58%, and an F1 score of 94.74%. The model's performance was comparable to or exceeded existing models with larger datasets, demonstrating its capability for effective feature extraction and reliable classification. The high area under the curve (AUC) value of 0.96 reinforced its potential for clinical application. While the model showed strong diagnostic capability, expanding the dataset size and incorporating a broader range of cases could further enhance generalizability. Future work should also consider integrating real-time image acquisition and optimizing computational processes for practical deployment. The findings underscore the promise of AI-driven diagnostic tools in supporting healthcare professionals by enabling timely, accurate, and scalable detection of oral malignancies, thereby contributing to improved patient care and outcomes. This study represents a significant step toward the practical application of AI in oral health diagnostics.

Keywords *– OSCC detection, Oral lesions, OPMD classification, VGG16 model, Malignancy prediction.*

Citation: Yılmaz, H., Özdem, M. (2024). AI-Powered Classification of Oral Lesions: Improving Early Detection and Diagnosis. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 151-158.

I. INTRODUCTION

Oral malignancies represent a significant health concern worldwide, encompassing a diverse range of neoplastic conditions that arise within the oral cavity. Among these, oral squamous cell carcinoma (OSCC) is the most prevalent, accounting for a substantial proportion of oral cancers. The World Health Organization has categorized various lesions as oral potentially malignant disorders (OPMDs), which include conditions such as leukoplakia, erythroplakia, and oral submucous fibrosis (OSMF) [1], [2]. These disorders are characterized by an increased risk of malignant transformation, with OSMF alone exhibiting a transformation rate of 3-13% into OSCC [3]. The recognition and management of these lesions are critical, as early detection can significantly improve patient outcomes and survival rates [4].

The clinical presentation of oral malignancies can be insidious, often leading to delayed diagnosis. For instance, oral malignant melanoma, although rare, is associated with a poor prognosis, with a 5-year survival rate ranging from 10% to 25% [5]. Symptoms may not manifest until the disease has progressed significantly, underscoring the importance of regular oral examinations and awareness of potential signs of malignancy [6]. Furthermore, the presence of oral lesions can severely impact patients' quality of life, affecting their ability to eat, speak, and maintain social interactions [7].

The etiology of oral malignancies is multifactorial, with risk factors including tobacco use, alcohol consumption, and chronic irritation from ill-fitting dentures or dental appliances [8].Additionally, systemic conditions such as hematological malignancies can present with oral manifestations, complicating the clinical picture [9]. The interplay between local and systemic factors necessitates a comprehensive approach to diagnosis and treatment, emphasizing the need for interdisciplinary collaboration among healthcare providers [10].

The application of artificial intelligence (AI) in the detection of oral malignancies has emerged as a pivotal advancement in modern dentistry and oncology. Traditional diagnostic methods, while valuable, often suffer from limitations such as reliance on subjective interpretation and the potential for human error, which can delay diagnosis and treatment [11].Recent studies have demonstrated that AI can significantly reduce the time required for diagnosis, thereby addressing critical delays that often occur in traditional diagnostic processes [12], [13]. Such AI-driven solutions could transform the current diagnostic landscape by enabling early detection in both clinical and community settings, thereby improving treatment outcomes and reducing mortality rates.

Warin et al. [14] evaluated deep convolutional neural network (CNN) algorithms for classifying and detecting OPMDs and OSCC using a dataset of 980 oral images. Various CNN architectures were used for image classification, with DenseNet-169 achieving the best performance. The model achieved high diagnostic accuracy metrics for detecting OSCC and OPMD in oral images. For OSCC, precision, sensitivity, and specificity were each 99%, with an F1 score of 98% and an area under the curve (AUC) of 1.0. For OPMD detection, the model recorded 95% precision, sensitivity, and F1 score, with 97% specificity and an AUC of 0.98. These results indicate that CNN models, particularly DenseNet-169, can perform at expert levels, making them promising tools for assisting general practitioners in early oral cancer detection.

Jubair et al. [15] aimed to develop a lightweight CNN for binary classification of oral lesions into benign or malignant/potentially malignant using real-time clinical images. The model utilized EfficientNet-B0, which is known for achieving state-of-the-art accuracy on large datasets while being smaller and faster than traditional CNNs, for transfer learning and was trained on 716 clinical images. The performance metrics included an accuracy of 85%, specificity of 84.50%, sensitivity of 86.70%, and an AUC of 0.93. These results suggest that CNN models can be effectively used to build cost-efficient, embedded AI devices with limited computational power for oral cancer screening and early detection, potentially expanding screening capabilities.

Huang et al. [16] presented a deep-learning model based on a metaheuristic approach for the accurate diagnosis of oral cancer, focusing on early detection to save lives. It used three preprocessing techniques—Gamma correction, noise reduction, and data augmentation—to enhance image quality and boost dataset size. Weights of the CNN were optimized using an improved version of the Squirrel Search Algorithm (ISSA) to increase accuracy. The model was tested on the "Oral Cancer (Lips and Tongue) Images" dataset from Kaggle, containing 131 images classified by ENT specialists. The dataset was split into 70% for training and 30% for testing. The proposed model achieved an accuracy of 97%, precision of 92.66%, sensitivity of 87.34%, and F1-score of 89.37%, demonstrating superior results compared to existing methods. However, the complex nature of both the CNN and the metaheuristic increases time complexity. Despite this, the model shows promise for adapting to different types of cancer diagnosis.

Fu et al. [17] aimed to develop a rapid, non-invasive, and cost-effective deep learning approach to identify oral cavity squamous cell carcinoma (OCSCC) using photographic images. The researchers employed cascaded convolutional neural networks, training them on 44,409 biopsy-proven OCSCC and normal control images from 11 hospitals in China collected over 13 years. The dataset was divided into development and internal validation sets, with an additional external validation set sourced from dental and oral surgery journals. It achieved an AUC of 0.98, a sensitivity of 94.90%, and a specificity of 88.70% on the internal validation dataset. The results suggest that this automated deep-learning approach is a viable clinical tool for fast screening, early detection, and assessment of therapeutic efficacy, demonstrating performance comparable to human specialists.

Bansal et al. [18] aimed to develop a new CNN model, termed "Oral_Cancer_Detection," to classify oral cancer images of lips and tongue into cancerous and non-cancerous categories. The model was trained using a small Kaggle dataset with 131 images (87 cancerous, 44 non-cancerous), incorporating data augmentation, feature extraction, and classification techniques. Implemented in MATLAB, the model achieved a 94% validation accuracy after 132 iterations. Key performance metrics showed a precision and specificity of 100%, sensitivity of 91%, and F1 score of 94%, indicating strong performance despite the dataset's limited size. The model is characterized by low computational requirements, making it effective for rapid cancer classification. While results are promising, the study suggests that increasing dataset size and training parameters could further improve accuracy, albeit with longer processing times.

Lin et al. [19] aimed to enhance the accuracy of smartphone-based deep learning methods for detecting oral diseases, focusing on improving diagnosis through systematic data collection and algorithm optimization. A centered image-capturing approach was developed to collect clear oral cavity images, leading to the creation of a medium-sized dataset with five disease categories: normal, ulcer, low-risk, high-risk, and cancer. A resampling method was also introduced to reduce variability from handheld smartphone cameras. The study employed the HRNet model, achieving a sensitivity of 83%, specificity of 96.60%, precision of 84.30%, and F1 score of 83.60% on 455 test images. The "center positioning" method improved the F1 score by about 8% over a simulated "random positioning" approach, while resampling added a further 6% performance boost. HRNet outperformed models like VGG16, ResNet50, and DenseNet169. The results highlight that smartphone-based imaging, when combined with targeted image capture, resampling, and HRNet, holds promise for primary oral cancer diagnosis.

Tanriver et al. [20] explored the potential of computer vision and deep learning for the automated detection of OPMDs using photographic images. With a two-stage pipeline, the model first detected lesions and then classified them as benign, OPMD, or carcinoma. Using EfficientNet-B4, the model achieved precision, sensitivity, specificity, and F1-score of 87%, 86%, 86%, and 86%, respectively, on the test set. The findings underscore the feasibility of this deep-learning approach as a low-cost, non-invasive tool that can support early screening processes and enhance OPMD detection, contributing to improved oral cancer outcomes. The model's real-time capabilities show potential for broader clinical use, aiding timely diagnosis and treatment.

In summary, oral malignancies encompass a spectrum of conditions that pose significant challenges in terms of diagnosis, treatment, and patient management. The incorporation of AI in the detection of oral malignancies represents a significant leap forward in the field of oral health. By improving diagnostic accuracy, facilitating early detection, and enabling personalized treatment approaches, AI technologies are poised to transform the landscape of oral cancer management.

This study aims to develop and evaluate a deep learning-based approach for the accurate classification of oral lesions as benign or malignant using publicly available image datasets. By employing a modified VGG16 architecture and optimized preprocessing techniques, the research seeks to improve diagnostic accuracy, sensitivity, and specificity compared to existing models. The overarching goal is to provide a reliable, efficient, and scalable tool that can assist healthcare professionals in early detection and diagnosis of oral

malignancies, thereby enhancing patient outcomes and facilitating timely treatment interventions.

II. Materials and Method

A. Dataset Used and Programming Environment

In this study, the publicly available "Oral Images Dataset," published by Chandrashekar et al., was utilized (Chandrashekar et al., 2021). This dataset contains color images of oral lesions captured using mobile cameras and intraoral cameras. Lesion areas within the dataset were subsequently labeled as benign or malignant by experts using the VGG Image Annotator (VIA) tool. VGG is an open-source, JavaScript-based application commonly used for image annotation (Dutta & Zisserman, 2019). The labeled image regions were cropped from the original images and resized to a resolution of 224x224 pixels. This process resulted in a dataset comprising a total of 330 images, with 162 labeled as benign and 168 as malignant, to be used in the study. A train-test split with a ratio of 0.3 was applied to the dataset, which comprises a total of 330 images (162 labeled as benign and 168 as malignant). Following the split, 234 images were allocated for training, while 96 images were set aside for testing Sample images from the resulting dataset are presented in Figure 1.

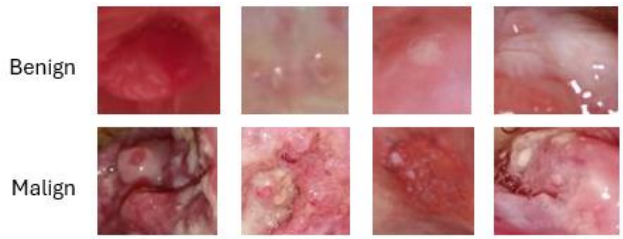


Fig. 1. Sample images from dataset.

The dataset was created, and preprocessing tasks on the images were carried out using a computer with an Intel i7 processor and 16 GB of RAM, alongside the Python programming language.

B. Convolutional Neural Network (CNN)

CNNs have emerged as a cornerstone of modern deep learning, particularly in the realm of image processing and computer vision. Their architecture is inspired by the human visual system, allowing them to effectively recognize patterns and features in visual data. CNNs utilize a hierarchical structure that includes convolutional layers, pooling layers, and fully connected layers, enabling them to learn increasingly abstract representations of input data as it progresses through the network (Bai & Li, 2023; O'Shea & Nash, 2015; Zakaria, 2023). This architecture is particularly adept at handling the spatial hierarchies inherent in images, making CNNs a preferred choice for tasks such as image classification, object detection, and medical image analysis (Kwiatkowska et al., 2021; Tian, 2020).

CNNs have emerged as a cornerstone of modern deep learning, particularly in the realm of image processing and computer vision. Their architecture is inspired by the human visual system, allowing them to effectively recognize patterns and features in visual data. CNNs utilize a hierarchical structure that includes convolutional layers, pooling layers, and fully connected layers, enabling them to learn increasingly abstract representations of input data as it progresses through the network [21], [22], [23]. This architecture is particularly adept at handling the spatial hierarchies inherent in images, making CNNs a preferred choice for tasks such as image classification, object detection, and medical image analysis [24], [25].

Among the various architectures developed for CNNs, the Visual Geometry Group (VGG) model, specifically VGG16, has garnered significant attention due to its depth and performance. The VGG16 architecture is a prominent model in the field of deep learning, particularly known for its application in image classification tasks. Developed by the Visual Geometry Group at the University of Oxford, VGG16 is characterized by its deep architecture, consisting of 16 layers with learnable parameters, which include 13 convolutional layers and 3 fully connected layers [23], [26]. The convolutional layers utilize small receptive fields of 3x3 pixels, which allows the network to capture fine-grained features in images while maintaining a manageable number of parameters [27]. This design choice is crucial as it enables the model to learn hierarchical representations of the input data, progressively extracting more complex features as the data passes through the layers [23]. VGG16's performance has been validated across numerous benchmarks, making it a popular choice for transfer learning in various applications, including medical imaging, where it has been employed for tasks. The application of VGG16 and similar CNN architectures has revolutionized fields, where automated systems can assist in the early detection, significantly improving diagnostic accuracy [24], [28].

C. Metrics Used

In the evaluation of machine learning models, several key performance metrics are commonly utilized to assess their effectiveness in classification tasks. These metrics include precision, accuracy, sensitivity (often referred to as recall), specificity, F1 score, confusion matrix, and the Receiver Operating Characteristic (ROC) curve. Each of these metrics provides unique insights into the model's performance [29].

Precision is defined as the ratio of true positive predictions to the total number of positive predictions made by the model. It indicates how many of the predicted positive cases were actually positive, thereby reflecting the model's ability to avoid false positives [30]. Accuracy, on the other hand, measures the overall correctness of the model by calculating the ratio of correctly predicted instances (both true positives and true negatives) to the total instances evaluated. While accuracy is a straightforward metric, it can be misleading in cases of imbalanced datasets where one class significantly outnumbers the other [31]. Sensitivity, or recall, quantifies the model's ability to correctly identify positive instances. It is calculated as the ratio of true positives to the sum of true positives and false negatives. High sensitivity is crucial in scenarios where missing a positive case (e.g., a disease diagnosis) could have severe consequences [32]. Conversely, specificity measures the proportion of true negatives correctly identified, providing insight into the model's ability to avoid false negatives. It is calculated as the ratio of true negatives to the sum of true negatives and false positives [33]. The F1 score is a harmonic mean of precision and recall, offering a single metric that balances the two. It is particularly useful in situations where there is a need to find an optimal balance between precision and recall, especially in imbalanced datasets [34]. The confusion matrix is a comprehensive tool that summarizes the

performance of a classification model by displaying the counts of true positives, true negatives, false positives, and false negatives. This matrix allows for a detailed analysis of the model's performance and helps identify areas for improvement [35]. Lastly, the ROC curve is a graphical representation that illustrates the trade-off between sensitivity (true positive rate) and specificity (false positive rate) at various threshold settings. The area under the ROC curve (AUC) serves as a single scalar value to summarize the model's performance across all thresholds, with higher AUC values indicating better model performance [36]. Together, these metrics provide a robust framework for evaluating the effectiveness of machine learning models in classification tasks, enabling practitioners to make informed decisions based on the specific requirements of their applications.

The formulas for calculating accuracy, precision, sensitivity (recall), specificity, and the F1 score are presented in Equations 1, 2, 3, 4, and 5, respectively.

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

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

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (3)$$

$$Specificity = \frac{TN}{TN+FP} \quad (4)$$

$$F_1 Score = \frac{2*TP}{2*TP+FP+FN} \quad (5)$$

III. Results

A. Network Design and Training Information

In this study, modifications were made to the final two layers of the VGG16 model to incorporate specific parameters. A dense layer with 256 neurons and a ReLU activation function was added, followed by an output layer with a sigmoid activation function. The "Adam" optimizer was selected, and the model was configured to run for 200 epochs; however, early stopping was triggered after the 138th epoch, concluding the training. The batch size was set to 32. Initially, the loss value was 3.32, which decreased to 0.15 by the end of training. Similarly, the accuracy began at 0.43 and gradually increased, reaching 0.95. Graphs illustrating the changes in loss and accuracy are shown in Figure 2.

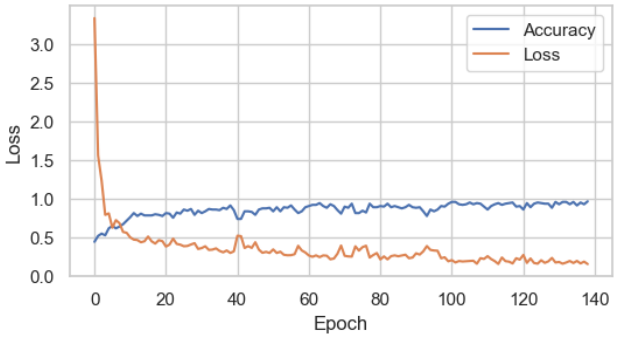


Fig. 2. Training information.

B. Evaluation of Training

In this study, a deep learning model was developed to classify intraoral lesion images captured by cameras as benign or malignant. Following the model evaluation, key performance metrics—including accuracy, precision, sensitivity, specificity, and F1 score—were calculated for each class. The findings are summarized below.

For benign lesions, the model achieved an accuracy of 94.79%, a precision of 97.62%, a sensitivity of 91.11%, and a specificity of 98.04%. The F1 score for the benign class was 94.25%. For malignant lesions, the model demonstrated similar robustness, with an accuracy of 94.79%, a precision of 92.59%, a sensitivity of 98.04%, and a specificity of 91.11%. The F1 score for the malignant class reached 95.23%.

These results indicate that the model performs well in distinguishing between benign and malignant lesions, showcasing strong precision and sensitivity across both classes. The obtained metrics are presented in Table 1. Following the model evaluation, a confusion matrix was generated to capture the classification performance for each class. This matrix enabled a detailed analysis of Type I and Type II errors across both classes, facilitating insights into the model's misclassification tendencies. Figure 3 presents the confusion matrix, which provides a comprehensive view of the model's ability to correctly identify benign and malignant cases and highlights areas for potential improvement in accuracy.

Table 1. Evaluation metrics of results.

Class	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1 Score (%)
Benign	94.79	97.62	91.11	98.04	94.25
Malign	94.79	92.59	98.04	91.11	95.23
Average	94.79	95.11	94.58	94.58	94.74

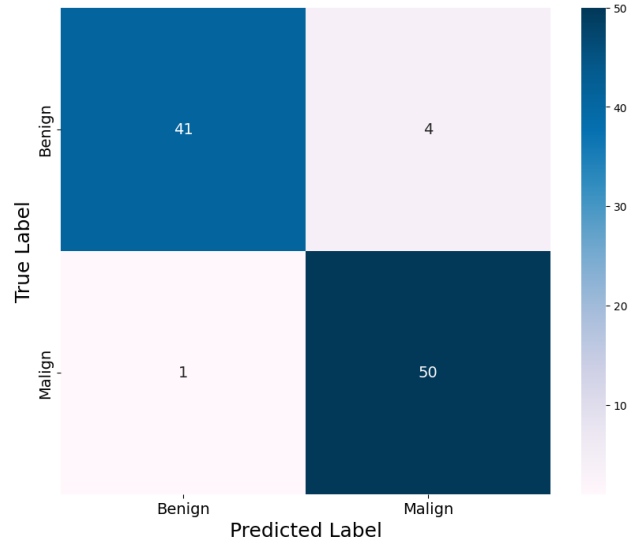


Fig. 3. Confusion matrix of results.

The True Positive Rate (TPR) matrix illustrates the proportion of actual positive cases accurately identified by the model in each class, providing insight into the model's effectiveness in detecting specific conditions. High TPR values across classes suggest that the model performs well in recognizing positive instances, thereby reducing the occurrence of false negatives and enhancing diagnostic reliability for each condition. The TPR matrix of the results is presented in Figure 4.

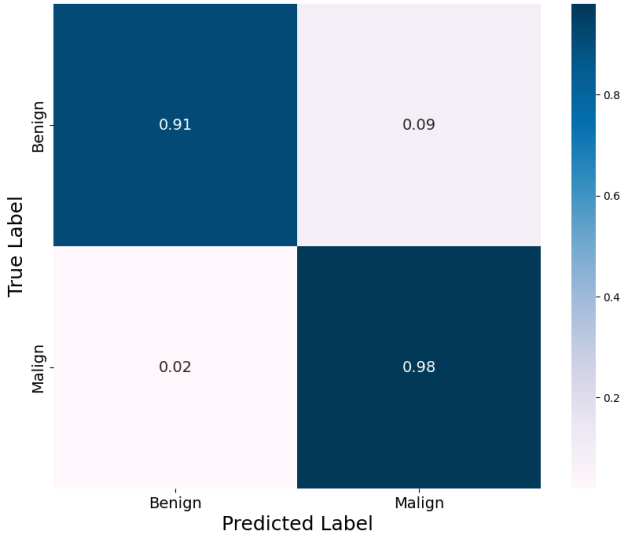


Fig. 4. TPR matrix of results.

Finally, the Area Under the Curve (AUC) value was calculated, and the Receiver Operating Characteristic (ROC) curve was generated to further evaluate the model's performance. This ROC curve, presented in Figure 5, visually represents the trade-off between sensitivity and specificity, providing an additional measure of the model's classification effectiveness.

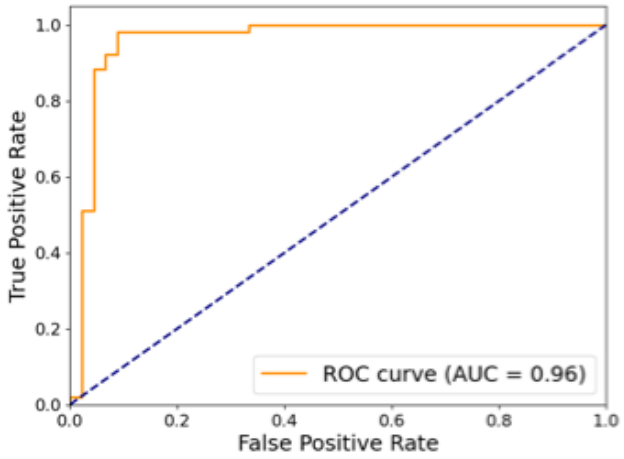


Figure 1. ROC curve and AUC value of results.

IV. DISCUSSION

The results from this study reveal a strong performance of the proposed deep learning model in classifying benign and malignant oral conditions from images. The model achieved an impressive accuracy of 94.79% on a dataset of 330 images, highlighting its capacity for reliable classification even with a relatively small dataset size. The relevant studies are given in Table 2 along with their performance metrics.

The precision of 95.11% signifies that the model maintains a low false positive rate, correctly identifying benign and malignant cases with high confidence. In terms of sensitivity, the model achieved 94.58%, indicating its effectiveness in detecting true positive cases of malignancy. This sensitivity is notable, especially when compared to models like EfficientNet-B4, which reached 86% sensitivity on a dataset of 684 images, suggesting that the proposed model better identifies malignant cases.

Moreover, the model's specificity was also 94.58%, reflecting a balanced capability to avoid false alarms by correctly identifying non-malignant cases. This level of specificity is comparable to the cascaded CNN's 88.7%, confirming that the model offers a significant reduction in false positives, making it suitable for clinical use where overdiagnosis can be problematic.

The F1 Score of 94.74% illustrates the model's balanced handling of precision and recall, showcasing its suitability for real-world application where both measures are critical. This score also suggests that the model performs well in maintaining a strong balance between false positives and false negatives, which is crucial in clinical diagnostics.

The AUC of 0.96 further reinforces the model's excellent discrimination capability between benign and malignant classes, approaching the maximum AUC value of 1.0. Compared to the AUC scores of other models, such as 0.983 for the cascaded CNN and 0.928 for EfficientNet-B0, the proposed model demonstrates highly competitive performance, despite the smaller dataset.

Table 2. Performance metrics of studies in the literature.

Study	Classification Type	CNN Model	Number of Images in Dataset	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1 Score (%)	AUC
Warin et al. (2022) [14]	Two-class	DenseNet-169	980		98% (OSCC) 95% (OPMD)	99% (OSCC) 95% (OPMD)	99% (OSCC) 97% (OPMD)	98% (OSCC) 95% (OPMD)	1.0 (OSCC) 0.98 (OPMD)
Jubair et al. (2020) [15]	Two-class	EfficientNet-B0	716	85.00%		86.70%	84.50%		0.93
Huang et al. (2023) [16]	Two-class	Unique Model	131	97.00%	92.66%	87.34%%		89.37%	
Fu et al. (2020) [17]	Two-class	Cascaded CNN Model	44409	92.40%		94.90%	88.70%		0.98
Bansal et al. (2023) [18]	Two-class	Unique Model	131	92.00%	100.00%	88.90%	100.00%	94.12%	
Tanriver et al. (2021) [20]	Multi-class	EfficientNet-b4	684		87.00%	86.00%		86.00%	
Lin et al. (2021) [19]	Multi-class	HRNet-W18	455		84.30%	83.00%	96.60%	83.60%	
Our study	Two-class	VGG16 based model	330	94.79%	95.11%	94.58%	94.58%	94.74%	0.96

While the model's performance metrics are promising, the relatively limited dataset size of 330 images may restrict generalizability. Expanding the dataset size, possibly incorporating a wider range of cases and image variations, could enhance the model's robustness and ensure its applicability across diverse populations.

Overall, the study confirms that AI-based image classification for oral diseases can achieve high accuracy and balance across key metrics, even with smaller datasets. The model's potential for clinical application is significant, as it provides accurate, efficient, and balanced diagnostic support, which could enhance early detection and patient outcomes.

V. CONCLUSION

In conclusion, this study demonstrates that deep learning models, particularly CNN architectures such as the modified VGG16, hold substantial potential for the classification of oral lesion images as benign or malignant. The proposed model, with a robust performance accuracy of 94.79%, precision of 95.11%, and sensitivity and specificity both at 94.58%, showcases its capability to effectively distinguish between different types of oral conditions. These results indicate that even with a dataset of moderate size, it is possible to develop a highly accurate model that can contribute significantly to early detection efforts in oral health diagnostics.

This study aimed to develop and evaluate a deep learning-based approach for the accurate classification of oral lesions using publicly available image datasets. By employing a modified VGG16 architecture and optimized preprocessing techniques, the research sought to improve diagnostic accuracy, sensitivity, and specificity compared to existing models. The overarching goal was to provide a reliable, efficient, and scalable tool that can assist healthcare professionals in early detection and diagnosis of oral malignancies, enhancing patient outcomes and facilitating timely treatment interventions.

The findings align well with existing literature, supporting the notion that AI-driven diagnostic tools can augment traditional methods and assist healthcare professionals by providing consistent, rapid, and reliable results. The high F1 score and AUC further underscore the model's balanced performance, indicating its readiness for potential integration into clinical workflows to support decision-making processes and improve patient care outcomes.

However, the study also recognizes that to achieve wider applicability and enhanced reliability, future research should focus on expanding dataset sizes and including more diverse and complex cases. This would aid in addressing

limitations related to generalizability and ensure that the model can be effectively employed in various clinical settings. In addition, incorporating real-time image acquisition techniques and optimizing computational efficiency could further enhance the practical deployment of the model.

In summary, this research provides a promising step forward in leveraging deep learning for the early detection and classification of oral malignancies. By bridging the gap between traditional diagnostic methods and modern AI capabilities, this study contributes to the broader effort of enhancing oral health management and ultimately improving patient outcomes.

ACKNOWLEDGMENT

Authors' Contributions

The authors' contributions to the paper are equal.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics.

REFERENCES

[1] S. Warnakulasuriya *et al.*, 'Oral Potentially Malignant Disorders: A Consensus Report From an International Seminar on Nomenclature and Classification, Convened by the WHO Collaborating Centre for Oral Cancer', *Oral Diseases*, vol. 27, no. 8, pp. 1862–1880, 2020, doi: 10.1111/odi.13704.

[2] S. Yang, Y. Lee, L. Chang, C. Yang, C. Luo, and P. Wu, 'Oral Tongue Leukoplakia: Analysis of Clinicopathological Characteristics, Treatment Outcomes, and Factors Related to Recurrence and Malignant Transformation', *Clinical Oral Investigations*, vol. 25, no. 6, pp. 4045–4058, 2021, doi: 10.1007/s00784-020-03735-1.

[3] C. B. More and N. R. Rao, 'Proposed Clinical Definition for Oral Submucous Fibrosis', *Journal of Oral Biology and Craniofacial Research*, vol. 9, no. 4, pp. 311–314, 2019, doi: 10.1016/j.jobcr.2019.06.016.

[4] S. Abati, C. Bramati, S. Bondi, A. Lissoni, and M. Trimarchi, 'Oral Cancer and Precancer: A Narrative Review on the Relevance of Early Diagnosis', *International Journal of Environmental Research and Public Health*, vol. 17, no. 24, p. 9160, 2020, doi: 10.3390/ijerph17249160.

[5] K. Matsuoka, 'Oral Malignant Melanoma Detected After Resection of Amelanotic Pulmonary Metastasis', *International Journal of Surgery Case Reports*, vol. 4, no. 12, pp. 1169–1172, 2013, doi: 10.1016/j.ijscr.2013.10.004.

[6] L. Cigic, 'Increased Prevalence of Oral Potentially Malignant Lesions Among Croatian War Invalids, a Cross-Sectional Study', *Journal of Clinical and Experimental Dentistry*, pp. e734-741, 2023, doi: 10.4317/jced.60715.

[7] F. M. Ghanaei, F. Joukar, M. Rabiei, A. Dadashzadeh, and A. K. Valeshabad, 'Prevalence of Oral Mucosal Lesions in an Adult Iranian Population', *Iranian Red Crescent Medical Journal*, vol. 15, no. 7, pp. 600–604, 2013, doi: 10.5812/ircmj.4608.

[8] A. M. Kavarodi, M. Thomas, and J. Kannampilly, 'Prevalence of Oral Pre-Malignant Lesions and Its Risk Factors in an Indian Subcontinent Low Income Migrant Group in Qatar', *Asian Pacific Journal of Cancer Prevention*, vol. 15, no. 10, pp. 4325–4329, 2014, doi: 10.7314/apjcp.2014.15.10.4325.

[9] G. Guan and N. Firth, 'Oral Manifestations as an Early Clinical Sign of Acute Myeloid Leukaemia: A Case Report', *Australian Dental Journal*, vol. 60, no. 1, pp. 123–127, 2015, doi: 10.1111/adj.12272.

[10] H. Mawardi *et al.*, 'Oral Epithelial Dysplasia and Squamous Cell Carcinoma Following Allogeneic Hematopoietic Stem Cell Transplantation: Clinical Presentation and Treatment Outcomes', *Bone Marrow Transplantation*, vol. 46, no. 6, pp. 884–891, 2011, doi: 10.1038/bmt.2011.77.

[11] N. Al-Rawi *et al.*, 'The Effectiveness of Artificial Intelligence in Detection of Oral Cancer', *International Dental Journal*, vol. 72, no. 4, pp. 436–447, Aug. 2022, doi: 10.1016/j.identj.2022.03.001.

[12] M. García-Pola, E. Pons-Fuster, C. Suárez-Fernández, J. Seoane-Romero, A. Romero-Méndez, and P. López-Jornet, 'Role of Artificial Intelligence in the Early Diagnosis of Oral Cancer. A Scoping Review', *Cancers*, vol. 13, no. 18, p. 4600, Sep. 2021, doi: 10.3390/cancers13184600.

[13] S. Nath, R. Raveendran, and S. Perumbure, 'Artificial Intelligence and Its Application in the Early Detection of Oral Cancers', *Clin Cancer Investig J*, vol. 11, no. 1, pp. 5–9, 2022, doi: 10.51847/h7wa0UHoIF.

[14] K. Warin, W. Limprasert, S. Suebnukarn, S. Jinaporntham, P. Jantana, and S. Vicharueang, 'AI-based analysis of oral lesions using novel deep convolutional neural networks for early detection of oral cancer', *PLoS ONE*, vol. 17, no. 8, p. e0273508, Aug. 2022, doi: 10.1371/journal.pone.0273508.

[15] F. Jubair, O. Al-karadsheh, D. Malamos, S. Al Mahdi, Y. Saad, and Y. Hassona, 'A novel lightweight deep convolutional neural network for early detection of oral cancer', *Oral Diseases*, vol. 28, no. 4, pp. 1123–1130, May 2022, doi: 10.1111/odi.13825.

[16] Q. Huang, H. Ding, and N. Razmjooy, 'Optimal deep learning neural network using ISSA for diagnosing the oral cancer', *Biomedical Signal Processing and Control*, vol. 84, p. 104749, Jul. 2023, doi: 10.1016/j.bspc.2023.104749.

[17] Q. Fu *et al.*, 'A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study', *EClinicalMedicine*, vol. 27, p. 100558, Oct. 2020, doi: 10.1016/j.eclinm.2020.100558.

[18] S. Bansal, R. S. Jadon, and S. K. Gupta, 'Lips and Tongue Cancer Classification Using Deep Learning Neural Network', in *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, India: IEEE, Mar. 2023, pp. 1–3. doi: 10.1109/ISCON57294.2023.10112158.

[19] H. Lin, H. Chen, L. Weng, J. Shao, and J. Lin, 'Automatic detection of oral cancer in smartphone-based images using deep learning for early diagnosis', *J. Biomed. Opt.*, vol. 26, no. 08, Aug. 2021, doi: 10.1117/1.JBO.26.8.086007.

[20] G. Tanriver, M. Soluk Tekkesin, and O. Ergen, 'Automated Detection and Classification of Oral Lesions Using Deep Learning to Detect Oral Potentially Malignant Disorders', *Cancers*, vol. 13, no. 11, p. 2766, 2021.

[21] M. Bai and M. Li, 'A Presentation of Structures and Applications of Convolutional Neural Networks', *Highlights in Science Engineering and Technology*, vol. 61, pp. 180–187, 2023, doi: 10.54097/hset.v61i.10291.

[22] K. O'Shea and R. R. Nash, 'An Introduction to Convolutional Neural Networks', 2015, doi: 10.48550/arxiv.1511.08458.

[23] N. Zakaria, 'Improved Image Classification Task Using Enhanced Visual Geometry Group of Convolution Neural Networks', *Joiv International Journal on Informatics*

Visualization, vol. 7, no. 4, p. 2498, 2023, doi: 10.30630/joiv.7.4.1752.

[24] D. Kwiatkowska, P. Kluska, and A. Reich, 'Convolutional Neural Networks for the Detection of Malignant Melanoma in Dermoscopy Images', *Advances in Dermatology and Allergology*, vol. 38, no. 3, pp. 412–420, 2021, doi: 10.5114/ada.2021.107927.

[25] Y. Tian, 'Artificial Intelligence Image Recognition Method Based on Convolutional Neural Network Algorithm', *Ieee Access*, vol. 8, pp. 125731–125744, 2020, doi: 10.1109/access.2020.3006097.

[26] K. Simonyan and A. Zisserman, 'Very Deep Convolutional Networks for Large-Scale Image Recognition', 2014, *arXiv*. doi: 10.48550/ARXIV.1409.1556.

[27] I. Fawwaz, T. Candra, D. A. M. Marpaung, A. Dinis, and M. R. Fachrozi, 'Classification of Beetle Type Using the Convolutional Neural Network Algorithm', *Sinkron*, vol. 7, no. 4, pp. 2340–2348, 2022, doi: 10.33395/sinkron.v7i4.11673.

[28] Akshitha and M. Veena, 'Melanoma Detection Using CNN', *International Research Journal of Modernization in Engineering Technology and Science*, 2023, doi: 10.56726/irjmets43733.

[29] H. Yılmaz, 'AI-Powered Healthcare Innovations: Rehabilitation, Education, And Early Diagnosis', Sep. 2024, *Serüven Yayınevi*. doi: 10.5281/ZENODO.13885904.

[30] M. Alehegn, 'Application of Machine Learning and Deep Learning for the Prediction of HIV/AIDS', *Hiv & Aids Review*, vol. 21, no. 1, pp. 17–23, 2022, doi: 10.5114/hivar.2022.112852.

[31] M. Sangeetha, 'Heart Disease Prediction Using ML', *International Journal of Innovative Science and Research Technology*, pp. 2630–2633, 2024, doi: 10.38124/ijisrt/ijisrt24mar2016.

[32] P. S. Mattas and I. Nadaan, 'Optimizing Cardiovascular Disease Diagnosis With Machine Learning: An Analysis', *International Journal of Research Publication and Reviews*, vol. 04, no. 02, pp. 430–434, 2023, doi: 10.55248/gengpi.2023.4217.

[33] C. S. Anita, P. Nagarajan, G. A. Sairam, P. Ganesh, and G. Deepakkumar, 'Fake Job Detection and Analysis Using Machine Learning and Deep Learning Algorithms', *Revista Gestão Inovação E Tecnologias*, vol. 11, no. 2, pp. 642–650, 2021, doi: 10.47059/revistageintec.v11i2.1701.

[34] U. Ramasamy and S. Santhoshkumar, 'Benchmark Datasets and Real-Time Autoimmune Disease Dataset Analysis Using Machine Learning Algorithms With Implementation, Analysis and Results', *Journal of Intelligent & Fuzzy Systems*, vol. 45, no. 2, pp. 2449–2463, 2023, doi: 10.3233/jifs-224115.

[35] H. T. Sihotang, M. K. Albert, F. Riandari, and L. A. Rendell, 'Efficient Optimization Algorithms for Various Machine Learning Tasks, Including Classification, Regression, and Clustering', *Idea*, vol. 1, no. 1, pp. 14–24, 2023, doi: 10.35335/idea.v1i1.3.

[36] S. A. Pane and F. M. Sihombing, 'Classification of Rock Mineral in Field X Based on Spectral Data (SWIR &Amp; TIR) Using Supervised Machine Learning Methods', *Iop Conference Series Earth and Environmental Science*, vol. 830, no. 1, p. 012042, 2021, doi: 10.1088/1755-1315/830/1/012042.

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 159-162
DOI: 10.36287/ijmsit.8.2.17
Received: December 13, 2024; Accepted: December 19, 2024

Efficient Time Allocation Strategies in Satellite Communication Networks

Peri Gunes [1*], Khadijeh Ali Mahmoodi [2] and Mert Ülkgün[3]

[1*] *Department of R&D, Infina Yazilim A.S., İstanbul, Türkiye (pgunes@infina.com.tr)*
[2]*ESYCOM Laboratory, Gustave Eiffel University, Paris, France (Khadijeh.alimahmoodi@esiee.fr)*
[3]*Department of R&D, Infina Yazilim A.S., İstanbul, Türkiye (mulkgun@infina.com.tr)*

Abstract – This paper introduces a novel time allocation algorithm designed for satellite communication networks, utilizing diverse communication links such as Radio Frequency (RF), Free Space Optical (FSO), and Long Range Wide Area Network (LoRaWAN). The proposed algorithm prioritizes gateway nodes based on key criteria, including priority levels and packet count, aiming to minimize delays and optimize overall throughput in satellite communication systems. LoRaWAN, known for its long-range communication capabilities, maximum device lifetime, multi-usage flexibility, bidirectional data transfer, low cost, and robust security, is particularly suited for such applications. Through a comprehensive study involving 200 gateway nodes, the algorithm's effectiveness in improving communication efficiency is demonstrated, offering a scalable solution for modern satellite networks.

Keywords – *Time allocation algorithm, satellite communication systems, communications efficiency.*

Citation: Gunes et.al., (2024). Efficient Time Allocation Strategies in Satellite Communication Networks. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 159-162.

I. INTRODUCTION

In an era marked by the relentless demand for seamless global connectivity, satellite communication has emerged as a cornerstone of modern telecommunications [1]. This technology enables coverage far beyond terrestrial networks, providing reliable communication across vast distances and making it indispensable for diverse applications, from global internet access to disaster recovery and remote sensing [2]. As the number of connected devices increases and data demands surge, optimizing satellite communication systems becomes critical to ensuring efficiency, reliability, and scalability.

Satellite communication systems have traditionally employed various multiplexing techniques to manage multiple users and optimize bandwidth usage. Among these, Time Division Multiple Access (TDMA) is a widely used method for improving system performance, allowing multiple earth stations to share satellite resources by dividing transmission time into distinct slots [3]. Early research demonstrated TDMA's potential to enhance channel efficiency, supporting various services over satellite links [4]. With the advent of emerging applications such as the Internet of Things (IoT) and 5G, there is mounting pressure to optimize TDMA-based satellite systems to handle higher traffic loads, reduce latency, and improve reliability.

Numerous studies have addressed these challenges by proposing solutions to enhance throughput, minimize latency, and improve resource management in satellite communication. For instance, recent work has reviewed state-of-the-art technologies, network architectures, and protocols designed to support direct-to-satellite (DtS) IoT applications, highlighting Long Range Wide Area Network (LoRaWAN) as a key enabler for satellite-based IoT in remote regions [3], [4]. Simulations have demonstrated the promise of combining Long Range (LoRa) technology with Low Earth Orbit (LEO) satellites for massive communications applications, showcasing the feasibility and reliability of packet reception in real-world trials [5], [6], [7].

Further advancements in satellite communication systems have focused on enhancing resource management techniques. One such method, dynamic TDMA (D-TDMA), adjusts time slots based on user demand and satellite capacity, optimizing bandwidth usage and reducing idle time [8]. Adaptive TDMA techniques have also been explored to balance resource allocation between multiple users while maintaining low-latency performance, particularly in satellite-ground communication systems [9]. These innovations are pivotal as the demand for satellite communication continues to grow with modern data-intensive applications.

Advances in coding and modulation techniques, such as Turbo Codes and Low-Density Parity-Check (LDPC) codes, have further improved satellite communication systems by increasing spectral efficiency and reducing bit error rates [10]. Additionally, the integration of satellite communication with emerging technologies like Software-Defined Networking (SDN) and Network Function Virtualization (NFV) has opened new pathways for dynamic resource allocation and network flexibility, particularly in multi-beam satellite systems [11], [12].

Building on this foundation, our paper delves into the intricacies of satellite communication systems, presenting a comprehensive overview of the communication process from initialization to reception. Our focus is on optimizing this process through the development and implementation of a prioritization algorithm for gateway nodes. This algorithm ranks gateway nodes based on priority levels and packet count, minimizing transmission delays and maximizing overall throughput in a satellite communication network.

To validate the effectiveness of the proposed algorithm, we conducted a detailed study involving 200 gateway nodes. Our methodology includes sorting nodes by priority and packet count, linking this prioritization process to fundamental steps in satellite communication. The results of our study highlight the efficiency gains achieved by the algorithm, revealing valuable patterns and insights for network management and resource optimization.

The paper is structured as follows: In Section II, we present the system model, providing details on the models for satellite communication networks. In Section III, we elaborate on the time allocation algorithm aimed at minimizing latency. Section IV presents numerical results along with a performance discussion. The paper concludes with Section V.

II. System Model

Satellite communication is a cornerstone of modern connectivity, enabling seamless data exchange across vast distances, as illustrated in Fig. 1. This section provides a detailed overview of the processes governing communication between ground gateways and satellites, covering each step from initialization to data reception. Additionally, we introduce a prioritization algorithm aimed at optimizing communication by intelligently sorting gateway nodes based on priority levels and packet counts, thereby reducing latency and improving overall system throughput.

The foundation of satellite communication lies in a precise initialization process. To achieve efficient frequency utilization, both Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA) techniques are employed [13]. FDMA divides the available frequency spectrum into channels, while TDMA allocates time slots within each channel to facilitate organized communication.

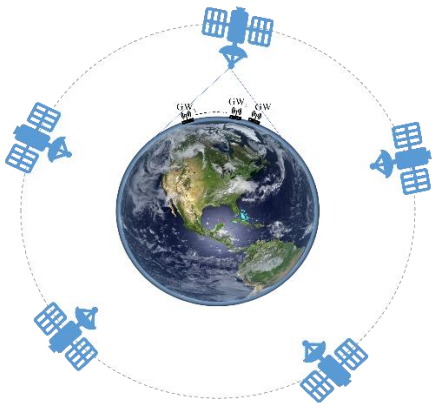


Fig. 1 Satellite Communication Network Under Consideration.

A. Key Components of the System Model:

- Gateway-Satellite Synchronization: Synchronization between ground gateways and satellites is critical for successful communication. This is typically achieved using Global Positioning System (GPS)-based synchronization methods, which ensure precise timing and coordination for data transfer between gateways and satellites.
- Satellite Wake-Up Signal (Beacon Transmission): As a satellite enters the range of a ground gateway, a wake-up signal (or beacon) is transmitted to notify the satellite of an upcoming data transmission. This signal acts as an alert for the satellite to prepare for communication.
- Satellite Wake-Up Confirmation: After receiving the wake-up signal, the satellite confirms its readiness to communicate by sending a Request to Send (RTS) signal, initiating the communication process between the satellite and the ground gateway.
- FDMA Channel Allocation: Efficient data transmission requires the allocation of suitable frequency channels. FDMA channels are assigned based on factors such as bandwidth availability, signal quality, and the communication system's operational requirements.
- TDMA Slot Allocation: To further optimize communication, specific time slots within the allocated FDMA channel are assigned for data transmission. This ensures an organized flow of data between the ground gateway and the satellite, minimizing the chances of interference and collisions.
- Data Encoding and Modulation: The data is then prepared for transmission by encoding and modulating it into a suitable format. Various modulation schemes, such as Binary Phase-Shift Keying (BPSK) or Quadrature Phase-Shift Keying (QPSK), are employed, along with error correction codes, to enhance data reliability during transmission.
- Signal Transmission: Once the data is encoded and modulated, it is transmitted via the assigned FDMA channel and TDMA time slot. Key considerations during this step include maintaining optimal power levels, ensuring adequate signal strength, and performing link budget calculations to achieve reliable communication.
- Reception at the Satellite: Upon reaching the satellite, the transmitted signals are received and demodulated using advanced techniques. The data is then decoded, and error correction mechanisms are applied if necessary, ensuring accurate data interpretation.
- Acknowledgment and Feedback: To confirm the successful reception of data, the satellite sends an acknowledgment signal back to the ground gateway. If errors are detected during reception, provisions for retransmission requests are in place to ensure data integrity.
- End of Transmission and Return to Idle State: Once the data transmission process is complete, both the ground gateway and the satellite return to an idle state, awaiting the next wake-up signal to initiate the next cycle of communication.
- Cyclic Nature of the Communication Process: This cyclical process of communication initialization,

transmission, reception, and acknowledgment continues in a loop. The prioritization algorithm introduced in this work adds an additional layer of optimization to this cycle by intelligently managing gateway node communication, thus enhancing the overall efficiency and throughput of the satellite network.

B. Time Allocation Algorithm:

Algorithm 1 describes a method for allocating time slots for communication from multiple gateways to a satellite, taking into account factors such as data transmission requirements and gateway priorities. The goal is to ensure an efficient allocation of time slots while meeting transmission needs. First, the gateways are sorted based on their data transmission requirements, and priority is given to those with higher transmission needs. Time slots are then allocated iteratively to gateways, starting with those with higher priority, and any remaining slots are distributed among lower-priority gateways. The result is a time slot allocation schedule for each service channel and slot, optimizing communication efficiency.

Algorithm 1: Time Slot Allocation for Communications from *N* Gateways to the Satellite

Inputs: Number of gateways, total time slots, service slots, gateway priorities, and data transmission

1. Initialize Time Slot Allocation;
2. Set the total number of time slots per satellite coverage slot;
3. Determine the number of service slots and time slots per service slot;
4. Initialize an empty time slot allocation schedule for each service channel and service slot;
5. Sort Gateways based on Data Transmission Requirements:
    a. Sort gateways in descending order based on payload data transmission needs;
6. Allocate Time Slots to Gateways:
    a. For each service channel and service slot do:
        i. Initialize available time slots to the total number per service slot;
        ii. Separate gateways with higher priorities and calculate required time slots;
        iii. Allocate initial time slots to higher-priority gateways;
        iv. Sort remaining gateways in ascending order;
        v. For each gateway do:
            1. Calculate required time slots based on data transmission needs;
            2. If required time slots are within available slots:
                a. Allocate time slots to the gateway;
                b. Subtract allocated time slots from available slots;
            3. If available slots become zero, break the loop and proceed to the next service slot;
        vi. If there are remaining unallocated gateways, distribute remaining available slots evenly among them;
7. Update the time slot allocation schedule for the current service channel and service slot;
8. Output Time Slot Allocation Schedule:
    a. Output the final time slot allocation schedule for each service channel and service slot.

Output: Time slot allocation schedule for each service channel and service slot, optimized based on the gateways' data transmission requirements and priorities.

III. NUMERICAL RESULTS

We consider 200 gateway nodes each with different priority levels and allocate different numbers of packets to each gateway. Fig.2 illustrates the time slot allocation schedule produced by the proposed algorithm. The algorithm's goal is to optimize time slot distribution across service channels and service slots based on each gateway's data transmission requirements and priority. The process begins by ranking gateways according to their payload transmission needs and assigning initial time slots to high-priority gateways. Remaining gateways are then sorted in ascending order of priority, and time slots are allocated accordingly, ensuring that higher-priority gateways receive sufficient resources to meet their transmission demands.

In this context, the figure visually represents three variables: gateway numbers (GW), the number of packets transmitted by each gateway, and their respective priority levels. Gateways with higher priority (depicted by the green dashed line) are allocated time slots first, ensuring their packets are transmitted promptly, as shown by the gradual stepwise increase in priority levels.

The result illustrates that higher-priority gateways, especially those with lower number of packets, receive earlier and more frequent time slot allocations. Gateways with lower priority have fewer time slots allocated, as indicated by the smaller number of packets and the overall decline in allocation toward the lower-priority gateways. This ensures that critical data transmission occurs with minimal delay, while also maximizing overall throughput and balancing the satellite network's resources.

The figure highlights the effectiveness of the algorithm in prioritizing gateways based on their packet count and priority levels, ensuring optimal use of available satellite time slots. By implementing this approach, the network improves both efficiency and scalability, essential for handling the growing demand for satellite communication in modern applications.

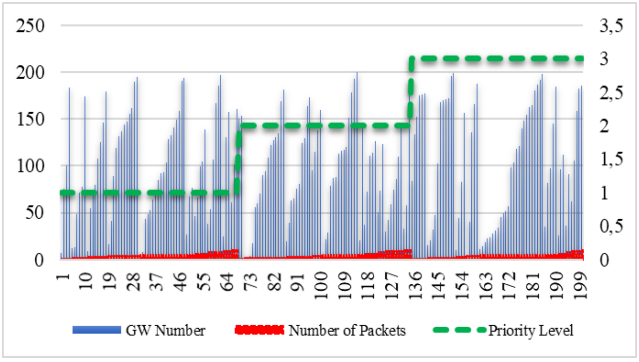


Fig. 2. Time Slot Allocation Schedule for Gateway Nodes

IV. CONCLUSION

In conclusion, this paper has presented a comprehensive approach to optimizing time slot allocation in satellite communication networks through the development of a prioritization algorithm. By systematically analyzing gateway data transmission requirements and prioritizing them based on defined criteria, the algorithm effectively enhances resource

allocation and minimizes latency. The results demonstrate that prioritization not only improves throughput but also ensures that critical data is transmitted promptly, addressing the growing demand for efficient satellite communication. Future work may explore further refinements to the algorithm and its application in diverse network environments to adapt to emerging technologies and increasing connectivity needs.

REFERENCES

[1] B. Yu, Y. Bao, Y. Huang, W. Zhan and P. Liu, "Modeling and Throughput Optimization of Multi-Gateway LoRaWAN," in IEEE Access, vol. 11, pp. 142940-142950, 2023, doi: 10.1109/ACCESS.2023.3343385.

[2] X. Luo, H. -H. Chen and Q. Guo, "LEO/VLEO Satellite Communications in 6G and Beyond Networks–Technologies, Applications, and Challenges," in IEEE Network, vol. 38, no. 5, pp. 273-285, Sept. 2024, doi: 10.1109/MNET.2024.3353806.

[3] J. A. Fraire, S. Céspedes and N. Accettura, "Direct-to-satellite IoT—A survey of the state of the art and future research perspectives: Backhauling the IoT through LEO satellites", *Proc. Ad-Hoc Mobile Wireless Netw. Conf.*, pp. 241-258, 2019.

[4] M. Centenaro, C. E. Costa, F. Granelli, C. Sacchi and L. Vangelist, "A survey on technologies standards and open challenges in satellite IoT", *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1693-1720, 3rd Quart. 2021.

[5] M. Asad Ullah, K. Mikhaylov and H. Alves, "Massive machine-type communication and satellite integration for remote areas", *IEEE Wireless Commun.*, vol. 28, no. 4, pp. 74-80, Aug. 2021.

[6] M. Asad Ullah, K. Mikhaylov and H. Alves, "Enabling mMTC in remote areas: LoRaWAN and LEO satellite integration for offshore wind farms monitoring", *IEEE Trans. Ind. Informat.*, Sep. 2021.

[7] M. A. Ullah, K. Mikhaylov and H. Alves, "Analysis and Simulation of LoRaWAN LR-FHSS for Direct-to-Satellite Scenario," in IEEE Wireless Communications Letters, vol. 11, no. 3, pp. 548-552, March 2022, doi: 10.1109/LWC.2021.3135984.

[8] B. Jabbari, "Cost-effective networking via digital satellite communications," in Proceedings of the IEEE, vol. 72, no. 11, pp. 1556-1563, Nov. 1984, doi: 10.1109/PROC.1984.13052.

[9] Y. Wang, M. Zeng and Z. Fei, "Efficient Resource Allocation for Beam-Hopping-Based Multi-Satellite Communication Systems," *Electronics*, vol. 12, no. 11, p. 2441, May 2023, doi: 10.3390/electronics12112441.

[10] P. Li, D. K. Borah, E. K. Tameh, and A. R. Nix, "Turbo Codes, LDPC Codes, and Polar Codes for Modern Satellite Communication Systems," *IEEE Communications Magazine*, vol. 50, no. 12, pp. 75-82, Dec. 2022, doi: 10.1109/MCOM.2022.6375934.

[11] Ferrus, R., Koumaras, H., Sallent, O., Agapiou, G., Rasheed, T., & Kourtis, M.-A. (2015). SDN/NFV-enabled satellite communications networks: Opportunities, scenarios and challenges. *Physical Communication*, 18, 95–112. doi: 10.1016/j.phycom.2015.08.007.

[12] Bertaux, L., Medjiah, S., Berthou, P., Abdellatif, S., Hakiri, A., & Gelard, P. (2015). Software-defined networking and virtualization for broadband satellite networks. *IEEE Communications Magazine*, 53(3), 54–60. doi: 10.1109/MCOM.2015.7060487.

[13] B. Jabbari, "Cost-effective networking via digital satellite communications," *Proc. IEEE*, vol. 72, no. 11, pp. 1556-1563, Nov. 1984. doi: 10.1109/PROC.1984.13052.

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 163-170
DOI: 10.36287/ijmsit.8.2.18
Received: November 8, 2024; Accepted: December 21, 2024

Word Frequency: New York Times Throughout the Times

Mehmet Aşıroğlu[1*] and Emre Atlıer Olca[2]

[1*]*Üsküdar American Academy, Istanbul, Turkey, (mehmet.asiroglu07@gmail.com) (ORCID: 0009-0006-1883-2245)*
[2]*Software Engineering Department, Maltepe University, Istanbul, Turkey (emreatlier@gmail.com) (ORCID: 0000-0001-6812-5166)*

Abstract – This paper investigates the evolution of the English language over the past century using a machine learning model trained on leading articles from The New York Times spanning from 1920 to 2020. The primary aim is to predict the year in which a given sentence could have been written based on linguistic patterns, including word usage and sentence structure. By analyzing these patterns, the model provides insights into the changing styles and trends in written English over time. The model's predictions are grounded in extensive data analysis and machine learning techniques, ensuring a high degree of accuracy. This study not only highlights the dynamic nature of language but also demonstrates the application of computational methods in linguistic research. The findings of this research are significant for historical linguistics and literature studies, as they provide a quantifiable method to track linguistic changes. Additionally, this work can aid in the development of tools for temporal text classification, benefiting fields such as digital humanities and archival studies. Understanding how language evolves is crucial for preserving cultural heritage and improving communication strategies in different communication platforms.

Keywords *– language evolution, machine learning, historical linguistics, text analysis, computational linguistics*

Citation: Aşıroğlu, M., Olca, E. (2024). Word Frequency: New York Times Throughout the Times, International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 163-170.

I. INTRODUCTION

Language is constantly changing, reflecting the cultural, social, and technological shifts that occur over time. As societies grow and evolve, so does the way they communicate. This evolution is particularly noticeable in written language, where changes in word choice, sentence structure, and writing style can be observed across different periods. Understanding these changes can offer valuable insights into historical events and broader societal trends, helping us better understand how communication develops alongside human progress.

This paper explores the evolution of the English language over the past century by analyzing leading articles from The New York Times, published between 1920 and 2020[7]. The New York Times, as a prominent publication, has documented key events and cultural shifts throughout the last hundred years, making it an ideal source for studying changes in language. By examining these articles, we aim to identify and measure shifts in writing style, word usage, and sentence structure, providing a detailed view of how journalistic language has changed over time.

The core of this research is a machine learning model designed to predict the publication year of a given sentence. This model was trained using a large dataset of New York Times articles, learning from patterns in word choice, sentence length, and sentence structure. By doing so, the model captures subtle changes in language that might otherwise go unnoticed, giving us a clearer picture of how language usage in journalism has evolved over the years.

The model works by first analyzing the input sentence, identifying important linguistic features, and then predicting the most likely year of publication based on the patterns it has learned. This predictive capability not only demonstrates the effectiveness of the model but also introduces a new way to study linguistic change. By using advanced computational methods, our research offers a fresh approach to examining how language evolves over time.

The findings from this study have implications beyond linguistics. They can be applied in literature studies, where understanding the historical context of language can enhance literary analysis. Additionally, in digital humanities, this research contributes to the development of tools that can classify texts by their time period, helping with the preservation and analysis of historical documents. Overall, this study provides a new perspective on the evolution of language, highlighting the close connection between language, society, and time.

The paper is structured into several sections, each detailing crucial aspects of the study on the evolution of language. Section 1 provides an introduction to the research, outlining the significance of studying language change and the methods used. Section 2 reviews related works, comparing and contrasting similar studies on language evolution and predictive modeling. Section 3 delves into the development of the linear SVC-trained language model, explaining the methodology behind predicting the origin date of a sentence based on linguistic features. Section 4 presents the results, supported by graphs, and discusses the findings in the context of language change over the past century. Finally, Section 5 concludes the paper by summarizing the key insights and discussing potential future research directions.

II. MATERIALS AND METHOD

Several studies have explored the evolution of written language and its stylistic changes over time. This section discusses six significant works that relate to our research, highlighting their methodologies and findings, and comparing them to our approach.

In the study "Modeling the Development of Written Language"(Wagner at all., 2011)[1], the authors used confirmatory factor analysis to test different models of written composition and handwriting fluency among first- and fourth-grade students. The study identified five key factors affecting written composition: macro-organization, productivity, complexity, spelling and punctuation, and handwriting fluency. The correlation between handwriting fluency and written composition factors was examined, revealing significant developmental differences between the two grade levels. While this study focuses on early developmental stages of writing skills, our research differs by analyzing a broader timespan of 100 years and emphasizing the evolution of language in published media. Moreover, our model uses machine learning techniques to predict the temporal origin of sentences based on linguistic patterns, rather than developmental differences in young writers.

The article "Change and Constancy in Linguistic Change: How Grammatical Usage in Written English Evolved in the Period 1931-1991"(Geoffrey and Nicholas, 2009)[2] examines the evolution of grammatical usage in British English through the Lanc-31 corpus, a trio of corpora spanning 1931, 1961, and 1991. By analyzing frequency counts of various grammatical features, the study identifies trends of increasing or decreasing usage, providing insights into grammaticalization, colloquialization, Americanization, and densification. This research closely aligns with our project, as both studies aim to trace linguistic changes over an extended period. However, our approach focuses on American English and employs machine learning to predict the publication year of sentences from The New York Times articles between 1920 and 2020. Additionally, while the Lanc-31 corpus provides a static analysis of grammatical features, our model dynamically predicts temporal origins based on a combination of word usage, sentence length, and syntactic patterns, offering a more comprehensive understanding of language evolution in journalistic writing.

Another relevant work is the research titled "Learning to Predict U.S. Policy Change Using New York Times Corpus with Pre-Trained Language Model."(Zhang at all., 2020)[3]. This study focuses on predicting policy changes in the United States by analyzing large-scale news data from The New York Times. The researchers built a comprehensive news corpus covering the period from 2006 to 2018 and fine-tuned the pre-trained BERT language model [9] to detect shifts in newspaper priorities, which they argue correspond to changes in U.S. policy. The study introduces a BERT-based Policy Change Index (BPCI)[15] to measure these changes, offering a novel approach to understanding and predicting policy shifts based on media analysis. This research closely relates to my project in that it also leverages machine learning and large-scale textual data from The New York Times. Both studies aim to uncover patterns and trends over time by analyzing language usage. However, while their focus is on predicting policy changes, my research is centered on examining the broader evolution of language in journalistic writing. Where their model seeks to identify specific policy shifts based on news priorities, the model that is presented in this study is designed to predict the publication year of sentences by analyzing linguistic features. This difference highlights how similar methodologies can be adapted to address distinct research questions, demonstrating the versatility and potential of machine learning in the study of language and its applications.

In the study titled "A Framework for Analyzing Semantic Change of Words Across Time"(Adam and Kevin, 2014)[4] the authors present a comprehensive approach to understanding how the meanings of words evolve over extended periods. The framework they propose utilizes word representations from distributional semantics to explore lexical changes at various levels, including individual word meaning, contrastive word pairs, and sentiment orientation. Their method allows for a detailed analysis of semantic transitions by leveraging large-scale diachronic corpora, which enables the visualization of a word's evolution over time. This research is particularly relevant for fields such as computational linguistics, historical linguistics, and natural language processing (NLP)[12], where understanding semantic change is crucial. The work relates to our project in its focus on language change over time, specifically through the lens of semantic evolution. Both studies utilize large datasets and computational methods to analyze linguistic trends. However, while their research is centered on the semantic shifts of individual words, our study takes a broader approach by examining changes in writing style, word usage, and sentence structure within journalistic writing over a century. The primary difference lies in the level of analysis: theirs is focused on word-level semantics, while ours considers sentence-level patterns and temporal trends. Additionally, their framework is designed to provide visual insights into word evolution, whereas our model aims to predict the publication year of sentences based on linguistic features. Despite these differences, both projects share a common goal of advancing our understanding of language change through computational means.

The research titled "Measuring News Sentiment"(Shapiro at all., 2017)[5] presents a novel approach to assessing economic sentiment by extracting it directly from newspaper articles rather than relying on traditional survey-based measures. The study introduces a news sentiment index developed using computational text analysis on a large corpus of economic and financial news articles. The researchers employ sentiment-scoring models, primarily utilizing lexical techniques, to analyze sentiment within these articles. By combining existing lexicons and creating a new lexicon tailored specifically for economic news, the study enhances the accuracy of sentiment prediction, achieving a rank correlation of approximately 0.5 with human ratings. The result is a national time-series measure of news sentiment, labeled the "News PMI model"[13] which correlates strongly with survey-based consumer sentiment indexes and is used to predict macroeconomic[16] outcomes. This work is relevant to our research as both studies leverage large datasets of news articles to analyze linguistic patterns over time. While their focus is on measuring sentiment in economic news and its impact on macroeconomic variables, our study examines broader linguistic changes, such as word usage and sentence structure, to predict the publication year of journalistic content. The primary similarity lies in the use of text analysis and machine learning techniques to extract meaningful information from large corpora. However, the key difference is that their project

is centered on sentiment analysis and its implications for economic forecasting, whereas our research is focused on tracking linguistic evolution and temporal shifts in language use within the context of news media. Despite these differences, both studies contribute to the growing body of work that uses computational methods to derive insights from textual data.

The research project titled "Understanding the Influence of News on Society Decision Making: Application to Economic Policy Uncertainty"(Trust at all., 2023)[6] focuses on the use of digital text data to analyze the impact of news on economic decision-making. With the rise of digital documentation, ranging from social media posts to news articles, the study explores how computational methods can be employed to understand the correlation between language usage and economic policy uncertainty (EPU). The project builds upon the Economic Policy Uncertainty (EPU)[14] index developed by Baker et al., which uses keyword-based methodologies to extract EPU-related news articles. However, this traditional approach is prone to false positives and negatives, prompting the need for more advanced techniques. To address these challenges, the authors propose a novel approach using weak supervision combined with neural language models, specifically BERT, for the automatic classification of news articles related to EPU. This method reduces the false positive rate significantly compared to traditional keyword-based approaches and is more efficient and cost-effective than fully supervised methods, which require extensive manual annotation. The study also introduces an Irish weak supervision-based EPU index and demonstrates its predictive power through econometric analysis with Irish macroeconomic indicators. This project shares similarities with our research in its use of computational techniques to analyze large-scale textual data for understanding broader social and economic phenomena. Both studies leverage machine learning models to process and categorize textual information, albeit with different goals. While their project focuses on extracting economic signals from news articles and predicting macroeconomic indicators, our study centers on tracking linguistic changes over time to predict the origin date of journalistic content. Both approaches highlight the importance of text analysis in deriving insights from digital documents, and their use of machine learning models to enhance the accuracy and efficiency of these analyses is directly relevant to our work.

An example of the table is given below.

Table 1. Content of different studies vs this study

Study/Project	Usage of Model	Usage of Dataset	Includes NLP	Tracks Change Over Time	Focuses on Word Usage	Sentiment Analysis	Predictive Analysis
1. Modeling the Development of Written Language	X	✓	X	✓	✓	X	X
2. Change and Constancy in Linguistic Change: How Grammatical Usage Evolved	✓	✓	✓	✓	✓	X	X
3. Learning to Predict U.S. Policy Change Using NYT Corpus	✓	✓	X	✓	✓	✓	✓
4. A Framework for Analyzing Semantic Change of Words Across Time	✓	✓	✓	✓	✓	X	X
5. Measuring News Sentiment	✓	✓	✓	X	X	✓	✓
6. Understanding the Influence of News on Society Decision-Making (EPU)	✓	✓	✓	X	X	✓	✓
Word Frequency: New York Times Throughout The Times	✓	✓	✓	✓	✓	X	✓

III. RESULTS

A. The Purpose of the project

The primary objective of this work is to build a language model capable of accurately classifying which decade a given sentence was likely written in. By doing so, we aim to establish a foundation for leveraging machine learning models in the analysis of language evolution over time. This project is driven by the need to understand how word usage, sentence structure, and overall linguistic trends change across different periods, offering insights into cultural, societal, and historical transformations reflected in written texts.

The contribution of this work lies not only in developing a functional model for decade classification but also in creating a scalable framework for future research. The model can be expanded to encompass broader linguistic features and larger datasets, eventually leading to more precise predictions and a deeper understanding of the mechanisms behind language change. Furthermore, this project could serve as a springboard for applications in digital humanities, historical linguistics, and automated text analysis, where tracking language shifts over time can provide valuable context for interpreting historical documents and understanding linguistic innovation.

We present the development of a Linear Support Vector Classifier (SVC)[10] model to predict the origin date of a sentence based on New York Times article titles[7] from 1920 to 2020. We leverage text data processing techniques, including TF-IDF (Term Frequency-Inverse Document Frequency[8]), to transform text into numerical vectors, which serve as input to our machine learning model.

TF-IDF (Term Frequency-Inverse Document Frequency) is a numerical statistic used to reflect the importance of a word in a document relative to a collection of documents or corpus. It is calculated by multiplying two components: Term Frequency (TF), which measures how often a word appears in a specific document, and Inverse Document Frequency (IDF), which gauges how common or rare the word is across all documents. The TF-IDF score increases with the number of times a word appears in a document while decreasing proportionally to the frequency of the word across the corpus. This approach highlights terms that are unique to each document, thus improving the model's ability to distinguish between different decades based on linguistic trends. TF-IDF is chosen in this study over other similar methods such as count vectorizer[17] or word embeddings[18] because it efficiently balances the significance of frequently occurring terms while downweighting common words that might not contribute to meaningful differentiation between decades. This results in a more refined feature set that enhances the performance of the Linear SVC model in predicting the temporal context of a sentence.

Our dataset is stored in Parquet format which is available on kaggle[8], which offers efficient storage and retrieval for large datasets, especially in structured or tabular data formats.

The process began with a clearly defined plan:

- load the dataset,
- preprocess the data to eliminate irrelevant entries,
- group the data by decades to reduce prediction complexity.

We intended to use a Linear SVC, a robust classification algorithm[19] well-suited for high-dimensional datasets like text data, and TF-IDF for feature extraction . By splitting the data into training and testing sets, we ensure model performance can be measured accurately. The code also allows users to input a new sentence and predict its likely decade.

B. Project Development Process

The development of this language model involved several iterations aimed at refining the model's performance and improving its ability to predict the decade in which a sentence was likely written. Initially, we began by training the model on a smaller dataset of New York Times article titles, which resulted in lower accuracy. The early trials used a random selection of fewer than 1,000 article titles per year. This yielded an accuracy of only 9%, largely due to the small training size and the lack of sufficient features for the model to generalize effectively.

We increased the sample size over time, eventually settling on 5,000 titles per year, as mentioned in the earlier code discussion. This improved the model's accuracy to 52%, indicating a more reliable relationship between the textual features of the article titles and their publication decade. The accuracy improvements were achieved by refining several steps, such as tweaking the TF-IDF vectorizer, using different random states, and optimizing the Linear SVC hyperparameters.

The code development in detail, from initial data processing to training the model and making predictions is explained as below.

Table 2. Getting data from New York Times

```
df = pd.read_parquet('/content/drive/MyDrive/
columbia-ml-data/nyt_data.parquet')
df = df.drop('excerpt', axis=1)
```

The code that is given in Table 2 starts by loading a dataset in Parquet format containing New York Times article titles from 1920 to 2020. The ‘drop('excerpt', axis=1)’ line removes the ‘excerpt’ column because it contains many empty values, which could confuse the model and negatively impact performance. By focusing on non-empty, relevant data (i.e., the article titles), the model is given cleaner, more consistent input.

Table 3. Grouping samples

```
sample=df.groupby('year').sample(n=5000,random_state
=84)
sample2 = sample['year'] // 10 * 10
sample2 = pd.concat([sample2, sample['title']], axis=1)
```

In table 3, the dataset is grouped by the year of publication, and a random sample of 5,000 titles per year is selected. The ‘random_state’ parameter ensures that the sampling process is reproducible. By dividing the years into decades (‘year // 10 * 10’), you simplify the prediction task: instead of predicting an exact year, the model predicts the decade in which a sentence might have been written. This reduces the complexity of the task and potentially increases the model’s accuracy.

Table 4. Labeling the axes

```
X = sample2['title']
y = sample2['year']
```

Table 4 shows the titles of the articles (‘X’) are used as the features, and the corresponding decades (‘y’) are used as the labels. These are the inputs and outputs that the model will learn to map during training.

Table 5. Splitting the data for train and test

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
```

In the code given in table 5, the dataset is split into training and testing subsets, with 80% of the data used for training the model and 20% reserved for testing its performance. The ‘random_state=42’ parameter ensures that the split is consistent each time the code is run, which is important for reproducibility.

Table 6. Using TF-IDF

```
tfidf = TfidfVectorizer()
X_train_vec = tfidf.fit_transform(X_train)
X_test_vec = tfidf.transform(X_test)
```

The TfidfVectorizer converts the text data into numerical vectors using the TF-IDF (Term Frequency-Inverse Document Frequency) method in table 6. This process transforms the words in the article titles into a form that the model can process. The ‘fit_transform’ method is applied to the training data, while ‘transform’ is used on the test data, ensuring that the same transformation is applied consistently across both

datasets.

Table 7. Choosing the model

```
model = LinearSVC()
model.fit(X_train_vec, y_train)
```

In table 7, a Linear Support Vector Classifier (LinearSVC) is instantiated and trained on the vectorized training data ('X_train_vec') and the corresponding labels ('y_train'). The SVC is a popular choice for text classification tasks due to its effectiveness in high-dimensional spaces, such as text data.

Table 8. Using predict

```
y_pred = model.predict(X_test_vec)
```

The trained model is used to predict the decades for the titles in the test dataset in table 8. These predictions ('y_pred') will be compared against the actual decades ('y_test') to evaluate the model's performance.

Table 9. Accuracy score

```
print(accuracy_score(y_test, y_pred))
```

Table 9 shows the 'accuracy_score' function calculates how often the model's predictions match the actual labels. This gives you a sense of how well the model is performing in terms of predicting the correct decade for unseen data.

Table 10. Testing the model

```
new_sentence = ["insert input here"]
new_sentence_vec = tfidf.transform(new_sentence)
predicted_year = model.predict(new_sentence_vec)
print(predicted_year)
```

This code in table 10 allows you to input a new sentence and predict the decade in which it might have been written. The input sentence is vectorized using the same TF-IDF model ('tfidf.transform(new_sentence)') and then passed through the trained SVC model to obtain the predicted decade ('predicted_year').

This code implements a linear SVC model trained on New York Times article titles to predict the decade in which a given sentence could have been written. The process includes loading and cleaning the data, sampling and grouping by decade, vectorizing the text data, training the model, and evaluating its accuracy. Additionally, the model can be used to predict the decade of new, unseen sentences.

C. *Accuracy Score and Its Limitations*

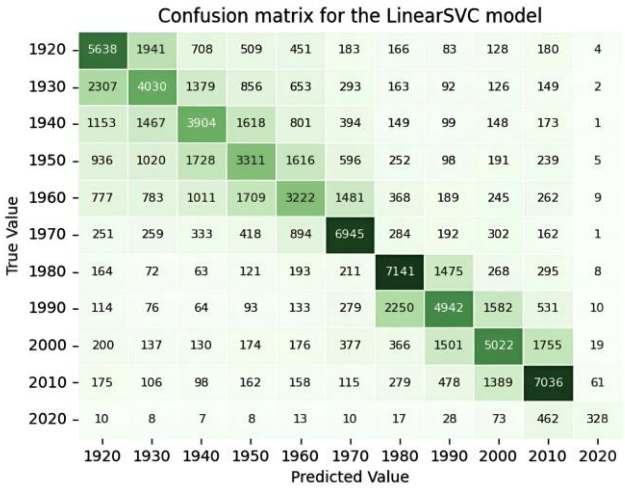
While an accuracy[20] of 52% is a significant improvement over the initial trials, it remains relatively low for real-world predictive applications. There are several reasons for this limitation. One major challenge lies in the overlapping nature of language use across decades. For example, while certain terms or phrases might be indicative of a specific era, many words and sentence structures remain constant or evolve slowly over time. This causes confusion in the model, especially for articles written in decades that are close to each other in time.

Another reason is the variability in the dataset. The New York Times articles cover a wide range of topics, from politics and economics to culture and science. Each topic may have its own unique linguistic features, and mixing them in a general model can dilute the predictive power. This variability contributes to the model's lower accuracy, as it must generalize over a vast array of subjects and styles.

IV. DISCUSSION

A. *Confusion Matrix Explanation*

Fig. 1. Confusion matrix for the model



The confusion matrix[11] shown in Figure 1 above provides a more detailed look at how well the model is performing across different decades. The matrix helps identify where the model tends to make incorrect predictions and gives insights into which decades are most easily confused with each other.

Diagonal values: These represent the correct predictions made by the model. For example, in the row corresponding to the 1920s, we see that 5,638 titles were correctly predicted to be from the 1920s. However, as we move down the diagonal, we notice that the number of correct predictions decreases for later decades, which is expected due to the increasing overlap in language use as we approach the present.

Off-diagonal values: These represent incorrect predictions, where the true value lies in one decade, but the model predicts another. For instance, the model frequently confuses the 1930s and 1940s, with 1,153 titles from the 1940s being incorrectly classified as from the 1930s. This indicates that the language in these two decades shares many similarities, making it

difficult for the model to distinguish between them.

Interestingly, the model performs much better for more recent decades, such as the 1980s and 2010s, where the diagonal values are much higher, indicating more accurate predictions. This suggests that the changes in language during these decades are more distinguishable from earlier periods, possibly due to the influence of modern technology and communication patterns.

B. Future Direction and Potential Applications

While the current model achieves moderate success, it sets the foundation for future work aimed at improving predictive accuracy and extending the analysis to other linguistic features, such as grammar patterns or word frequencies.

The long-term goal is to create a model that not only predicts the decade of a sentence but also identifies more granular linguistic shifts, such as the adoption of specific phrases or syntactic structures. Such a model could be useful in historical linguistic research, helping scholars track the evolution of language in various fields.

This study's model offers several promising applications in digital humanities and related fields. One notable contribution is its potential to aid archival research by serving as a temporal text classification tool for historians. By accurately predicting the decade a piece of text originates from, the model could help researchers organize large, unstructured corpora of historical documents, making it easier to identify patterns, trends, and shifts in language use over time. This would be particularly useful for analyzing underexplored periods or detecting chronological inconsistencies in archives.

In education, the model could be used to teach students about linguistic evolution and historical context. By analyzing texts from different decades, learners could explore how societal changes influenced language use, gaining insights into both history and linguistics. Additionally, this approach could enhance the development of historical language models, which often struggle to account for the nuances of older texts. Incorporating temporal context from models like this could improve their ability to process and understand historical documents, expanding their utility in fields like computational linguistics and cultural heritage preservation.

C. Example Predictions

To further illustrate how the model operates, let's consider a few sample predictions:

Input sentence: "The government enacted new legislation to regulate financial markets."

Predicted decade: 1980

Reasoning: The language model likely associates the terms "financial markets" and "regulate" with the economic policies and financial reforms of the 1980s.

Input sentence: "New technologies are shaping the future of communication."

Predicted decade: 2000

Reasoning: The prominence of "new technologies" and "communication" in this sentence aligns with the digital revolution and the rise of the internet, which began in earnest in the early 2000s.

These examples demonstrate the model's ability to link certain phrases and keywords with specific time periods. However, they also highlight some limitations. For example, if a sentence were written in a more ambiguous style, the model might struggle to provide an accurate prediction.

D. The Reasoning Behind TF-IDF

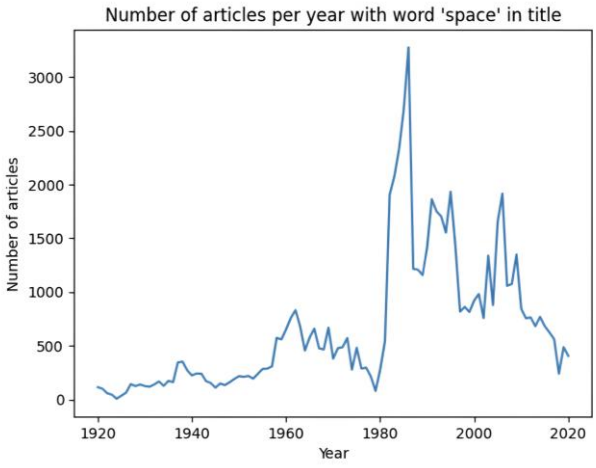
TF-IDF was selected as the feature extraction method for its simplicity, interpretability, and alignment with the study's objectives. Unlike advanced word embeddings such as Word2Vec or GloVe, which focus on capturing contextual semantics, TF-IDF emphasizes the relative importance of terms within the corpus. This makes it particularly suited for analyzing lexical and stylistic shifts over time, as it highlights changes in word usage patterns without introducing semantic complexities.

TF-IDF is computationally efficient, allowing the analysis of a large dataset like The New York Times corpus without excessive resource demands. While word embeddings could capture deeper relationships between words, they often introduce complexity and context that might obscure the stylistic and lexical trends this study seeks to investigate. Future research could include a comparative analysis of TF-IDF and embedding-based approaches to evaluate their respective impacts on capturing historical language trends and predicting temporal origins.

E. Cultural and Societal Factors

The model's accuracy variations across decades can be attributed to linguistic trends shaped by cultural and societal factors, which are often reflected in The New York Times' coverage of trending topics.

Fig. 2. Year by Number of Articles Graph



For example, as seen in figure 2, the sharp rise in the use of the word "space" during the 1960s and 1970s corresponds with the Space Race and the Apollo program, signaling a societal focus on space exploration. During these periods, the vocabulary, syntactic patterns, and frequency of related terms became more distinct, providing the model with stronger temporal markers and improving its predictive accuracy.

Conversely, in decades where trends are less sharply defined or topics overlap significantly with prior periods, the model may struggle to distinguish linguistic shifts, leading to reduced accuracy. For instance, the decline in articles about "space" after the 1980s reflects a cultural shift in priorities, resulting in less distinct linguistic signals for the model to learn from.

Since the corpus focuses on trending topics, it mirrors societal attention spans rather than steady linguistic evolution, which may introduce biases in the model. Fluctuations in accuracy highlight how the temporal coverage of events influences the distinctiveness of language features, reinforcing the importance of contextualizing model performance within cultural and historical dynamics.

F. Sampling Bias and Dataset Limitations

The dataset for this study was curated by selecting 5,000 article titles per year from the New York Times archive, creating a balanced representation of textual data across a century. While this sampling method ensured consistency in dataset size for each year, it may have introduced certain biases. The New York Times, as a major publication, tends to prioritize topics and writing styles that reflect its readership and editorial focus, potentially overrepresenting subjects of significant cultural, political, or economic importance at the expense of less mainstream topics.

This selection method also raises concerns about the diversity of linguistic styles captured in the dataset. Since the New York Times typically employs formal journalistic language, the model may have limited exposure to informal, regional, or genre-specific linguistic patterns that are also part of language evolution. Consequently, the model's predictions and insights may be skewed toward reflecting trends and styles unique to the New York Times rather than broader linguistic shifts.

Recognizing this limitation, future research could incorporate additional sources, such as regional newspapers or less formal publications, to create a more diverse and representative dataset. This would not only mitigate potential bias but also improve the generalizability of the model's findings to a wider range of linguistic contexts.

V. CONCLUSION

In conclusion, this study presents the development of a machine learning-based language model designed to predict the decade in which a given sentence from New York Times articles was written. We employed a Linear Support Vector Classifier (SVC) and used TF-IDF (Term Frequency-Inverse Document Frequency) to convert text into numerical vectors, focusing on analyzing language evolution over time. The core goal was to track shifts in word usage, sentence structures, and linguistic trends between 1920 and 2020, thereby offering insights into how language reflects cultural, societal, and historical changes.

Our methodology, which included data cleaning, decade-based grouping, and feature extraction using TF-IDF, allowed us to significantly improve the model's performance over time. By expanding the dataset to 5,000 titles per year and optimizing the model's parameters, we achieved an accuracy of 52%. Although this accuracy is moderate, it marks a substantial improvement from initial trials and demonstrates the potential of machine learning in understanding language change. TF-IDF was particularly effective in highlighting distinctive features that aided in decade classification, a decision that proved valuable in making the model more efficient.

Despite these advancements, several limitations persist. The overlap in language use across decades, especially between adjacent periods like the 1930s and 1940s, presented challenges for the model. Additionally, the wide range of topics covered by New York Times articles made it difficult for the model to generalize across different subjects. However, the model showed higher accuracy in more recent decades, suggesting that language changes were more discernible during periods of technological and communicative shifts.

Looking forward, future research should focus on incorporating additional linguistic features such as grammar patterns and word frequency analysis to improve predictive power. Expanding the dataset to include other text sources, like literature or social media, could offer a broader perspective on language evolution. Additionally, refining evaluation metrics beyond accuracy could help identify areas where the model struggles and provide more detailed insights into misclassifications.

While the current model is a foundational step, it opens the door to more advanced applications in digital humanities and historical linguistics. By tracking language shifts over time, future studies could offer deeper insights into how language adapts to societal changes, providing valuable tools for historians, linguists, and researchers in the field.

REFERENCES

[1] Wagner, Richard K., et al. "Modeling the development of written language." Reading and writing 24 (2011): 203-220.

[2] Leech, Geoffrey, and Nicholas Smith. "Change and constancy in linguistic change: How grammatical usage in written English evolved in the period 1931-1991." Corpus Linguistics. Brill, 2009.

[3] Zhang, Guoshuai, et al. "Learning to predict US policy change using New York Times corpus with pre-trained language model." Multimedia Tools and Applications 79 (2020): 34227-34240.

[4] Jatowt, Adam, and Kevin Duh. "A framework for analyzing semantic change of words across time." IEEE/ACM joint conference on digital libraries. IEEE, 2014.

[5] Shapiro, Adam Hale, Moritz Sudhof, and Daniel J. Wilson. "Measuring news sentiment." Journal of econometrics 228.2 (2022): 221-243.

[6] Trust, Paul, Ahmed Zahran, and Rosane Minghim. "Understanding the influence of news on society decision making: application to economic

policy uncertainty." Neural Computing and Applications 35.20 (2023): 14929-14945.
[7] https://www.kaggle.com/datasets/tumanovalexander/nyt-articles-data?resource=download Accessed 20 July 2024.
[8] Yun-tao, Zhang, Gong Ling, and Wang Yong-cheng. "An improved TF-IDF approach for text classification." Journal of Zhejiang University-Science A 6.1 (2005): 49-55.
[9] Sabharwal, Navin, et al. "Bert algorithms explained." Hands-on Question Answering Systems with BERT: Applications in Neural Networks and Natural Language Processing (2021): 65-95.
[10] https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html Accessed 6 Aug. 2024.
[11] Ohsaki, Miho, et al. "Confusion-matrix-based kernel logistic regression for imbalanced data classification." IEEE Transactions on Knowledge and Data Engineering 29.9 (2017): 1806-1819.
[12] Nadkarni, Prakash M., Lucila Ohno-Machado, and Wendy W. Chapman. "Natural language processing: an introduction." Journal of the American Medical Informatics Association 18.5 (2011): 544-551.
[13] Rao, Prahalad K., et al. "Process-machine interaction (PMI) modeling and monitoring of chemical mechanical planarization (CMP) process using wireless vibration sensors." IEEE Transactions on Semiconductor Manufacturing 27.1 (2013): 1-15.
[14] Yu, Jian, et al. "Economic policy uncertainty (EPU) and firm carbon emissions: evidence using a China provincial EPU index." Energy economics 94 (2021): 105071.
[15] Taylor, Joshua A., and Johanna L. Mathieu. "Index policies for demand response." IEEE Transactions on Power Systems 29.3 (2013): 1287-1295.
[16] Ramasamy, Ravindran, and Soroush Karimi Abar. "Influence of macroeconomic variables on exchange rates." Journal of economics, Business and Management 3.2 (2015): 276-281.
[17] Turki, Turki, and Sanjiban Sekhar Roy. "Novel hate speech detection using word cloud visualization and ensemble learning coupled with count vectorizer." Applied Sciences 12.13 (2022): 6611.
[18] Wadud, Md Anwar Hussen, M. F. Mridha, and Mohammad Motiur Rahman. "Word embedding methods for word representation in deep learning for natural language processing." Iraqi Journal of Science (2022): 1349-1361.
[19] Canty, Morton John. Image analysis, classification and change detection in remote sensing: with algorithms for Python. Crc Press, 2019.
[20] Li, Hongjian, et al. "The impact of protein structure and sequence similarity on the accuracy of machine-learning scoring functions for binding affinity prediction." Biomolecules 8.1 (2018): 12.
[21] Chen, Yuanyuan, Xuan Wang, and Xiaohui Du. "Diagnostic evaluation model of English learning based on machine learning." Journal of Intelligent & Fuzzy Systems 40.2 (2021): 2169-2179.
[22] Qi, Shi, et al. "An English teaching quality evaluation model based on Gaussian process machine learning." Expert Systems 39.6 (2022): e12861.
[23] Chang, Hui-Tzu, and Chia-Yu Lin. "Improving student learning performance in machine learning curricula: A comparative study of online problem-solving competitions in Chinese and English-medium instruction settings." Journal of Computer Assisted Learning (2024).
[24] Georgiou, Georgios P. "Comparison of the prediction accuracy of machine learning algorithms in crosslinguistic vowel classification." Scientific Reports 13.1 (2023): 15594.