INNOVATIVE APPROACHES IN APPLIED ENGINEERING FIELDS

EDITORS:
Prof. Dr. Nurettin DOĞAN
Assist. Prof. Dr. Selahattin ALAN
Assist. Prof. Dr. Züleyha YILMAZ ACAR

SELÇUK UNIVERSITY PRESS

INNOVATIVE APPROACHES IN APPLIED ENGINEERING FIELDS

Editors

Prof. Dr. Nurettin DOĞAN
Assist. Prof. Dr. Selahattin ALAN
Assist. Prof. Dr. Züleyha YILMAZ ACAR

✱ New Perspectives on Artificial Intelligence, Optimization, and Data Analytics in Computational Engineering

✱ Pioneering Perspectives in Modern Electrical and Electronics Engineering

✱ Innovative Studies in Advanced Materials, Communication Protocols, and Robotic Applications

✱ Cost Efficiency and Schedule Predictability in Construction Projects

SELÇUK
UNIVERSITY
PRESS

Innovative Approaches in Applied Engineering Fields

FOREWORD AND ACKNOWLEDGMENTS

Engineering sciences are continuously adapting to rapidly changing technological advancements and emerging challenges, striving to deliver solutions that enhance human well-being. In this context, the convergence of engineering disciplines contributes significantly to solving complex problems with innovative approaches. Innovative Approaches in Applied Engineering Fields aims to present the latest developments, theoretical frameworks, and practical applications in applied engineering fields. This comprehensive work, prepared by experts and researchers in the field, highlights the multidisciplinary nature of engineering.

This book encompasses a variety of studies categorized under four main sections: New Perspectives on Artificial Intelligence, Optimization, and Data Analytics in Computational Engineering explores applications of artificial intelligence and data analytics, while Pioneering Perspectives in Modern Electrical and Electronics Engineering provides innovative insights into modern electrical and electronics engineering. Additionally, Innovative Studies in Advanced Materials, Communication Protocols, and Robotic Applications presents groundbreaking research on advanced materials, communication protocols, and robotic applications, and Cost Efficiency and Schedule Predictability in Construction Projects focuses on cost efficiency and schedule management in construction projects.

In the first chapter, the integration of Opposition-Based Learning (OBL) into Particle Swarm Optimization (PSO) is examined to enhance metaheuristic performance. By evaluating both current and opposite solutions, OBL improves exploration, prevents premature convergence, and balances exploration and exploitation. This study, involving tests on 23 benchmark functions, demonstrates OBL's ability to enhance solution accuracy, convergence speed, and robustness. This chapter was prepared by B. S. DOĞAN and T. SAĞ.

In the second chapter, a novel hybrid algorithm, Cuckoo Search-Tuna Swarm Optimization (CSTSO), is introduced to address the limitations of Tuna Swarm Optimization (TSO), such as early convergence and local minima issues. By combining the fish swarm behavior of TSO with the global search capabilities of Cuckoo Search (CS), CSTSO demonstrates improved performance in solving complex optimization and engineering design problems. This chapter was prepared by N. ŞAHİN and S. SERVİ.

In the third chapter, an extensive analysis of mobility prediction mechanisms, models, and architectures for Location-based Services (LBS) is presented. The paper delves into various prediction models, including Bayesian Network (BN), Markov Chain (MC), Artificial Neural Networks (ANN), and Hidden Markov Model (HMM), which play key roles in improving forecast accuracy and optimizing network performance. By predicting the next client position, these models contribute significantly to resource management and LBS applications. This work, written by Shiva Derogar KHACHEKINI, Gürcü ÖZ, and Ali Hakan ULUSOY, paves the way for future research and the development of LBS-specific mobility prediction techniques.

In the fourth chapter, a detailed study and comparison of VANET routing protocols in Intelligent Transportation Systems (ITS) is provided. The paper investigates three types of topology-based routing protocols: Proactive, Reactive, and Hybrid models. It focuses on protocols like AODV, DSDV, and OLSR, analyzing their features, advantages, and limitations based on network performance metrics. This work, authored by Mojtaba Ayoubi MOBARHAN and Muhammaed

SALAMAH, offers valuable insights for optimizing network performance and lays the groundwork for future development of new routing protocols for VANETs in ITS applications.

In the fifth chapter, a deep learning model using DenseNet121 is applied to skin cancer detection with the HAM10000 dataset, achieving 93.4% accuracy, 100% sensitivity, and 99.39% specificity. Grad-CAM is utilized to improve transparency by visualizing the areas the model focuses on. This approach highlights AI's potential for early detection and better diagnosis of skin cancer. The chapter is written by Havva Hazel ARAS and Nurettin DOĞAN.

In the sixth chapter, the study focuses on improving synthetic data generation for financial applications by enhancing the effectiveness of Generative Adversarial Network (GAN)-based approaches. The authors propose a comprehensive evaluation framework to assess the quality of synthetic data based on fidelity, privacy, utility, outlier management, and feature relationship preservation. To improve the generation process, two preprocessing strategies are introduced: a scaling layer that selects the most suitable scaling technique for each feature and an outlier removal layer using the Interquartile Range (IQR) method. These methods are applied to financial datasets, demonstrating their ability to generate high-quality synthetic data. This chapter is authored by Ahmet Yasin AYTAR and Abdurrahman DEMİRLİ.

In the seventh chapter, the study explores the application of deep learning algorithms, specifically Convolutional Neural Networks (CNNs), for crack detection in Magnetic Particle Testing (MPT) images. MPT is widely used in the metals and materials industry to identify surface defects, but traditional methods rely on human interpretation, making it difficult to detect cracks accurately and quickly. The authors evaluate the performance of deep learning models in automating crack detection, using a dataset of cracked and non-cracked images. The study demonstrates that the CNN model achieved a high test accuracy of 83.12%. This work aims to improve the efficiency and accuracy of crack detection in industrial applications, minimizing human error. The chapter is authored by Yasin ULUS and Mehmet Akif ŞAHMAN.

In the eighth chapter, the authors review the development and applications of the Artificial Hummingbird Algorithm (AHA), a metaheuristic method introduced in 2022. This algorithm, inspired by the behavior of hummingbirds, uses iterative techniques to balance exploration and exploitation in solving complex problems. The review covers various studies on AHA, analyzing its performance, advantages, and limitations. AHA is widely applied in continuous function optimization, and its efficiency is enhanced when combined with other algorithms. The chapter is authored by E. ÖZEL and O. İNAN.

In the ninth chapter, the study investigates the use of time, frequency, and time-frequency domain feature sets for deep learning methods in predicting epileptic seizures. By extracting features from raw EEG data using techniques from these domains, the study evaluates their effectiveness in seizure prediction using deep learning algorithms. The findings highlight the contributions of these feature sets in improving prediction accuracy. This work was authored by A. DAŞDEMİR and H. K. ÖRNEK.

In the tenth chapter, the authors examine the use of alternative distributions-Rayleigh, Gamma, and Weibull-in wind speed forecasting. The study uses daily wind speed data from five regions in Turkey between 1998 and 2002 to forecast wind speeds for 2003. The forecasting performance of these distributions is compared using the Least Squares method and Mean Square Error (MSE) values. Results show that the Rayleigh distribution outperforms the

Weibull distribution in terms of error rate, highlighting its advantages for wind energy project planning and optimization. The chapter is authored by F. YILMAZ, A. PEKGÖR, İ. ARSEL, S. SERVİ, O. İNAN, A. GENÇ, and G. OTURANÇ.

In the eleventh chapter, the authors focus on time-frequency analysis of biomedical signals, particularly ECG and SpO2, in relation to sleep apnea. The study uses data from the PSG-Audio dataset, with ECG sampled at 200 Hz and SpO2 at 1 Hz. Statistical analysis reveals an average SpO2 of 94.64% and an ECG mean of 40.94 mV. Frequency domain analyses using FFT and STFT identify a stable SpO2 frequency of 0 Hz and an ECG frequency of 0.0125 Hz, which is linked to bradycardia during apnea episodes. The findings offer critical biomarkers for real-time sleep apnea monitoring, aiding in the development of automated diagnostic systems. The chapter is authored by G. DEMİRCİ and A. GÖLCÜK.

In the twelfth chapter, this study focuses on the development and use of a Medical Expert System (MES) for the diagnosis of nephritis. In collaboration with experts in the field, 128 rules were initially formulated using 7 different risk factors associated with the disease. These rules were then streamlined and simplified using Boolean Function Simplification methods, resulting in a final set of 6 rules that form the basis of the MES rule base. The effectiveness of MES was evaluated using a dataset of 120 patient cases and the results indicated a 100% success rate in nephritis diagnosis. This study demonstrates the potential of using advanced technologies such as expert systems in the medical field, especially for nephritis diagnosis. This study was prepared by F. BAŞÇİFTÇİ, K. A. GÜNDÜZ, and H. S. ÖZGÖKÇE.

In the thirteenth chapter, the performance issues and scheduling algorithms in Multi-path Transmission Control Protocol (MPTCP) systems are examined. The chapter analyzes schedulers such as Default, Blocking Estimation-Based, and Earliest Completion First, focusing on their role in addressing packet loss and path heterogeneity. This study was prepared by A. A. KHAJEKINI, A. H. ULUSOY, E. EVER, and M. A. MOBARHAN.

In the fourteenth chapter, a novel microwave sensor design based on directional couplers and complementary split ring resonators (CSRR) is presented for characterizing the dielectric properties of materials. Operating in the 1.5–3 GHz range, the sensor is optimized using CST Studio and evaluated for sensitivity through simulation tests on materials with varying dielectric constants. This chapter was prepared by İ. GENÇ, M. A. GÖZEL, and M. KAHRİMAN.

In the fifteenth chapter, the analysis of boost converter circuits and switching types is discussed, focusing on the comparison between 1-stage hard switching, 2-stage hard switching, and soft switching boost converter circuits. The research shows that 2-stage boost converter circuits are more efficient, as they experience lower losses compared to 1-stage circuits. Additionally, soft switching circuits offer more reasonable losses and provide a more stable system than hard switching circuits. This chapter, which contributes to optimizing circuit selection for electric vehicle charging stations, was prepared by S. M. KILINÇARSLAN and H. E. KOÇER.

In the sixteenth chapter, the study examines the use of Dual Tone Multi Frequency (DTMF) signaling over fixed telephone lines to remotely control electrical devices. It introduces a microcontroller-based encryption algorithm to address security vulnerabilities commonly associated with DTMF systems. The research highlights practical applications, such as centralized call-to-prayer systems in Turkey, and proposes an efficient method to enhance security in remote control systems. This chapter is authored by Şerafetdin BALOĞLU.

In the seventeenth chapter, Mehmet Fırat DOĞAN explores the role of polymeric membranes in PEM fuel cells, focusing on their importance for energy security, particularly in energy-dependent countries like Türkiye. The study investigates the mass losses and material changes of these membranes at different temperatures in various environments using DTA-TG experimental methods, aiming to address the challenges of fuel cell durability.

In the eighteenth chapter, the practical evaluation of VANET routing protocols in modern transportation systems is discussed, focusing on their role in improving vehicle communication, traffic efficiency, and road safety. The study compares protocols like AODV and OLSR, assessing network performance in parking systems and various traffic scenarios. This research provides insights into optimizing VANET performance in real-world applications. The chapter is authored by J. MOSENGO, M. M. MOBARHAN, A. A. KHAJEKINI, B. PONT, and K. TUZOLANA.

In the nineteenth chapter, an innovative modular end effector design for robotic DNA extraction is presented, addressing challenges in traditional isolation methods. The system, compatible with standard pipetting devices, utilizes magnetic bead technology and a customizable magnetic field to improve precision and efficiency. Experimental validation with *E. coli* DNA samples demonstrates its high precision and speed, with future potential for integration into advanced lab workflows. The chapter is authored by B. YILMAZ and V. FIRAT.

In the twentieth chapter, a spreadsheet application is developed to help small-scale construction companies manage costs and schedules using Earned Value Analysis (EVA). The tool calculates cost and schedule performance indexes based on entered data, providing an efficient way to control project progress. The application was successfully tested on a small project, offering practical benefits for contractors with limited resources. The chapter is authored by A. N. ŞENGÜL, N. DOĞAN, and Ö. H. BETTEMİR.

In the twenty-first chapter, the schedule variance of construction projects is assessed using Monte Carlo Simulation to account for the uncertainties arising from internal and external disruptions. By applying this method to a construction schedule prepared with the critical path method, the study estimates the potential risks to project duration. The developed software, written in C++, significantly improves the speed of analysis compared to traditional spreadsheet applications. This chapter is authored by M. KAYA, Ö. H. BETTEMİR, and M. F. ACAR.

We believe this book will serve as a valuable resource for researchers, students, and professionals working in the engineering field. It aims to inspire readers with new ideas and solutions, contributing to both scientific and applied studies. We extend our gratitude to all the authors for their invaluable contributions and hope this work will open new horizons in the field of engineering.

EDITORS

Prof. Dr. Nurettin DOĞAN
Assist. Prof. Dr. Selahattin ALAN
Assist. Prof. Dr. Züleyha YILMAZ ACAR

TABLE OF CONTENTS

NEW PERSPECTIVES ON ARTIFICIAL INTELLIGENCE, OPTIMIZATION, AND DATA ANALYTICS IN COMPUTATIONAL ENGINEERING

OPPOSITION-BASED LEARNING METHODS TO IMPROVE EXPLORATION IN METAHEURISTICS

Büşra Seval DOĞAN[1], Tahir SAĞ[2]

INTRODUCTION

The solution of complex optimization problems, particularly in real-world applications that involve high-dimensional and nonlinear challenges, presents significant computational demands. Traditional deterministic methods are often inadequate for such problems, prompting a shift toward metaheuristic algorithms [1]. Metaheuristic algorithms, inspired by natural phenomena, offer heuristic approaches for effectively exploring vast search spaces in optimization processes. However, these algorithms are prone to limitations, including susceptibility to local optima and premature convergence, especially in highly complex solution spaces [2].

Particle Swarm Optimization (PSO), a widely used metaheuristic method, exemplifies such approaches through its model of swarm behavior to iteratively converge on optimal solutions[3]. The simplicity and computational efficiency of PSO make it effective across various optimization problems; however, its performance can be hindered by early convergence and local optima entrapment [4]. To address these issues, alternative strategies such as Opposition-Based Learning (OBL) have been introduced. OBL represents a prominent enhancement to optimization algorithms, designed to expand the exploration capacity of the solution space [5].

OBL is a strategy that allows algorithms to explore solution spaces more comprehensively by evaluating the opposites of current solutions, thus balancing exploration and exploitation phases for broader search capability [6]. By considering both existing and opposite solutions, OBL enables coverage of a larger portion of the search space. This approach helps metaheuristic algorithms avoid local optima and reduces premature convergence risks. When integrated with PSO, OBL has been shown to significantly enhance [7]. In this context, the influence of OBL on PSO algorithms is further reinforced through hybrid approaches, such as Opp-PSOGWO, which not only accelerate the solution search but also enhance the ability to avoid local optima [8].The researchers analyze and compare five advanced PSO algorithms (CCPSO2, LSS-PSO, OBL-PSO, SPSO, and VCPSO) for solving large-scale optimization problems. The tests evaluate the efficiency and applicability of the algorithms, showing that CCPSO2 and OBL-PSO outperform the others, particularly for multimodal problems. These results highlight that OBL-PSO and CCPSO2 are competitive algorithms for solving large-scale and complex optimization challenges [9].This study proposes two improved PSO variants, ORIW-PSO-TF

[1] Selcuk University, Konya/Türkiye, Orcid: https://orcid.org/0000-0002-5969-2860, doganbusraseval@gmail.com
[2] Selcuk University, Konya/Türkiye, Orcid: https://orcid.org/0000-0001-8266-7148, tahirsag@selcuk.edu.tr

(Threefry-based) and ORIW-PSO-P (Philox-based), to address premature convergence. Key innovations include the use of pseudo-random sequences for population initialization, opposition-based learning to increase diversity, and a novel opposition-based rank inertia weight to accelerate convergence. The variants outperform conventional methods on 16 benchmark functions. Additionally, they achieve superior results in training artificial neural networks (ANN) using 15 UCI benchmark datasets, demonstrating their effectiveness in both optimization and machine learning applications[10].This paper addresses the task scheduling problem in distributed, heterogeneous, and multiprocessing environments like cloud computing. It proposes a Particle Swarm Optimization (PSO) approach with opposition-based learning (OBL) to minimize makespan and enhance task execution. The method, compared to existing strategies such as PSO, mPSO, GA, and others, aims to overcome premature convergence and improve efficiency in dynamic settings. Experimental results demonstrate that the proposed OBL-PSO scheduling method outperforms traditional approaches [11]. This paper introduces novel schemes to accelerate the convergence of Evolutionary Algorithms (EAs) by using opposition-based learning for population initialization and generation jumping. The methods are tested using Differential Evolution (DE) and are applicable to other population-based algorithms like Genetic Algorithms, Swarm Intelligence, and Ant Colonies. Experimental results with unimodal and multimodal benchmark functions show promising performance improvements, suggesting the effectiveness of the proposed approaches in enhancing algorithm efficiency [12]. This paper proposes an adaptive mutation-based opposition particle swarm optimization (AMOPSO) to address premature convergence in PSO. Building on the generalized opposition-based PSO (GOPSO), AMOPSO incorporates two strategies: (1) an adaptive mutation selection (AMS) to enhance exploration, and (2) an adaptive nonlinear inertia weight (ANIW) to balance exploration and exploitation. Experimental results on 14 benchmark functions show that AMOPSO outperforms or competes well with other opposition-based PSO variants, offering improvements in both convergence speed and solution quality [13].

This study aims to systematically investigate the effects of Opposition-Based Learning on PSO and to evaluate the performance of this hybrid approach through statistical analysis. To this end, the study examines the OBL-PSO integration's impact on solution accuracy, convergence speed, and overall algorithm resilience, utilizing 23 standard test functions, each executed independently 30 times. Results indicate that OBL enhances PSO's global search capabilities, improves solution quality, and reduces convergence time, aligning with existing literature that highlights the advantages of OBL in optimizing metaheuristic algorithms [6, 7].

Beyond focusing solely on PSO-OBL integration, this study also suggests that future research could further improve optimization performance by developing hybrid approaches that incorporate OBL with other metaheuristic algorithms. Thus, the advantages OBL offers in optimization processes present promising opportunities for tackling large-scale and complex problems more effectively.

The key contributions and motivations of the study can be summarized as follows:

- This research provides an in-depth examination of how the OBL strategy influences the performance of PSO algorithm, focusing on solution accuracy, convergence speed, and algorithm resilience.

- The performance of the OBL-PSO hybrid approach is statistically evaluated across 23 standard test functions, each conducted 30 times independently. Findings reveal that OBL enhances PSO's global search capacity, yielding higher-quality solutions and shortening convergence times.

- Results align with existing literature suggesting that OBL integration benefits metaheuristic algorithms, supporting the role of OBL as an effective strategy in the advancement of optimization algorithms.

- In addition to examining OBL's integration with PSO, this study highlights the potential for OBL to be incorporated into other metaheuristic algorithms in future research, potentially enhancing optimization performance for complex, large-scale problems.

The paper is organized to comprehensively address the research scope. The second section provides an overview of the OBL strategy and its theoretical foundation for improving metaheuristic algorithms, with detailed explanations of OBL techniques. The third section reviews the principles and operation of PSO algorithm. Subsequent sections present PSO's performance on 23 standard test functions, supported by experimental results, followed by a comparative evaluation of performance improvements achieved through OBL-PSO integration. The study concludes with a discussion of the results and recommendations for future research.

Opposition Based Learning

OBL is an innovative methodology designed to enhance the efficiency of optimization algorithms. By evaluating the opposites of existing solutions, OBL facilitates a broader exploration of the search space. This approach helps balance exploration and exploitation processes, thereby enabling algorithms to interact with a wider solution domain effectively [5, 6].

OBL stands out as a critical strategy for expanding the solution space. In traditional optimization methods, the risk of convergence to local optima is a significant challenge. By considering both existing solutions and their opposites, OBL enriches the search process with additional information, increasing the likelihood of identifying high-quality solutions. Integrating OBL into various metaheuristic algorithms has been shown to substantially enhance their overall performance [1].

Research in the literature highlights that OBL significantly improves the global search capabilities of algorithms, particularly in high-dimensional and complex optimization problems. Algorithms enhanced with OBL demonstrate faster convergence rates and improved solution quality, achieving notable success in optimization tasks [14].

Recognized as a pivotal strategy in optimization, OBL enables more effective exploration of the solution space. Future research holds considerable potential in further integrating this approach with a wider range of metaheuristic algorithms to enhance optimization performance.

The core mechanism of OBL involves calculating the opposite of a given solution, effectively expanding the search domain. By incorporating this concept, algorithms can escape local optima more effectively. The contributions of OBL to the optimization process are evident across various applications. These include foundational OBL strategies for symmetric solution

spaces, dynamic OBL methods for adapting to environmental changes in dynamic systems, multi-objective OBL techniques for achieving diverse solution sets, and applied OBL approaches in fields such as engineering, financial modeling, and machine learning. These methodologies collectively improve solution quality, accelerate optimization processes, and yield more balanced solution sets [1, 5, 15, 16].

A wide range of OBL variants has been introduced in the literature. These methods are generally categorized into two primary groups: (i) those that utilize a mapping function derived from decision variables, and (ii) those that rely on objective function values to identify solutions exhibiting opposing characteristics [17]. This study employs four types of OBL: Type-I, center-based sampling, generalized OBL, and quasi-reflection. The descriptions of each of these methods are provided below, in the following order.

Definition 1: Type-1

Indeed, this variant is regarded as the fundamental form of OBL and is mathematically defined in Eq.(1) [5].

$$\tilde{x}_i = lb_i + ub_i - x_i \tag{1}$$

where $x_i \in R$ is the i.th candidate solution of an optimization problem with n-dimension and $x_i \in [lb_i, ub_i]$.

Definition 2: Quasi-Opposite and Quasi-Reflection

Let x be an arbitrary real number within the interval $[a, b]$. Its opposite, denoted as x_0 , is defined as $x_0 = a + b - x$. And x be any real number within the interval $[a, b]$ [18]. The quasi-reflected point, x_{qr}, is then defined in Eq (3).

$$x_{qr} = rand(c, x) \tag{3}$$

where c represents the center of the interval $[a, b]$, which can be calculated as $(a + b)/2$. $rand(c, x)$ represents a random number uniformly distributed within the interval $[c, x]$.

Definition 3: Center-Based Sampling (CBS)

CBS aims to calculate an opposite solution $\tilde{x}$, consisting of parameters that are closer to the center of each variable [19]. It is defined as in Eq. (4).

$$\tilde{x}_i = rand_i(lb_i + ub_i - 2 * x_i) + x_i \tag{4}$$

where $randi$ is a uniformly distributed random number in the range of [0,1].

Definition 4: Generalized OBL(GOBL)

GOBL is an alternative form of OBL that seeks to obtain candidate solutions that are closer to the global optimum [20]. It is defined in Eq. (5).

$$\tilde{x}_i = k * (lb_i + ub_i) - x_i \tag{5}$$

where $k \in [0,1]$ is a random number.

PARTICLE SWARM OPTIMIZATION

PSO is a population-based optimization algorithm in which the position of each particle within the solution space is updated based on its current velocity and previous position. This update process considers both the individual best position (*pBest*) of each particle and the global best position (*gBest*) of the entire swarm[3, 4].

In PSO, the velocity of each particle is updated during each iteration according to the following equation.

$$v_i = \omega . v_{i-1} + c_1 . rand . (pBest_i - x_i) + c_2 . rand . (gBest - x_i) \qquad (7)$$

where, v_i represents the current velocity of the particle, ω is the inertia weight coefficient, c_1 and c_2 are the cognitive and social learning coefficients, **rand** denotes random values, $pBest_i$ is the particle's personal best position, and $qBest$ represents the global best position of the swarm.

This equation enables the particles to converge towards the optimal solution in the search space by balancing their velocities between the individual and collective components.

The position of the particles is updated by adding the velocity to their current position after the velocity update. This process is expressed as the sum of the current position x_i and the newly calculated velocity v_i for each particle. The formula $x_i = x_i + v_i$ determines the new position the particle will reach in the subsequent step. Here, x_i represents the particle's position from the previous iteration, and v_i denotes the velocity calculated for that iteration. This update allows the algorithm to navigate through the solution space and converge toward the optimal solution.

These equations assist the algorithm in maintaining a balance between exploration and exploitation. The inertia weight, ω, plays a crucial role in this balance; a larger ω value allows the particles to explore a broader search space, whereas a smaller ω value accelerates the convergence towards the optimal solution[21, 22].

PSO algorithm typically maintains a balance between exploration and exploitation throughout the search process. However, under certain specific conditions, the overall performance of PSO may be insufficient. Such conditions are commonly associated with phenomena such as low fitness values, stagnation, and poor exploration performance. OBL variants can intervene to address these issues, enabling PSO to search for solutions more effectively.

PSO uses the fitness value of each particle to determine the quality of the solution. If a particle's fitness value falls below a certain threshold, the particle can employ opposition-based strategies to move away from local optima. A low fitness value typically indicates that the particle is positioned in a worse region of the solution space. In such cases, the OBL method allows the particle to compute the opposite of its current solution, facilitating the search for a more efficient solution [5]. Solutions with low fitness values are generally found in regions distant from the global optimum, which can lead to the algorithm becoming trapped in local optimum.

In certain cases, PSO can become trapped within the solution space and converge to local optima. This situation typically arises when particles become positioned too close to one another, leading to a narrowing of the exploration process. Stagnation may cause the PSO

algorithm to fail in exploring a sufficiently wide area of the solution space. In such cases, OBL variants can assist the particles in exploring a broader solution space. By calculating the opposite of the current solution, OBL facilitates the rapid redirection of particles to new regions, effectively preventing stagnation and enabling the algorithm to progress more efficiently towards the optimal solution[3].

In PSO, when particles cluster too closely together and become confined to a narrow region of the solution space, the exploration performance weakens. This situation can lead to a loss of the algorithm's ability to discover new solutions. Weak exploration performance causes particles to concentrate in the same regions, resulting in difficulties in reaching the global optimum. OBL variants are highly effective in overcoming such challenges. OBL facilitates the exploration of a broader area within the solution space, thereby increasing the likelihood that PSO will find better solutions [1].

EXPERIMENTAL RESULTS

This study presents an experimental analysis aimed at evaluating the performance of PSO. In the experiments, the PSO algorithm and its various variants were applied to 23 standard test functions. The algorithm's performance under different conditions was systematically investigated based on the parameters defined for each test function.

The key parameters used in the study include a population of 50 particles, a maximum of 100 iterations, and 30 independent runs for each experiment. This configuration ensures the collection of reliable data regarding the algorithm's stability and overall performance. Furthermore, based on the results of each run, the best, worst, mean, and standard deviation values were calculated, enabling a comparative analysis of the PSO and its variants. The parameter settings are also provided in Table 1.

Table 1: Used Parameters.

Parameter	Value
Number of Particles	50
Number of Iterations	100
Number of Runs	30

The list and abbreviations of the variants used in this study are given in Table 2.

Table 2: Descriptions of Variants.

Variant Name	OBL Method Used
PSOv1	Type-1
PSOv2	Center-Based Sampling
PSOv3	Generalized OBL
PSOv4	Quasi-Reflection

OBL methods are integrated into PSO algorithm as shown in Pseudocode given in Figure 1. Each particle's current fitness is compared with its personal best score. If the current fitness is better, the personal best score and position are updated, and the fitness change counter, which tracks stagnation, is reset. If the current fitness does not improve, the counter is incremented to reflect the lack of progress. When the fitness change counter reaches or exceeds a predefined tolerance threshold, the OBL method is applied. This method generates a new position for the particle using an OBL-based function, which leverages opposition-based principles to explore the search space more effectively. By introducing this mechanism, the algorithm improves its ability to escape local optima and maintain diversity in the search process.

```
IF current fitness < personal best
    Update personal best score and position
    Reset the fitness change counter for the particle
ELSE
    Increment the fitness change counter for the particle
ENDIF
IF fitness change counter ≥ tolerance threshold THEN
    Apply OBL method:
        - Update particle position using:
          position = OBLFunction()
ENDIF
```

Figure 1: The Pseudocode of Integration Strategy of OBL methods in this study.

The statistical results given in Table 3 provide a comparative analysis of the performance of several variants of the PSO algorithm based on OBL methods. The statistical analysis presents the best value obtained, along with the mean, standard deviation (std), and the average runtime across 30 independent executions. Furthermore, the table includes the ranking of each algorithm for the respective test functions. The rank value allows for the comparison of algorithms relative to each other across test functions. This ranking, based on the consideration of mean values, enables a comparative evaluation of the performance of the algorithms. A high rank value indicates that the algorithm demonstrates strong overall performance[23]. Moreover, to illustrate the effectiveness of the proposed algorithms, convergence plots for all 23 functions are presented from Figs. 2 to 24. These figures allow for a comparative view of the performance of all the algorithms on the test problem.

In this study, the performances of various PSO and OBL variants have been systematically investigated across a range of test functions. The results obtained demonstrate how the PSO and OBL techniques perform differently across the tested functions.

The original PSO algorithm generally yields satisfactory results on certain benchmark functions, particularly those that are simple and unimodal. However, its performance tends to be relatively weaker on more complex test functions, especially when compared to OBL

variants. This disparity is particularly pronounced in the case of multimodal and high-dimensional functions. The benchmark functions on which OBL outperforms PSO are as follows: F6, F7, F12, and F13.

Table 3: Statistical results of classical benchmark functions

		PSO	PSOv1	PSOv2	PSOv3	PSOv4
F1	best	3.61e+02	4.01e+02	4.62e+02	6.44e+02	7.57e+02
	worst	5.31e+03	1.16e+04	1.13e+04	1.20e+04	1.13e+04
	mean	1.92e+03	2.41e+03	2.30e+03	2.53e+03	2.27e+03
	std	9.56e+02	2.15e+03	1.96e+03	2.66e+03	1.92e+03
	rank	1	5	3	2	4
F2	best	2.27e+01	1.71e+01	1.85e+01	2.09e+01	1.88e+01
	worst	8.64e+01	8.70e+01	8.52e+01	7.87e+01	7.69e+01
	mean	4.85e+01	5.11e+01	5.07e+01	4.59e+01	5.04e+01
	std	1.33e+01	1.70e+01	1.82e+01	1.38e+01	1.21e+01
	rank	4	1	5	3	2
F3	best	1.51e+04	1.98e+04	1.50e+04	2.38e+04	2.50e+04
	worst	5.78e+04	5.07e+04	5.12e+04	4.38e+04	5.37e+04
	mean	3.60e+04	3.43e+04	3.57e+04	3.26e+04	3.57e+04
	std	9.47e+03	6.67e+03	8.17e+03	6.20e+03	6.74e+03
	rank	4	2	5	3	1
F4	best	3.99e+01	4.66e+01	4.60e+01	3.95e+01	4.12e+01
	worst	6.81e+01	7.81e+01	7.32e+01	7.20e+01	7.72e+01
	mean	5.67e+01	5.96e+01	5.71e+01	5.46e+01	5.58e+01
	std	7.42e+00	7.04e+00	7.92e+00	8.85e+00	7.25e+00
	rank	4	5	1	3	2
F5	best	2.03e+05	2.14e+05	1.64e+05	1.18e+05	2.19e+05
	worst	3.97e+06	2.28e+06	4.84e+06	8.52e+06	7.91e+06
	mean	1.35e+06	8.75e+05	1.55e+06	1.44e+06	1.34e+06
	std	9.81e+05	5.60e+05	1.29e+06	1.58e+06	1.41e+06
	rank	2	5	1	4	3
F6	best	4.51e+02	4.05e+02	6.17e+02	5.22e+02	5.08e+02
	worst	1.33e+04	6.27e+03	1.28e+04	1.07e+04	3.99e+03
	mean	2.94e+03	2.16e+03	2.99e+03	2.56e+03	1.71e+03
	std	3.33e+03	1.27e+03	3.23e+03	2.50e+03	7.45e+02
	rank	5	2	4	1	3
F7	best	4.16e-01	4.76e-01	4.80e-01	4.35e-01	3.35e-01
	worst	1.15e+01	1.46e+01	9.00e+00	7.54e+00	1.71e+01
	mean	2.36e+00	2.63e+00	2.42e+00	2.26e+00	2.71e+00
	std	2.24e+00	2.88e+00	2.03e+00	2.11e+00	3.81e+00
	rank	4	1	3	2	5
F8	best	-1.00e+04	-1.01e+04	-1.01e+04	-1.01e+04	-1.01e+04
	worst	-7.74e+03	-7.83e+03	-7.05e+03	-6.90e+03	-7.16e+03
	mean	-8.80e+03	-8.70e+03	-8.62e+03	-8.55e+03	-8.63e+03
	std	5.38e+02	4.93e+02	6.74e+02	6.88e+02	6.94e+02
	rank	1	2	5	3	4
F9	best	1.41e+02	1.50e+02	1.45e+02	1.61e+02	1.25e+02
	worst	2.84e+02	3.07e+02	2.82e+02	3.02e+02	2.99e+02
	mean	2.21e+02	2.32e+02	2.16e+02	2.22e+02	2.21e+02
	std	2.92e+01	3.78e+01	3.29e+01	3.84e+01	3.84e+01
	rank	3	5	1	4	2
F10	best	8.21e+00	8.28e+00	6.92e+00	7.18e+00	6.44e+00

	worst	2.04e+01	1.76e+01	1.66e+01	2.04e+01	1.68e+01
	mean	1.19e+01	1.16e+01	1.14e+01	1.07e+01	1.10e+01
	std	3.45e+00	2.42e+00	2.84e+00	2.95e+00	2.50e+00
	rank	4	5	3	2	1
	best	7.50e+00	5.41e+00	6.35e+00	8.74e+00	8.29e+00
	worst	3.83e+01	4.04e+01	1.00e+02	2.84e+01	3.12e+01
F11	mean	1.88e+01	1.94e+01	2.41e+01	1.83e+01	1.79e+01
	std	7.54e+00	9.90e+00	2.16e+01	5.90e+00	7.24e+00
	rank	5	4	1	2	3
	best	8.62e+01	5.97e+03	2.49e+03	9.20e+01	5.84e+01
	worst	1.68e+06	2.37e+06	3.05e+06	2.45e+06	8.75e+06
F12	mean	3.59e+05	5.06e+05	5.67e+05	3.25e+05	7.29e+05
	std	4.88e+05	6.48e+05	7.26e+05	5.87e+05	1.64e+06
	rank	4	1	2	3	5
	best	5.27e+04	2.18e+05	2.13e+05	1.99e+05	6.13e+05
	worst	2.39e+07	2.24e+07	9.63e+06	8.13e+06	1.41e+07
F13	mean	2.96e+06	2.99e+06	2.03e+06	2.77e+06	2.38e+06
	std	5.15e+06	3.98e+06	1.90e+06	2.05e+06	2.58e+06
	rank	3	5	4	1	2
	best	9.98e-01	9.98e-01	9.98e-01	9.98e-01	9.98e-01
	worst	9.98e-01	9.98e-01	9.98e-01	9.98e-01	9.98e-01
F14	mean	9.98e-01	9.98e-01	9.98e-01	9.98e-01	9.98e-01
	std	1.33e-12	1.58e-12	5.15e-13	1.29e-13	1.90e-12
	rank	4	3	5	2	1
	best	5.35e-04	4.26e-04	4.78e-04	4.33e-04	5.36e-04
	worst	2.04e-02	2.04e-02	2.04e-02	2.04e-02	2.04e-02
F15	mean	5.47e-03	2.88e-03	5.25e-03	2.55e-03	4.99e-03
	std	8.36e-03	5.93e-03	7.92e-03	5.03e-03	7.83e-03
	rank	4	2	5	3	1
	best	-1.03e+00	-1.03e+00	-1.03e+00	-1.03e+00	-1.03e+00
	worst	-1.03e+00	-1.03e+00	-1.03e+00	-1.03e+00	-1.03e+00
F16	mean	-1.03e+00	-1.03e+00	-1.03e+00	-1.03e+00	-1.03e+00
	std	8.94e-16	9.83e-16	1.31e-15	1.93e-15	3.75e-15
	rank	1	2	3	4	5
	best	5.56e+01	5.56e+01	5.56e+01	5.56e+01	5.56e+01
	worst	5.56e+01	5.56e+01	5.56e+01	5.56e+01	5.56e+01
F17	mean	5.56e+01	5.56e+01	5.56e+01	5.56e+01	5.56e+01
	std	2.89e-14	2.89e-14	2.89e-14	2.89e-14	2.89e-14
	rank	1	2	3	4	5
	best	3.00e+00	3.00e+00	3.00e+00	3.00e+00	3.00e+00
	worst	3.00e+00	3.00e+00	3.00e+00	3.00e+00	3.00e+00
F18	mean	3.00e+00	3.00e+00	3.00e+00	3.00e+00	3.00e+00
	std	1.29e-14	1.64e-14	2.50e-14	1.85e-14	1.77e-14
	rank	1	5	2	4	3
	best	-3.86e+00	-3.86e+00	-3.86e+00	-3.86e+00	-3.86e+00
	worst	-3.86e+00	-3.85e+00	-3.86e+00	-3.86e+00	-3.85e+00
F19	mean	-3.86e+00	-3.86e+00	-3.86e+00	-3.86e+00	-3.86e+00
	std	1.57e-14	1.44e-03	1.74e-11	5.20e-14	1.44e-03
	rank	1	4	3	2	5
	best	-3.32e+00	-3.32e+00	-3.32e+00	-3.32e+00	-3.32e+00
	worst	-3.14e+00	-3.14e+00	-3.08e+00	-3.13e+00	-3.14e+00
F20	mean	-3.29e+00	-3.26e+00	-3.25e+00	-3.28e+00	-3.26e+00
	std	6.09e-02	7.28e-02	7.37e-02	6.30e-02	7.01e-02

		1	4	5	2	3
	rank	1	4	5	2	3
	best	-1.02e+01	-1.02e+01	-1.02e+01	-1.02e+01	-1.02e+01
	worst	-2.63e+00	-2.63e+00	-2.63e+00	-2.63e+00	-2.63e+00
F21	mean	-6.63e+00	-5.98e+00	-5.80e+00	-5.22e+00	-6.31e+00
	std	3.06e+00	3.37e+00	3.08e+00	2.97e+00	3.33e+00
	rank	1	5	2	3	4
	best	-1.04e+01	-1.04e+01	-1.04e+01	-1.04e+01	-1.04e+01
	worst	-1.84e+00	-1.84e+00	-1.84e+00	-1.84e+00	-2.75e+00
F22	mean	-5.61e+00	-5.40e+00	-5.57e+00	-6.03e+00	-6.75e+00
	std	3.53e+00	3.25e+00	3.50e+00	3.54e+00	3.53e+00
	rank	5	4	1	3	2
	best	-1.05e+01	-1.05e+01	-1.05e+01	-1.05e+01	-1.05e+01
	worst	-1.68e+00	-2.42e+00	-1.86e+00	-1.68e+00	-2.42e+00
F23	mean	-5.80e+00	-7.26e+00	-7.00e+00	-5.75e+00	-7.61e+00
	std	3.52e+00	3.55e+00	3.74e+00	3.77e+00	3.62e+00
	rank	5	2	3	1	4
Mean Rank		2.86	3.27	3.14	**2.68**	3.05

OBL variants exhibit superior performance relative to PSO, particularly with respect to their ability to escape local minima. This advantage is especially pronounced in functions such as F7 and F12. This benefit arises from OBL's enhanced ability to explore the solution space more efficiently and avoid becoming trapped in local minima.

Considering the standard deviation values of the variants, it is evident that OBL algorithms produce more stable and consistent solutions. The basic form of PSO often exhibits higher variance, leading to a loss of solution consistency. This is particularly noticeable in functions such as F1 and F5. The lower standard deviation values of OBL suggest that it yields more stable results under the same parameters.

OBL variants have managed the exploration-exploitation balance more effectively than PSO. While OBL conducts a broad exploration of the search space during the initial iterations, it achieves better convergence in later iterations, resulting in higher-quality solutions. In contrast, PSO occasionally disrupts this balance and prematurely shifts towards exploitation. This characteristic of OBL has been particularly evident in complex functions such as F6 and F12.

Overall, OBL variants have demonstrated superior performance across most test functions compared to PSO, particularly in their ability to avoid local minima and efficiently explore the solution space in multimodal and complex problems. Upon analyzing the rank values, OBL variants generally exhibit lower average ranks across most functions, suggesting that they tend to achieve superior results and demonstrate greater consistency in performance. While PSO yields competitive results in certain simpler test functions (such as F1), OBL's enhanced ability to balance exploration and exploitation enables it to outperform PSO in more complex functions.

Unimodal functions typically evaluate the exploitation capacity of algorithms, as they represent problems with a smooth and distinct optimum, where local minima are sparse. The basic form of PSO has yielded strong results on such functions, particularly those with low complexity, such as F1. However, in functions like F2 and F5, OBL has demonstrated superior performance.

OBL variants have achieved faster convergence on these functions and resulted in lower mean and standard deviation values compared to PSO.

Multimodal functions are characterized by the presence of numerous local minima, making the exploration capacity of the algorithm critical. In these problems, OBL has generally shown superior performance. For functions such as F6 and F12, the basic form of PSO often becomes trapped in local minima, whereas OBL algorithms are observed to explore more solution points, thereby improving the overall quality of optimization. Specifically, the success of OBL can be attributed to its ability to generate more diverse initial populations and to maintain greater diversity throughout the solution search process.

Fixed-shape multimodal functions are problems that not only require optimization algorithms to effectively address local minima, but also demand solutions that are close to the global optimum. In this context, OBL variants have demonstrated lower mean and standard deviation values compared to PSO. By more efficiently exploring the solution space, OBL has provided solutions with reduced error margins and enhanced stability. Despite the inherent complexity of these functions, OBL algorithms can be seen to improve solution quality by fostering greater diversity.

Overall, it can be concluded that OBL variants demonstrate superior performance compared to PSO, particularly in complex, multimodal test functions, where they offer a distinct advantage. The success of OBL algorithms is particularly evident in their ability to escape local minima and effectively balance exploration and exploitation. However, for unimodal and simpler functions, the basic form of PSO has also been capable of yielding competitive results. While OBL achieves better outcomes by incorporating diversity and flexibility into the solution search process, the simpler structure of PSO has proven advantageous in certain test functions.

In conclusion, the integration of OBL methods with PSO has generally been successful in enhancing optimization performance. Nevertheless, it can be asserted that both algorithms offer distinct advantages depending on the problem type, and the choice between them should be made based on the specific application requirements. While OBL tends to provide more effective results for complex problems, PSO can deliver competitive outcomes for simpler, low-complexity problems.

OBL-based methods have proven to be effective in addressing the limitations of PSO and enhancing its overall optimization performance. Specifically, variants such as Generalized OBL and Center-Based Sampling OBL exhibit strong capabilities in terms of global minimum achievement and convergence speed. Conversely, Type1 OBL demonstrates more limited success, while Quasi-Reflection OBL delivers rapid improvements but loses its effectiveness in subsequent iterations.

These results suggest that the integration of OBL methods into PSO offers an effective strategy for achieving improved outcomes, particularly in more complex optimization problems. However, it is important to note that the effectiveness of each variant may vary depending on the specific use case scenario.

Figure 2: Convergence plot for the F1 function

Figure 3: Convergence plot for the F2 function

Figure 4: Convergence plot for the F3 function
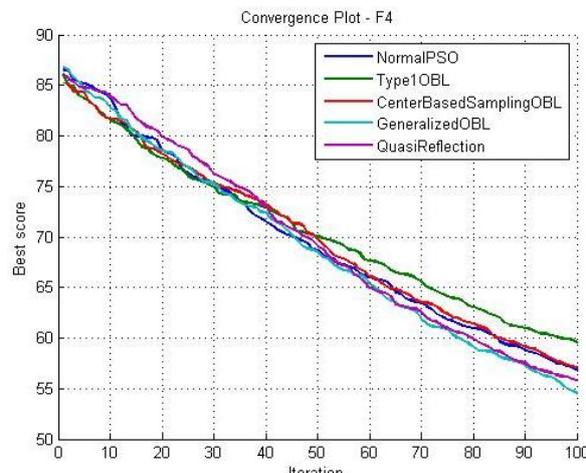
Figure 5: Convergence plot for the F4 function
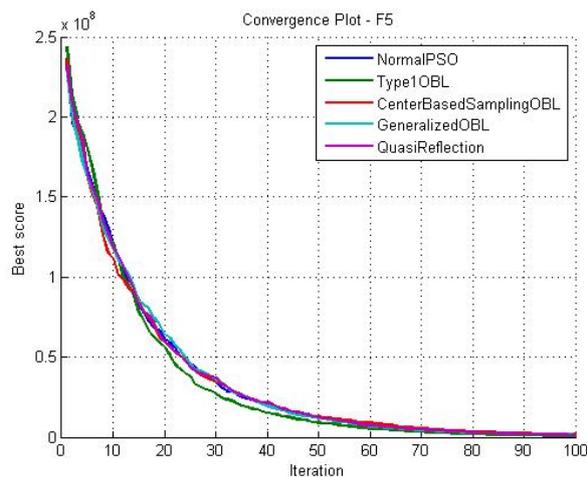
Figure 6: Convergence plot for the F5 function
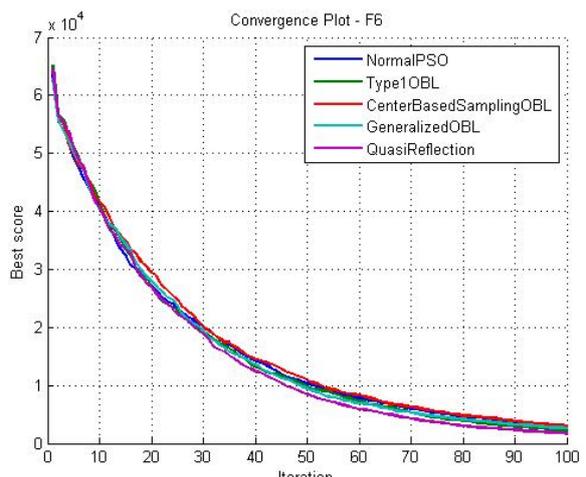
Figure 7: Convergence plot for the F6 function
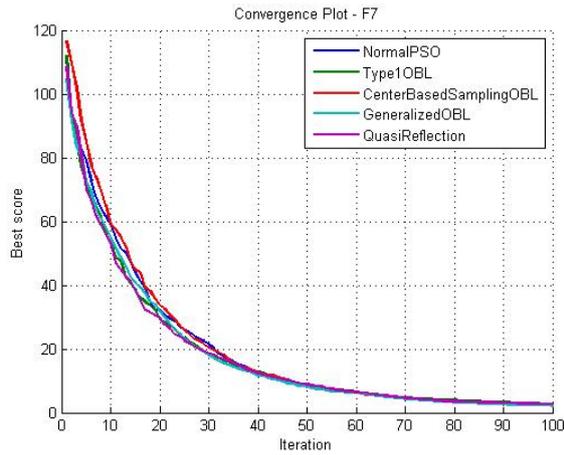
Figure 8: Convergence plot for the F7 function
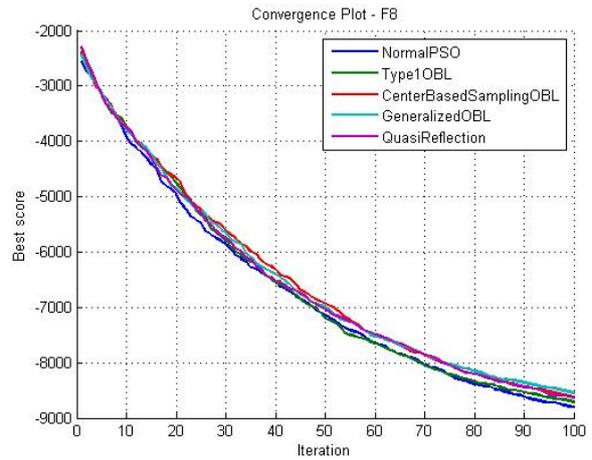
Figure 9: Convergence plot for the F8 function
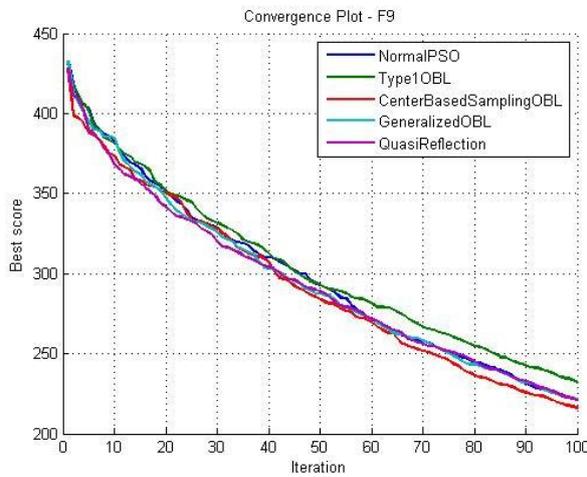
Figure 10: Convergence plot for the F9 function
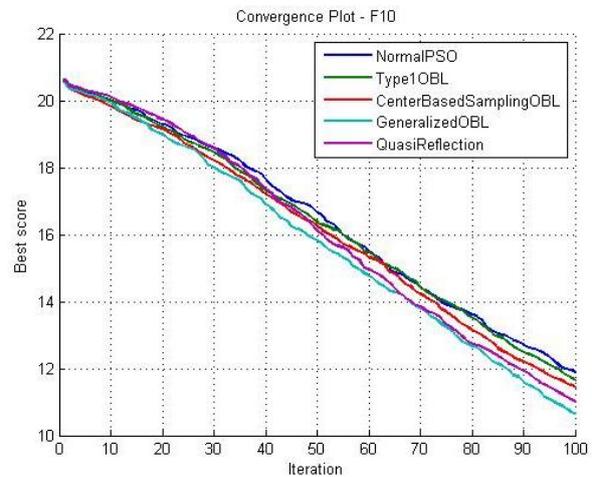
Figure 11: Convergence plot for the F10 function
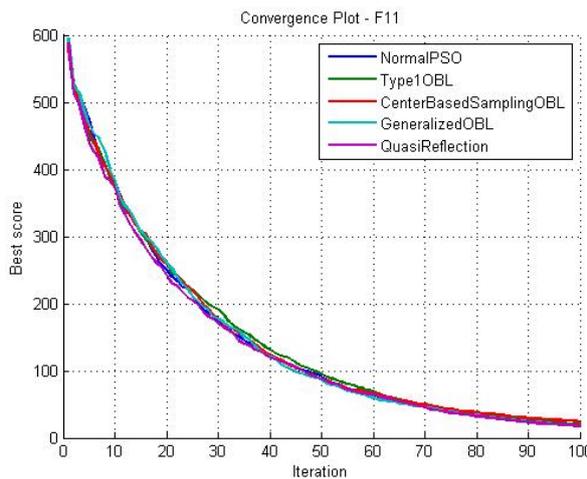
Figure 12: Convergence plot for the F11 function
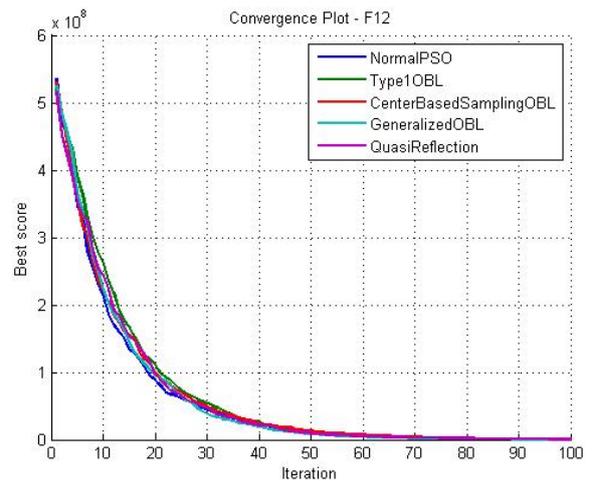
Figure 13: Convergence plot for the F12 function
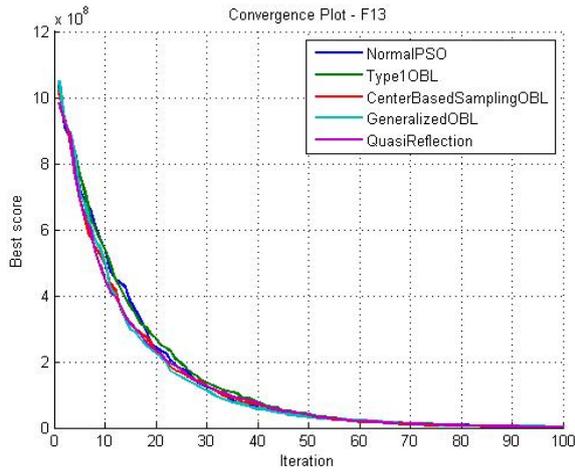
Figure 14: Convergence plot for the F13 function
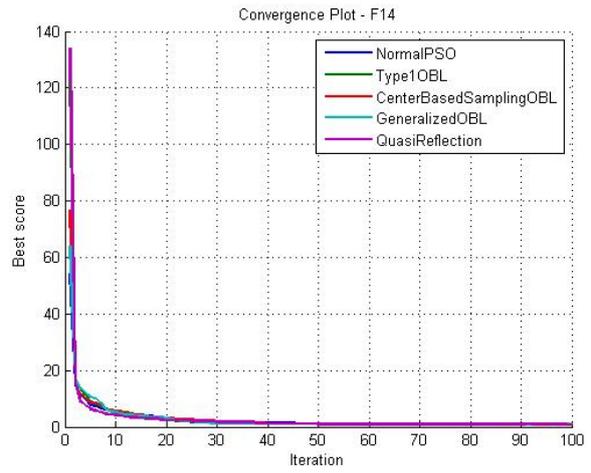
Figure 15: Convergence plot for the F14 function
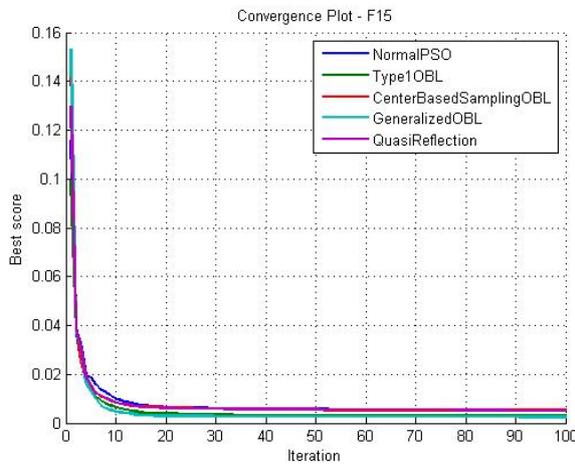
Figure 16: Convergence plot for the F15 function
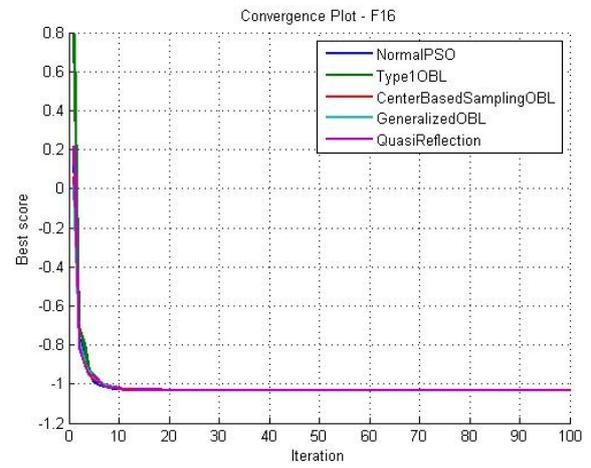
Figure 17: Convergence plot for the F16 function
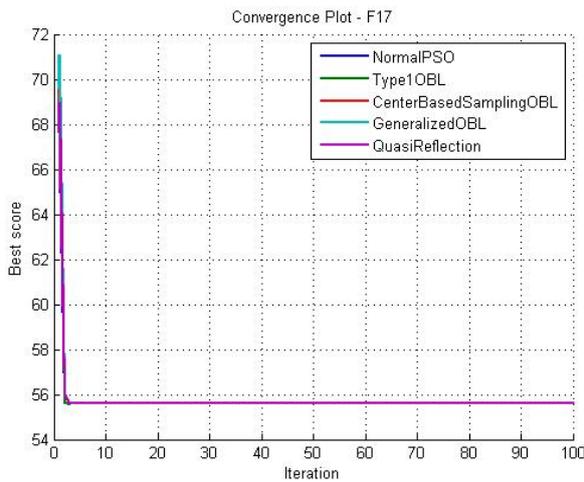
Figure 18: Convergence plot for the F17 function
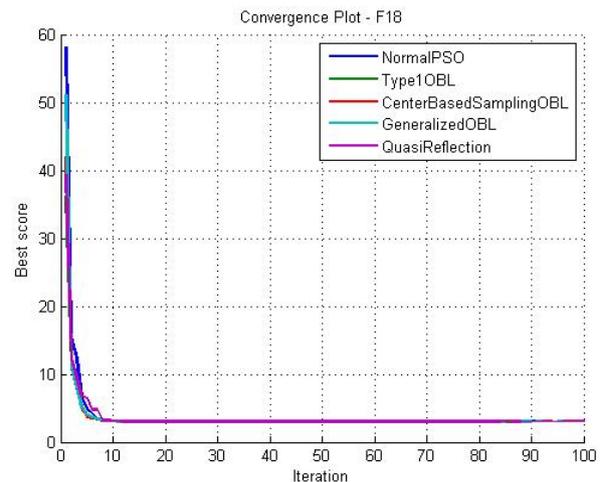
Figure 19: Convergence plot for the F18 function
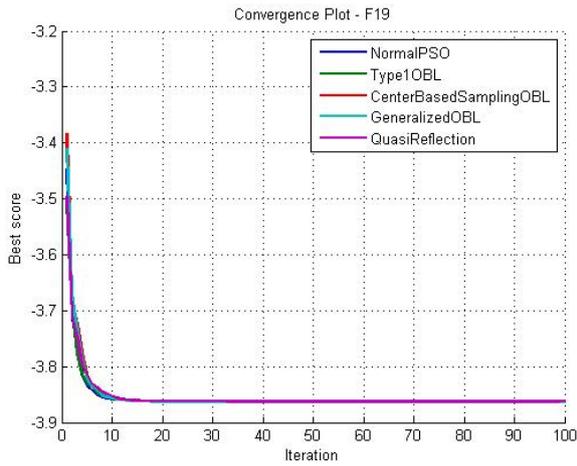
Figure 20: Convergence plot for the F19 function

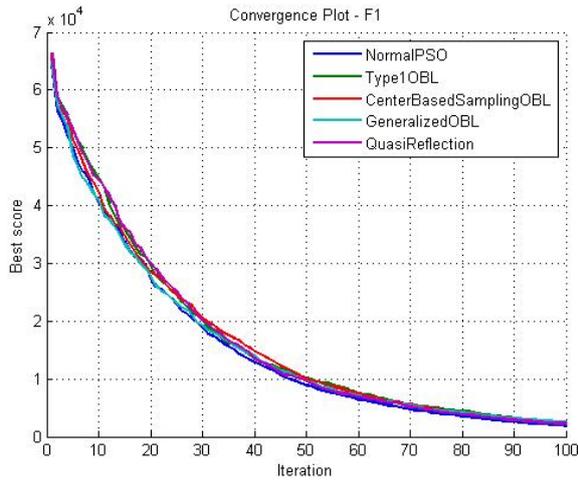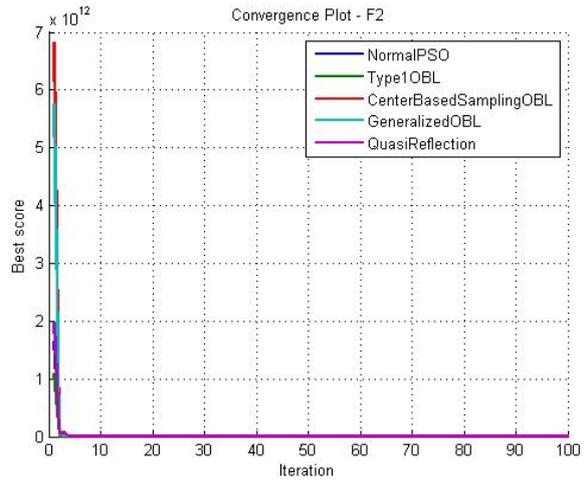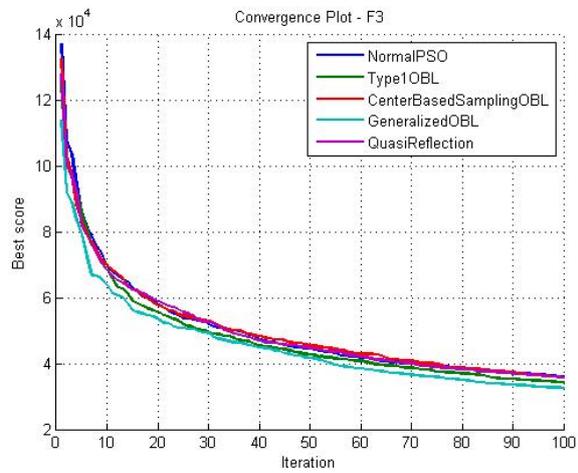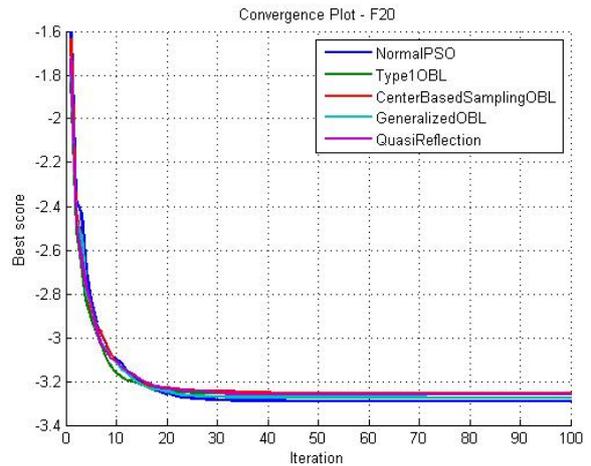Figure 21: Convergence plot for the F20 function
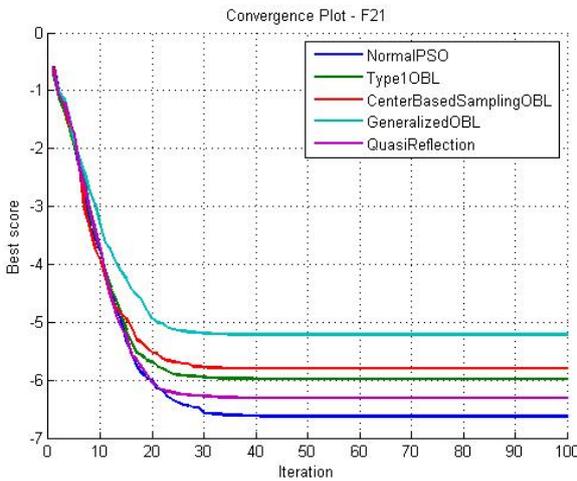
Figure 22: Convergence plot for the F21 function
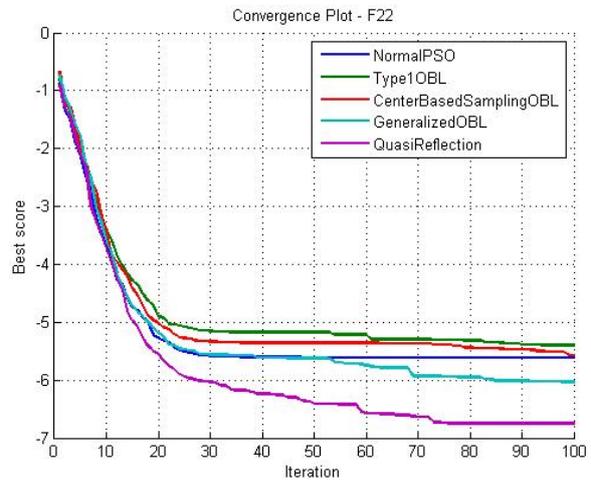
Figure 23: Convergence plot for the F22 function
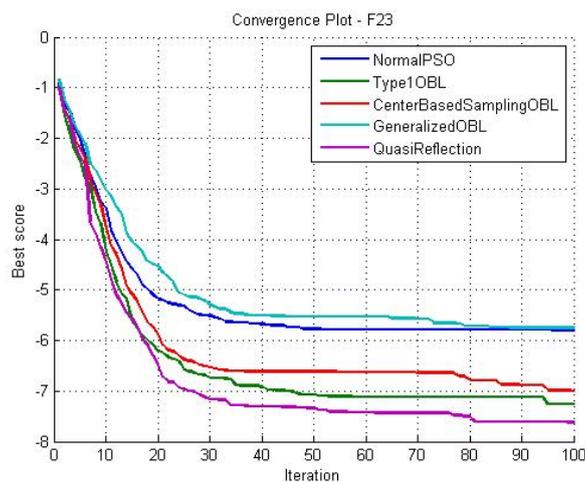
Figure 24: Convergence plot for the F23 function

CONCLUSION

The findings of this study underscore the efficacy of integrating OBL techniques with PSO to enhance optimization performance. OBL consistently demonstrated its ability to improve exploration by evaluating opposite solutions, helping the PSO algorithm avoid premature convergence and escape local minima. Statistical analysis on 23 benchmark functions revealed that OBL-PSO variants outperform the standard PSO algorithm, especially in multimodal and high-dimensional optimization problems. Among the variants tested, Generalized OBL and Center-Based Sampling OBL were particularly effective, achieving better global optimization and faster convergence. Conversely, Type-1 and Quasi-Reflection OBL methods showed more limited improvements but still contributed to diversifying the search process. While PSO remains competitive for simpler problems, the integration of OBL presents a promising strategy for tackling more complex optimization challenges. The study confirms the potential of OBL in advancing metaheuristic algorithms, emphasizing its versatility in balancing exploration and exploitation during optimization.

For future work, several directions are proposed to build upon the insights gained from this study: (i) Hybrid approaches could leverage the strengths of each method to address specific problem domains effectively. (ii) Investigating adaptive mechanisms to fine-tune OBL parameters during the optimization process may enhance performance across diverse and dynamic problem landscapes. (iii) Applying OBL-PSO to practical optimization problems in engineering, healthcare, or financial modeling could validate its efficacy in solving complex, domain-specific challenges. (iv) Expanding the scope to include large-scale optimization problems and exploring the scalability of OBL-PSO under increased problem dimensions would provide valuable insights into its robustness.

REFERENCES

[1] X.-S. Yang, *Nature-inspired metaheuristic algorithms*. Luniver press, 2010.

[2] I. Boussaïd, J. Lepagnot, and P. Siarry, "A survey on optimization metaheuristics," *Information sciences,* vol. 237, pp. 82-117, 2013.

[3] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-international conference on neural networks*, 1995, vol. 4: ieee, pp. 1942-1948.

[4] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence (Cat. No. 98TH8360)*, 1998: IEEE, pp. 69-73.

[5] H. R. Tizhoosh, "Opposition-based learning: a new scheme for machine intelligence," in *International conference on computational intelligence for modelling, control and automation and international conference on intelligent agents, web technologies and internet commerce (CIMCA-IAWTIC'06)*, 2005, vol. 1: IEEE, pp. 695-701.

[6] S. Rahnamayan, H. R. Tizhoosh, and M. M. Salama, "Opposition-based differential evolution," *IEEE Transactions on Evolutionary computation,* vol. 12, no. 1, pp. 64-79, 2008.

[7] S. Mahdavi, S. Rahnamayan, and K. Deb, "Opposition based learning: A literature review," *Swarm and evolutionary computation,* vol. 39, pp. 1-23, 2018.

[8] T. Khosla and O. P. Verma, "An Efficient Particle Swarm Optimization with Fusion of Figurate Opposition-based learning and Grey Wolf Optimizer," in *2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 2022: IEEE, pp. 208-213.

[9] M. R. Ullmann, K. F. Pimentel, L. A. de Melo, G. da Cruz, and C. Vinhal, "Comparison of PSO variants applied to large scale optimization problems," in *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 2017: IEEE, pp. 1-6.

[10] N. Ul Hassan, W. H. Bangyal, M. S. Ali Khan, K. Nisar, A. A. Ag. Ibrahim, and D. B. Rawat, "Improved opposition-based particle swarm optimization algorithm for global optimization," *Symmetry,* vol. 13, no. 12, p. 2280, 2021.

[11] M. Agarwal and G. M. S. Srivastava, "Opposition-based learning inspired particle swarm optimization (OPSO) scheme for task scheduling problem in cloud computing," *Journal of Ambient Intelligence and Humanized Computing,* vol. 12, no. 10, pp. 9855-9875, 2021.

[12] S. Rahnamayan, H. R. Tizhoosh, and M. M. Salama, "Opposition-based differential evolution algorithms," in *2006 IEEE international conference on evolutionary computation*, 2006: IEEE, pp. 2010-2017.

[13] W. Dong, L. Kang, and W. Zhang, "Opposition-based particle swarm optimization with adaptive mutation strategy," *Soft Computing,* vol. 21, pp. 5081-5090, 2017.

[14] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in engineering software,* vol. 95, pp. 51-67, 2016.

[15] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Advances in Engineering Software,* vol. 69, pp. 46-61, 2014/03/01/ 2014, doi: https://doi.org/10.1016/j.advengsoft.2013.12.007.

[16] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on evolutionary computation,* vol. 11, no. 6, pp. 712-731, 2007.

[17] N. Rojas-Morales, M.-C. R. Rojas, and E. M. Ureta, "A survey and classification of opposition-based metaheuristics," *Computers & Industrial Engineering,* vol. 110, pp. 424-435, 2017.

[18] M. Ergezer, D. Simon, and D. Du, "Oppositional biogeography-based optimization," in *2009 IEEE international conference on systems, man and cybernetics*, 2009: IEEE, pp. 1009-1014.

[19] S. Rahnamayan and G. G. Wang, "Center-based sampling for population-based algorithms," in *2009 IEEE Congress on Evolutionary Computation*, 2009: IEEE, pp. 933-938.

[20]   H. Wang, Z. Wu, S. Rahnamayan, and L. Kang, "A scalability test for accelerated DE using generalized opposition-based learning," in *2009 ninth international conference on intelligent systems design and applications*, 2009: IEEE, pp. 1090-1095.

[21]   M. Clerc, "Particle Swarm Optimization, vol. 93 John Wiley & Sons," ed: France, 2010.

[22]   R. Poli, "Analysis of the publications on the applications of particle swarm optimisation," *Journal of Artificial Evolution and Applications,* vol. 2008, no. 1, p. 685175, 2008.

[23]   M. Mavrovouniotis, C. Li, and S. Yang, "A survey of swarm intelligence for dynamic optimization: Algorithms and applications. Swarm and Evolutionary Computation, 33: 1-17," ed, 2017.

RESUME

Büşra Seval DOĞAN

Büşra Seval DOĞAN graduated from Selçuk University, Faculty of Engineering, Department of Computer Engineering. She has gained extensive experience in the design, development, and integration of ERP and CRM systems as a full-stack developer. She has worked on web and mobile application development, user-centered design, and system integration. During her master's studies, she conducted and published research on the application of evolutionary algorithms. Her expertise includes software architecture, database management, and modern design patterns. Currently, she is conducting research focused on the development of software tools to enhance the optimization performance of evolutionary algorithms and the integration of opposition-based learning (OBL) methods.

Assoc. Prof. Dr. Tahir SAĞ

Tahir SAĞ graduated from Selçuk University, Faculty of Engineering and Architecture, Department of Computer Engineering, in 2005. He completed both his Master's and Ph.D. degrees at Selçuk University, Graduate School of Natural and Applied Sciences, Department of Computer Engineering. Since 2005, Dr. Sağ has been serving at Selçuk University, Faculty of Technology, and in 2023, he earned the title of Associate Professor. He has authored numerous national and international scientific papers and publications in the fields of evolutionary computation and artificial intelligence. Dr. Sağ is married and has a daughter.

- 21 -

A NOVEL HYBRID ALGORITHM: ENHANCING TUNA SWARM OPTIMIZATION WITH CUCKOO SEARCH FOR GLOBAL OPTIMIZATION

Nesibe ŞAHİN[1], Sema SERVİ[2]

INTRODUCTION

Metaheuritic are flexible methods proposed to solve complex optimization problems with large-sized solution spaces. Without being strictly bound by a method, they can thus arrive at effective results on practically any field. Each nature mechanism inspires the development of metaheuristic algorithms, which perform a widespread search aimed at locating the global optimum of the problem's solution space. The preferred algorithms for solving a number of problems, their capability to bring in fast and efficient results, has drawn much attention towards them in recent years.

Among these, Tuna Swarm Optimization is developed based on the hunting behavior of tuna fishes collectively. Just like fish shoals that move around together in order to catch food, TSO tries to reach the best solution. However, one of the major disadvantages with TSO is its tendency to converge prematurely and get stuck in local minima. This makes it less effective to reach the global optimum.

Another nature-inspired metaheuristic algorithm is the heuristic algorithm Cuckoo Search, developed based on the strategy of cuckoo birds laying their eggs in other birds' nests. This algorithm effectively explores the solution space and stands out in the capability of reaching the global optimum without getting stuck in local minima. This strong search capability allows one to gain a considerable advantage over other optimization problems that can have more complexity.

The hybrid algorithms should realize better and more efficient solutions by incorporating strengths from various algorithms. This improves their capability of conducting a wide search in the solution space while avoiding entrapment in local solutions. Hybrid metaheuristic algorithms find broad applications in solving engineering and scientific problems.

It is seen within the literature that most of the metaheuristic algorithms contain solutions for various optimization problems. Hybrid versions obtained by using different algorithms together have gained wide success in solving those problems that contain wide solution spaces. example Hybrid versions of algorithms such as PSO and GA are among the frequently used algorithms in solving engineering problems.

---

[1] Selcuk University, Konya/Türkiye, Orcid: https://orcid.org/0009-0005-3053-9482, 218273002015@lisansustu.selcuk.edu.tr
[2] Selcuk University, Konya/Türkiye, Orcid: https://orcid.org/0000-0003-2069-9085semaservi@selcuk.edu.tr

For instance, Lei Xie and Tong Han, in the 2021 study, did an extensive analysis of the TSO algorithm and showed that this algorithm has better global and local searching capabilities as compared to other traditional algorithms. In this current study, the performance of the TSO was tested from the point of view of various benchmark functions. It has been shown from the results that the TSO is efficient in converging to the global optimum; hence, it proves to be a novel technique in solving optimization problems [12].

Kumar and Magdalin proposed techniques that could help in improving the phases of exploration and exploitation in TSO. In this study, a new algorithm known as Chaotic Tuna Swarm Optimization was developed. They integrated a Chaotic Tent Map into the basic TSO version to prevent the system from being stuck in a local minimum. Because of this new mechanism, an improvement in TSO's exploration and exploitation capability resulted in an improved outcome [13].

Another effective solution for image segmentation of forestry canopy is ITSO, proposed by Wang and Zhu. ITSO has successfully found the optimal threshold value by speeding up the entropy-solving procedure in image segmentation. Also, this hybrid method performs better than other segmentation algorithms [14].

Besides, the foraging behavior has been simulated into two different strategies that TSO used: spiral and parabolic food search. The research showed how TSO provided successful results on large-scale engineering problems and benchmark functions. Computer simulations by Xie et al. 2021 showed that the TSO outperforms other metaheuristic algorithms and makes such an algorithm an effective solution to optimization problems [12]

In this respect, the cuckoo search-tuna swarm optimization algorithm may be pointed out as a new approach, with its benefits from the strengths of both TSO and CS algorithms. CSTSO proposes fast and efficient solutions in broad solution spaces by merging the collective behavior of the fish swarms from TSO into the strong search capabilities of CS. This hybrid algorithm is prominent because it does not get stuck in local minima, finding the global optimum, and further has produced successful results within many areas, such as engineering applications.

The conference paper, after the introduction, will be composed of five main sections: The Materials and Methods section will explain the algorithms and research methods that have been used. The section on Findings and Discussion will present the performance of the CSTSO algorithm and compare it with other methods. Where appropriate, the general findings from the study and recommendations for further research will be summarized in the Conclusion and Recommendations section. The Acknowledgments and Notes section will be used to acknowledge individuals and institutions that assisted in the research. Finally, the References section will indicate the literature cited in the study.

MATERIALS AND METHODS

Tuna Swarm Optimization (TSO) Algorithm

Tuna Swarm Optimization is a newly developed metaheuristic algorithm, inspired by tuna fish navigating to find food as a group. The main objective of this approach is to simulate the

movement of tuna toward food to find the best solution for optimization problems. The TSO search process can be divided into two main phases: exploration and exploitation.

- The exploratory phase makes the algorithm search through different solution points in a broad solution space. In this stage, the swarm of fish activates at dispersed positions and tries to attain the position of the optimal food source. The movement of fish is based on interference among themselves towards the best positions of individuals in the group. Such a strategy offers a wide investigation of different space regions, giving prominence to the approaching attitude concerning the global minimum of the algorithm.

- Exploitation Phase The swarm tends to converge to the best-identified region in the case of exploitation. The fishes move their positions toward the nearest optimum source of food. Solutions detected in previously inspected areas are refined. This however poses a danger of locking into local optima. A weakness of the algorithm would be that the exploration stage may not be long-lasting enough thus causing it to go into prematurity convergence.

The TSO algorithm aims to solve by drawing inspiration from the collective behavior of fish. As each individual (fish) moves among solutions, information is shared between individuals moving collectively. This allows for the improvement of the solution by leveraging the advantages of collective movement.

Cuckoo Search (CS) Algorithm

Cuckoo Search algorithm is one of the recently proposed nature-inspired metaheuristic algorithms inspired by the strategy of cuckoo birds laying their eggs in other birds' nests. CS tries to reach the global optimum efficiently by making use of a random search strategy called Levy flight.

- Levy Flight: The core of the CS algorithm's search mechanism is done by Levy Flight, which allows a wide exploration in the space of solutions and employs a random but long step. This kind of search methodology can jump quickly from one point to another within the solution space so that it falls into the global optimum with higher probabilities. In this way, large steps over less frequent intervals rather than small, frequent steps are taken, thus enabling an enhanced capability of exploration in Levy flights.

- Egg-Laying Strategy: The CS algorithm makes a random position for each bird-like solution to lay an improved egg in its nest. If the new solution outperforms the current solutions, the old solution must be replaced by the new one. This strategy increases the convergence rate of the algorithm toward the global optimum.

- Renewal of the nests of birds: The algorithm rejects bad solutions (the eggs) with a specified probability and generates a new solution (nests). This allows the algorithm to avoid local minima and conduct a wide-ranging search for better solutions. The adaptive nature of the CS algorithm supplies apt search capabilities over extensive solution spaces.

Cuckoo Search-Tuna Swarm Optimization (CSTSO) Algorithm

CSTSO is a hybrid optimization algorithm that embeds the swarm behavior of TSO into the Levy flight-based search mechanism of CS. CSTSO aims to create a balance between an extensive exploration to discover the global optimum without being trapped in any of the local optima. This logic combines two major mechanisms:

- Solution Improvement with TSO's Swarm Behavior: CSTSO utilizes swarm behavior from the TSO algorithm to execute a collective search among solutions. The advantage of fish swarm behavior is that it offers improvements in exploration and exploitation phases more effectively as solutions to share information. As the swarm members move towards optimum as a group, they share information concerning the best positions with each other.

- Global Search Capability via Levy Flight of CS: While improving the exploration and exploitation phases of TSO, CSTSO also benefits from the powerful global search capability of CS. Levy flight enables the algorithm with longer steps through the solution space and increases the possibility of reaching the global optimum. Such long-distance steps prevent the algorithm from getting stuck in local minima, thus improving the solution quality.

- Adaptive Solution Strategy: Combine adaptive features of TSO and CS algorithms in CSTSO. This algorithm rejects bad solutions with some specific probability and generates a new solution to get faster and more efficient searches in optimization problems. This hybrid technique attains very strong performances for complex problems with large solution spaces.

By incorporating these two powerful mechanisms, the probability of CSTSO getting stuck in a local optimum is at a minimum so that faster convergence and quick results are obtained. The swarm-based search capabilities of TSO interacting with the global search advantages of CS confer effective solutions to the most challenging optimization problems, even those from engineering applications.

Figure 1 presents the pseudo-code of the main steps of the CSTSO algorithm. The code explains the basic functioning and the search strategies of the algorithm sequentially.

```
Algorithm: Cuckoo Search-Tuna Swarm Optimization (CSTSO)

Input:
    - MaxIterations: Maximum iterations
    - PopulationSize: Number of tuna (agents)
    - SearchSpace: Solution space boundaries
    - FitnessFunction: Objective function to optimize

Output:
    - BestSolution: Best solution found
    - BestFitness: Fitness value of the best solution

Step 1: **Initialize Population**
    - Randomly initialize tuna positions in the search space
    - Evaluate fitness for each tuna, set BestSolution and BestFitness

Step 2: **Main Loop** (until MaxIterations reached)
    For iteration = 1 to MaxIterations do:
        - Update exploration-exploitation parameters (a1, a2)

        For each tuna i:
            - **Ensure Boundaries:** Adjust position if out of bounds
            - **Levy Flight:** Apply Cuckoo Search's Levy flight to generate new solutions
            - **Adaptive TSO Movement:** Move tuna towards best solution using TSO's group behavior
            - **Random Reset:** With probability z, randomly reset tuna position

        - **Update Best Solution:** If better solution is found, update BestSolution and BestFitness

Step 3: **Return Best Solution and Fitness**
```

Figure 1: Pseudo-code of CSTSO

FINDINGS AND DISCUSSION

In this work, the performance of the proposed CSTSO algorithm is evaluated on a set of well-known benchmark functions. For the provided test functions, there are 7 unimodal and 6 multimodal functions. Functions F1-F7 with only one global optimal are utilized to measure the local exploitation capability of the algorithm. Functions F8-F13 are multimodal functions with multiple local optima, testing the algorithm's capability of [17] global exploration and its success in avoiding local optima. The parameters used in these evaluations include a population size of 30 and a maximum of 500 iterations for each function. Moreover, each function was tested independently with 30 runs. Table 1 and Table 2 present the average results of 30 independent runs of these functions, while Table 3 shows their three-dimensional plots.

Results for Unimodal Functions F1-F7

Table 1: F1 – F7 test results

Function	f_{min}		TSO	CSTSO
F1	0	Mean	2.7654e-225	**1.9739e-231**
		S.D	0	**0**
F2	0	Mean	2.1192e-114	**6.6185e-119**
		S.D	9.7184e-114	**4.3248e-118**
F3	0	Mean	8.6824e—214	**1.8730e-215**
		S.D	0	**0**
F4	0	Mean	2.7460e-114	**1.1047e-115**
		S.D	1.0470e-113	**4.4405e-115**
F5	0	Mean	0.9020	**0.3226**
		S.D	4.3983	**1.6421**
F6	0	Mean	**1.8923e-04**	2.9700e-04
		S.D	**2.6082e-04**	5.0998e-04
F7	0	Mean	**3.8339e-04**	3.8627e-04
		S.D	**3.6827e-04**	3.9101e-04

CSTSO generally provided better average results and lower standard deviation values compared to TSO. This indicates that CSTSO produces more consistent solutions closer to the optimum.

It was observed that CSTSO performed better than TSO, particularly on challenging functions (F1-F4), and improved the consistency of the results.

In the F5 function, the CSTSO algorithm demonstrated superior and more consistent performance, with both a lower average and lower standard deviation values.

For the F6 and F7 functions, TSO outperformed CSTSO in terms of both average and standard deviation, indicating that TSO provided more accurate and consistent results on these functions.

Results for Multimodal Functions F8-F13

Table 2: F8 – F13 test results

Function	f_{min}		TSO	CSTSO
F8	-12569.487	Mean	**-1.2569e+04**	**-1.2569e+04**
		S.D	**0.0166**	**0.0121**
F9	0	Mean	**0**	**0**
		S.D	**0**	**0**
F10	0	Mean	**4.4409e-16**	**4.4409e-16**
		S.D	**0**	**0**
F11	0	Mean	**0**	**0**
		S.D	**0**	**0**
F12	0	Mean	7.4254e-06	**5.9826e-06**
		S.D	1.3104e-05	**1.1416e-05**
F13	0	Mean	**7.6814e-04**	0.0028
		S.D	**0.0038**	0.0049

In the F8 function, both algorithms achieved results very close to the target minimum value. However, CSTSO provided more consistent results with a lower standard deviation.

For the F9, F10, and F11 functions, both algorithms achieved ideal results (with 0% error), and thus no significant difference was observed between them.

In the F12 function, CSTSO outperformed TSO in terms of both average and standard deviation values, meaning that CSTSO demonstrated more accurate and consistent performance in this function.

For the F13 function, TSO provided better results in terms of average value and standard deviation, producing a solution closer to the target.

Table 3: 3D visualization for 2D benchmark functions
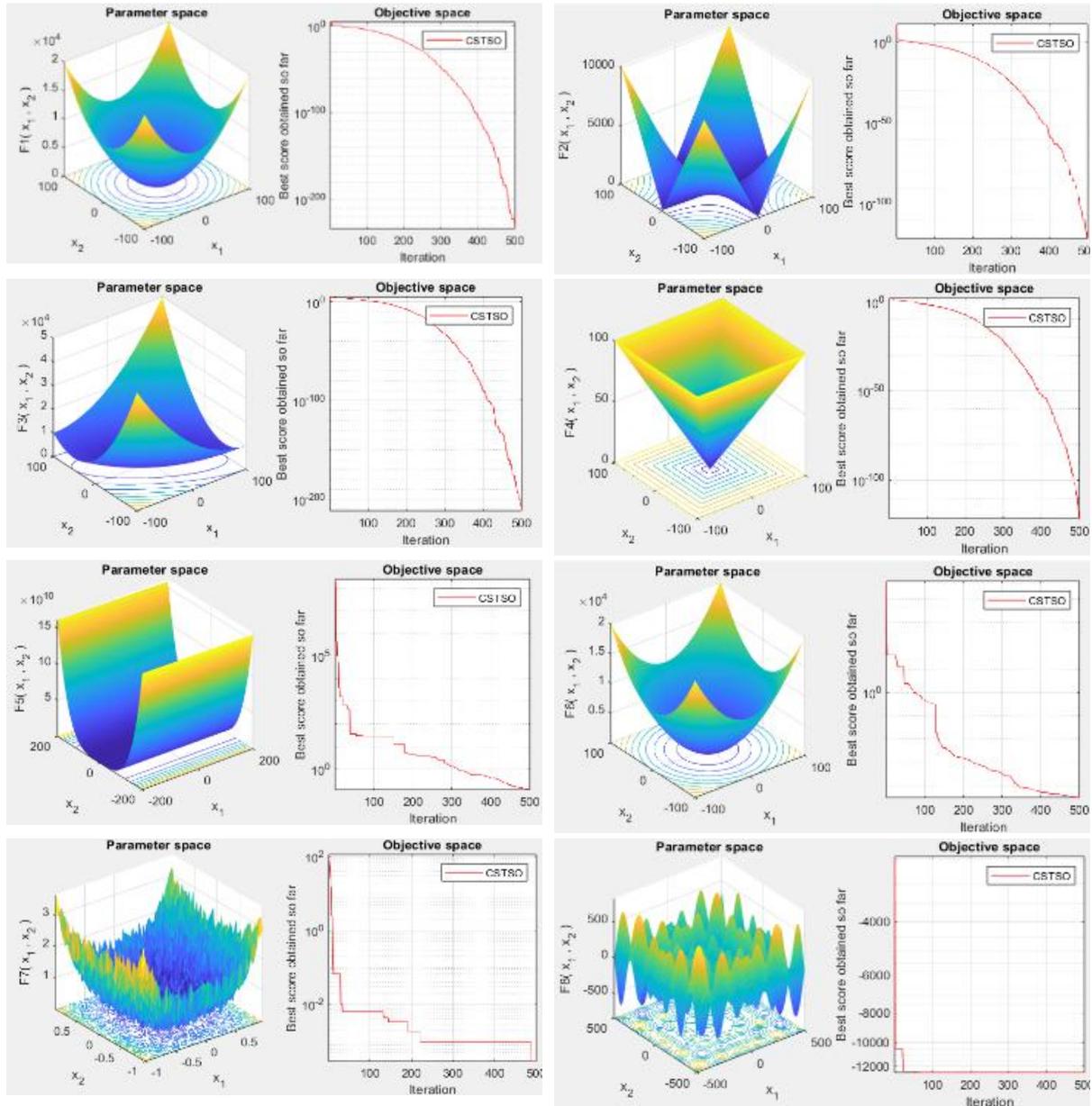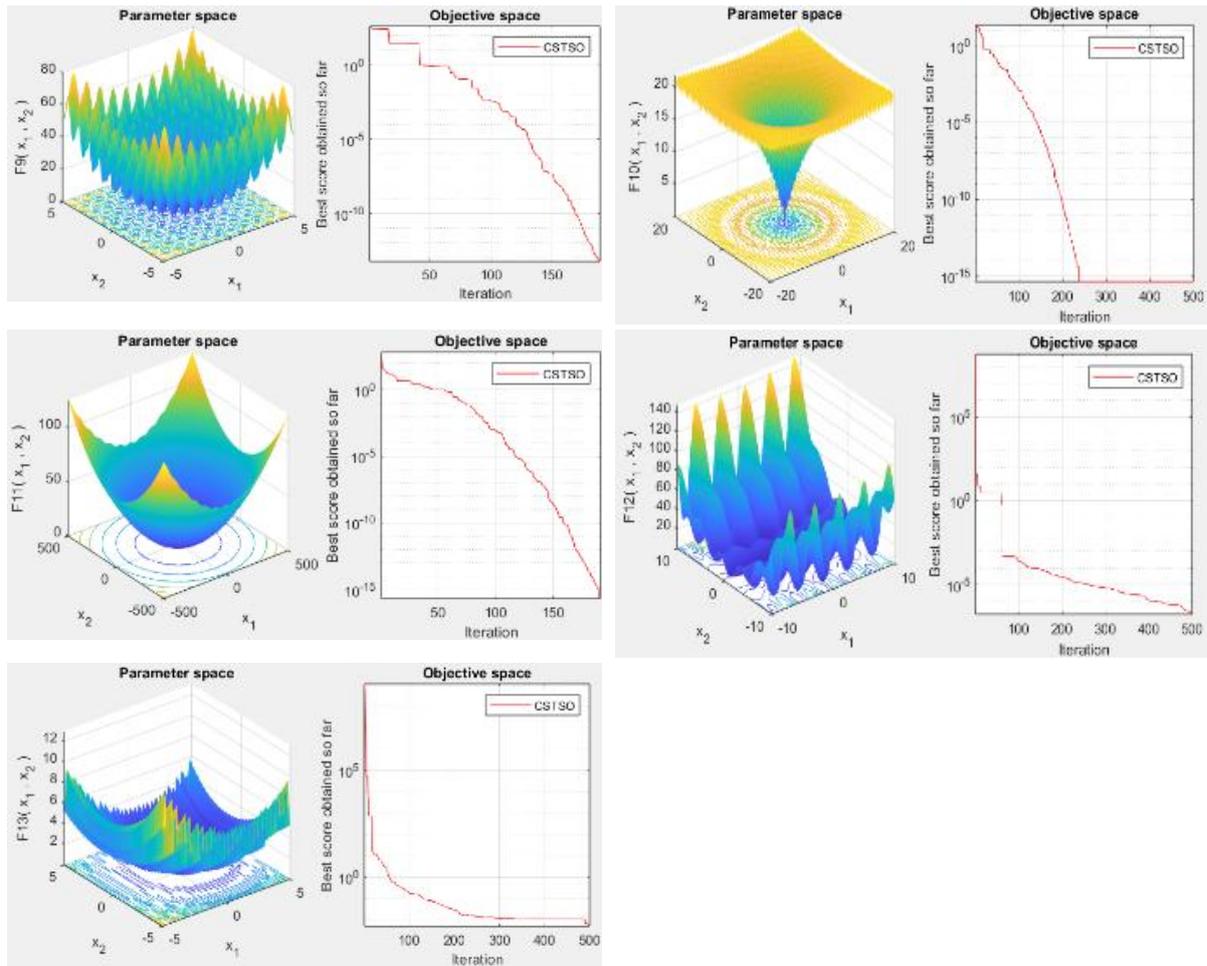
CSTSO for Engineering Design Problem

The capability of CSTSO in solving real-world problems has been assessed on considering an engineering design problem. The welded beam problem was considered. This kind of engineering design problem has been solved by CSTSO by utilizing a population of 50 coupled with 1000 iterations. Each of the problems was run individually for 30 runs, and the statistical results obtained were compared to the results provided by other methods in the literature. The approach proposed herein provides a firm basis for the optimization performance assessment of the CSTSO approach and its comparison with other approaches.

Welded Beam

The welded beam problem is one of the classic structural optimization problems in engineering. The major objective of this problem is to minimize the production cost of the beam. The optimization variables are defined as:

➢ (x_1) : thickness of weld
➢ (x_2) : joint length
➢ (x_3): height of beam
➢ (x_4) : Thickness of the beam

The problem of the design is given in the following mathematical model [16]:

$$\min f(x_1, x_2, x_3, x_4) = 1.10471x_1^2 x_2 + 0.04811x_3 x_4(14.0 + x_2)$$

Thus, the problem can be considered as an optimization problem with a variety of structural and material constraints represented by the following equations:

$$g_1(X) = \tau_d - \tau(X) \geq 0,$$
$$g_2(X) = \sigma_d - \sigma(X) \geq 0,$$
$$g_3(X) = x_4 - x_1 \geq 0, \tag{1}$$
$$g_4(X) = P_c(X) - P \geq 0,$$
$$g_5(X) = \delta_d - \delta(X) \geq 0,$$

The terms and equations used here are as follows:

Shear stress ($\tau(X)$): (2)

$$\tau(X) = \sqrt{(\tau'(X))^2 + (\tau''(X))^2 + \frac{x_2 . \tau'(X).\tau''(X)}{0.25x_2^2 + (x_1 + x_3)^2}}$$

Bending stress ($\sigma(X)$): (3)

$$\sigma(X) = \frac{50400}{x_3^2 x_4}$$

Critical buckling force ($P_c(X)$): (4)

$$P_c(X) = 64746.002(1 - 0.0282346x_3)\, x_3 x_4^3$$

Deflection ($\delta(X)$) (5)

$$\delta(X) = \frac{2.1952}{x_3^3 x_4}$$

Shear force components ($\tau'(X)\ and\ \tau''(X)$): (6)

$$\tau'(X) = \frac{6000}{\sqrt{2x_1 x_2}}$$

$$\tau''(X)) = \frac{6000(14 + 0,5x_2)}{\sqrt{0,25x_2^2 + (x_1 + x_3)^2}}$$

The developed mathematical model and the constraints mentioned predefined a basis on which a welded beam can achieve an optimum in terms of structural integrity and cost. The minimum cost also considers parameters such as the size of the beam and the limits on stress for optimization.

This problem has been solved in several algorithms like DDSCA, HGA, MGWO-III, IAPSO, TEO, hHHO-SCA, HPSO, CPSO, WCA, and TSO. Table 4 summarizes the best results found by the above algorithms and then compares those with the best obtained results by CSTSO. The results indicate that CSTSO achieved a parameter design plan at a lower cost in comparison to other algorithms. CSTSO gave the best solution with minimum cost 1.4731 and the design variables as 0.18296, 2.4083, 9.5818, and 0.18299.

Table 4: Comparison of the best solutions [12] returned by the optimizers on the Welded Beam Design problem [15].

Table 4: Welded Beam results

Optimum values for variables					
Algorithm	x_1	x_2	x_3	x_4	Optimal cost
DDSCA [3]	2.0516e-1	34.759e-1	90.797e-1	2.0552e-1	17.305e-1
HGA [4]	2.05712e-1	34.70391e-1	90.39693e-1	2.05716e-1	17.25326e-1
MGWO-III [5]	2.05667e-1	34.71899e-1	90.36679e-1	2.05733e-1	17.24984
IAPSO [6]	2.05729e-1	34.70886e-1	90.36623e-1	2.05729e-1	17.24852e-1
TEO [7]	2.05681e-1	34.72305e-1	90.35133e-1	2.05796e-1	17.25284e-1
hHHO-SCA [8]	1.90086e-1	36.96496e-1	93.86343e-1	2.04157e-1	17.79032e-1
HPSO [9]	2.0573e-1	34.70489e-1	90.36624e-1	2.0573e-1	17.24852e-1
CPSO [10]	2.02369e-1	35.44214e-1	90.48210e-1	2.05723e-1	17.28024e-1
WCA [11]	2.05728e-1	34.70522e-1	90.36620e-1	2.05729e-1	17.24856e-1
TSO [12]	2.05729e-1	34.70490e-1	90.36626e-1	2.05729e-1	17.24854e-1
CSTSO	1.8296e-1	24.083e-1	95.818e-1	1.8299e-1	**14.731e-1**

CONCLUSION AND RECOMMENDATIONS

In this study, successful results were obtained in the solution of engineering problems with the CSTSO algorithm formed by combining the Tuna Swarm Optimization (TSO) and Cuckoo Search algorithm in a hybrid structure. In particular, tests carried out on the problem of welded beam design evaluated the performance of the algorithm in the case of minimizing cost and fulfilling all constraints, and results better than the results of other algorithms in the literature were obtained [2]. CSTSO has a smaller local minimum problem, or it found the global optimum faster.

Some recommendations for the further improvement of this algorithm in future studies are as follows: first, CSTSO can be tested on various engineering problems, enabling a wider application area. Adaptive adjustments to the parameters of the algorithm could be made to enhance its performance further. It could also be adapted for multi-objective optimization problems and hybridized with other metaheuristic algorithms to reach even more powerful solutions. In that way, better results can be obtained for more complex and high-dimensional problems.

THANKS, AND INFORMATION NOTE

REFERENCES

[1] Rakita, D., Mutlu, B., Gleicher, M., & Hiatt, L. M. (2019). Shared control–based bimanual robot manipulation. Science Robotics, 4(30), eaaw0955.

[2] Huang, Q., Wang, R., Yan, D., & Zhang, J. (2018). AAC double compression audio detection algorithm based on the difference of scale factor. Information, 9(7), 161.

[3] Li, Y., Zhao, Y., & Liu, J. (2021). Dimension by dimension dynamic sine cosine algorithm for global optimization problems. Applied Soft Computing, 98, 106933.

[4] Yan, X., Liu, H., Zhu, Z., & Wu, Q. (2017). Hybrid genetic algorithm for engineering design problems. Cluster Computing, 20, 263-275.

[5] Kumar, V., & Kumar, D. (2017). An astrophysics-inspired grey wolf algorithm for numerical optimization and its application to engineering design problems. Advances in Engineering Software, 112, 231-254.

[6] Guedria, N. B. (2016). Improved accelerated PSO algorithm for mechanical engineering optimization problems. Applied Soft Computing, 40, 455-467.

[7] Kaveh, A., & Dadras, A. (2017). A novel meta-heuristic optimization algorithm: thermal exchange optimization. Advances in engineering software, 110, 69-84.

[8] Kamboj, V. K., Nandi, A., Bhadoria, A., & Sehgal, S. (2020). An intensify Harris Hawks optimizer for numerical and engineering optimization problems. Applied Soft Computing, 89, 106018.

[9] He, Q., & Wang, L. (2007). A hybrid particle swarm optimization with a feasibility-based rule for constrained optimization. Applied mathematics and computation, 186(2), 1407-1422.

[10] He, Q., & Wang, L. (2007). An effective co-evolutionary particle swarm optimization for constrained engineering design problems. Engineering applications of artificial intelligence, 20(1), 89-99.

[11] Eskandar, H., Sadollah, A., Bahreininejad, A., & Hamdi, M. (2012). Water cycle algorithm–A novel metaheuristic optimization method for solving constrained engineering optimization problems. Computers & Structures, 110, 151-166.

[12] Xie, L., Han, T., Zhou, H., Zhang, Z. R., Han, B., & Tang, A. (2021). Tuna swarm optimization: a novel swarm-based metaheuristic algorithm for global optimization. Computational intelligence and Neuroscience, 2021(1), 9210050.

[13] Kumar, C., & Mary, D. M. (2022). A novel chaotic-driven Tuna Swarm Optimizer with Newton-Raphson method for parameter identification of three-diode equivalent circuit model of solar photovoltaic cells/modules. Optik, 264, 169379.

[14] Wang, J., Zhu, L., Wu, B., & Ryspayev, A. (2022). Forestry canopy image segmentation based on improved tuna swarm optimization. Forests, 13(11), 1746.

[15] Selim, H., Haikal, A. Y., Labib, L. M., & Saafan, M. M. (2024). MCHIAO: a modified coronavirus herd immunity-Aquila optimization algorithm based on chaotic behavior for solving engineering problems. Neural Computing and Applications, 1-85.

[16] Yildiz, A. R. (2012). A new hybrid particle swarm optimization approach for structural design optimization in the automotive industry. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 226(10), 1340-1351.

[17] Liu, Z., & Chen, H. (2023, January). An improved Harris Hawk optimization algorithm and its application to Extreme Learning Machine. In 2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE) (pp. 842-846). IEEE.

RESUME

Nesibe ŞAHİN

Nesibe ŞAHİN graduated from Harran University Computer Engineering in 2021. In 2022, she started her master's degree in Computer Engineering at Selçuk University and is still continuing. She is working on metaheuristic algorithms.

Assist. Prof. Dr. Sema SERVİ

Sema Servi is a Assistant Professor Dr. in Computer Engineering at Selcuk University, Faculty of Technology. She completed her master's and doctorate degrees at Selcuk University, Faculty of Science, Department of Mathematics. Servi has published articles on various national and international journals on topics such as applied mathematics and artificial intelligence. Her main research interests include numerical approximation solution methods, optimization algorithms and data mining.

A COMPREHENSIVE OVERVIEW OF MOBILITY PREDICTION MECHANISMS, MODELS AND ARCHITECTURE

Shiva Derogar KHACHEKINI[1], Gürcü ÖZ[2], Ali Hakan ULUSOY[3]

INTRODUCTION

Even though users in mobile networks prioritize mobility, there are several significant issues with mobile communication systems, including handover challenges, traffic capacity, user location updates, registration processes, and multilayer network management. The impact of handover is often one of the prevalent issues in mobile networks. This means that when a user leaves one coverage area but is unable to connect to a new cell without any problems, it can result in an unacceptable latency or even call-dropping, which calls for appropriate handover management. Thus, mobility is very beneficial for user comfort, but if it is not properly regulated and controlled, it may also lead to some degradation. As a result, mobility aspects have garnered interest in numerous academic and industrial fields. Using a mobility prediction system is one of the effective and useful ways to handle mobility challenges. An important aspect of future mobile network systems will be the ability to predict a user's next location or even the route they will take. Mobility prediction applications in mobile networks can typically be categorized into three groups: location-based applications, resource management, and handover management. Additionally, if an effective mechanism can accurately evaluate users' future movements, the network will be able to anticipate traffic conditions and develop more efficient plans for mobile clients. Additionally, certain location-based information can be transmitted. The Hierarchical Dynamic Bayesian Network (HDBN) and Hidden Markov Model (HMM) and are among the numerous processes and techniques created to identify human mobility patterns and predict human movement across various domains using mobility data [1]. The Markov Chain (MC) predictor is a widely used and significant mobility prediction method. Based on a location's past, this method can be used to forecast the next place that people will visit. One kind of predictor that examines user mobility is called MC. Based on the current or past location; it forecasts the upcoming one [2–3]. Even though the Markov model mobility prediction mechanism has been the subject of numerous studies and publications and may yield satisfactory outcomes, this approach has certain drawbacks and restrictions that might be addressed and improved. In order to forecast a user's mobility, the majority of related work focuses on location aspects of the user's data.

[1] Eastern Mediterranean University, North Cyprus, via Mersin 10, Turkey, Orcid: https://orcid.org/0009-0001-7335-9175, shiva.derogar@yahoo.com
[2] Eastern Mediterranean University, North Cyprus, via Mersin 10, Turkey, Orcid: https://orcid.org/0000-0002-6744-7427, gurcu.oz@emu.edu.tr
[3] Eastern Mediterranean University, North Cyprus, via Mersin 10, Turkey, Orcid: https://orcid.org/0000-0002-2992-9265, alihakan.ulusoy@emu.edu.tr

This article's structure continues as follows: A number of related studies and papers that are centered on mobility prediction processes are explained and presented in the following section as related work. The mobility prediction architecture is fully explained in Section 3. Section 4 delves deeply into a variety of mobility prediction models and algorithms. Lastly, the next section presents the conclusion.

RELATED WORKS

In the last thirty years, mobility prediction theory has been thoroughly researched to anticipate a user's future location. For varied purposes, such as resource management, handover management, and LBSs, these studies used a variety of strategies and procedures. In order to execute and maximize the mobility predictions that are given, a variety of works and publications used and presented various ideas, procedures, and approaches [4]. Researchers have proposed a model with several distinct phases to forecast human movement behavior in an article [5]. Markov chains serve as the basis for the prediction process, focusing on the extraction of the region of interest as the main phase of the model. Furthermore, it was demonstrated that second-order Markov chains are more effective than first-order Markov chains in predicting human motion. In a different study, the authors used the GeoLife data set to forecast user movement using the Gaussian Hidden Markov (GHM) algorithm. Five phases of implementation were used to conduct this study. Accuracy was the primary evaluation criterion in this work, and its implementation achieved an accuracy of 0.769 [6]. The researchers introduced a unique technique that combines the use of the Probability Density Function (PDF) for departure time prediction with the Markov model for position prediction. For mobility behavior with great regularity, as indicated by low entropy, this model had the maximum prediction accuracy. The authors investigated the occurrences of missing places in [7]. On the two user subgroups with and without missing locations, they attempted to apply mobility prediction based on n-order Markov chain models. At least 26.88% more predictions were made for the second subset than for the missing point. The prediction performance is improved by proposing an optimized Markov chain model. Additionally, prediction accuracy can be raised by up to 12.93%. The authors presented a novel technique for an n-Markov model based on future place prediction that remembers the n prior places visited. Analyses on three different databases demonstrate that this prediction algorithm's accuracy rises from 70% to 95%. In 2017, a mobility prediction mechanism based on HMM was developed for an efficient Wi-Fi AP selection system [8]. According to the evaluations, this model enhanced the connections to high-signal Wi-Fi access points while reducing the connections to low-signal Wi-Fi access points. Despite having a lot of handovers, the suggested model enhanced user experience and quality of service. To determine the shortest path from a source node to a destination node, researchers developed a prediction method for Wireless Sensor Networks (WSN). The most dependable forwarding nodes regarding network connectivity were selected based on the reliability parameter of this prediction technique [9]. Researchers concentrated on showcasing an approach for predicting multimodal user mobility patterns along with a distinctive multiuser multivariate multi-order Markov model in their study [10]. They evaluated the effectiveness of the suggested methods using a set of user's GPS trajectory data. The results of the experiment showed that, in comparison to the Z-eigen-based method, the proposed multi-order Markov-

based multimodal prediction model could enhance trajectory prediction accuracy by as much as 31.10 percentage points. To describe the traffic conditions in two-dimensional (2D) space during peak hours, a novel approach based on HMM and Contrast was put forth [11]. To record the variation in traffic patterns, researchers used mean speed and contrast. The results demonstrated that, when compared to neuro-fuzzy approaches and work linked to HMM, this model had better prediction error. A state transition probability model based on HMM was suggested by the authors of the paper [12] and is capable of predicting the lifetime for a variety of endurance failure mechanisms. They were able to get an accuracy metric of 83.6%. In order to save the failing cells throughout endurance, they also create an optimal programming algorithm. The memory chip's failure rate decreased by 38%. A particular mobility prediction and error analysis method was put into place in article [13]. In order to compute the primary phase of this algorithm and give a probability transition matrix, it includes a questionnaire survey of college students' autonomous learning. The pertinent data is introduced into the prediction process once the model has been put into practice using this matrix. The average relative error in the results' error analysis is 6%, demonstrating the model's dependability. MATLAB was also used to anticipate the real data, and appropriate outcomes were obtained. A trajectory prediction method was put forth that uses a user's past mobility data to forecast their future trajectory while also accounting for users with comparable mobility data. This model used an Extended Mobility Markov Chain (EMMC) approach to obtain a prediction model for a single user and an optimized DBSCAN algorithm to extract stay points from a user's mobility data [14]. Using comprehensive, semantically rich spatial-temporal data from location-based social networks, researchers in [15] put forward a semantics-aware mobility model that accounts for human mobility motivations. In this study, they utilized the Hidden Markov Model (HMM) to learn latent states and their transitions within the embedding space. They verified the method's capacity to generate high-quality mobility models by analyzing it on two various datasets. According to this evaluation, the optimized model works better than the most advanced mobility models, exhibiting much greater efficiency and higher prediction accuracy. To address the issues of sparsity and enhance prediction performance, the authors proposed a multi-task learning approach designed to predict users' mobility. This system tracks the mobility patterns of multiple users concurrently, leveraging their shared characteristics. To characterize the spatiotemporal mobility patterns of users, they put into practice a Bayesian mixture model. Additionally, they performed joint mobility prediction for all users using the hierarchical technique. They used mobile applications and data gathered by a cellular network operator to perform thorough assessments. According to their findings, this system can predict with an average accuracy of 53.9% and 73.2% for users who have high-quality mobility data [16]. Using a wireless LAN system and a mobility robot prototype, the authors demonstrated a self-status-based wireless link quality prediction and assessed its effectiveness. The dataset collected in an indoor experiment is used to assess prediction performance in offline processing [17]. A cell selection method for handover has been developed by researchers to enable simultaneous adaptive mobility and signal strength prediction [18]. Adaptive order Lagrange interpolation was utilized to predict the user's path, with the order being modified based on the user's moving condition. Based on numerical findings, the proposed method reduces the number of handovers by 16.8% and handover failures by 56.9% when compared to the contrast method.

Mobility prediction architecture

The use of mobility prediction is one of the effective and useful strategies to manage mobility and preserve users' ongoing connections. Future mobile network systems and apps must have the capability to anticipate a user's next cell [1]. One of the primary tools for predicting users' future locations is mobility prediction. However, it is challenging to forecast users' future positions, particularly from a network standpoint. This section outlines a general architecture for implementing mobility prediction, which can be divided into six components, as shown in Fig. 1. These elements include the data source, necessary data, prediction outputs, prediction algorithms, performance indicators, and application categories. Every section is presented in detail, and some significant examples of each are also discussed in general terms.
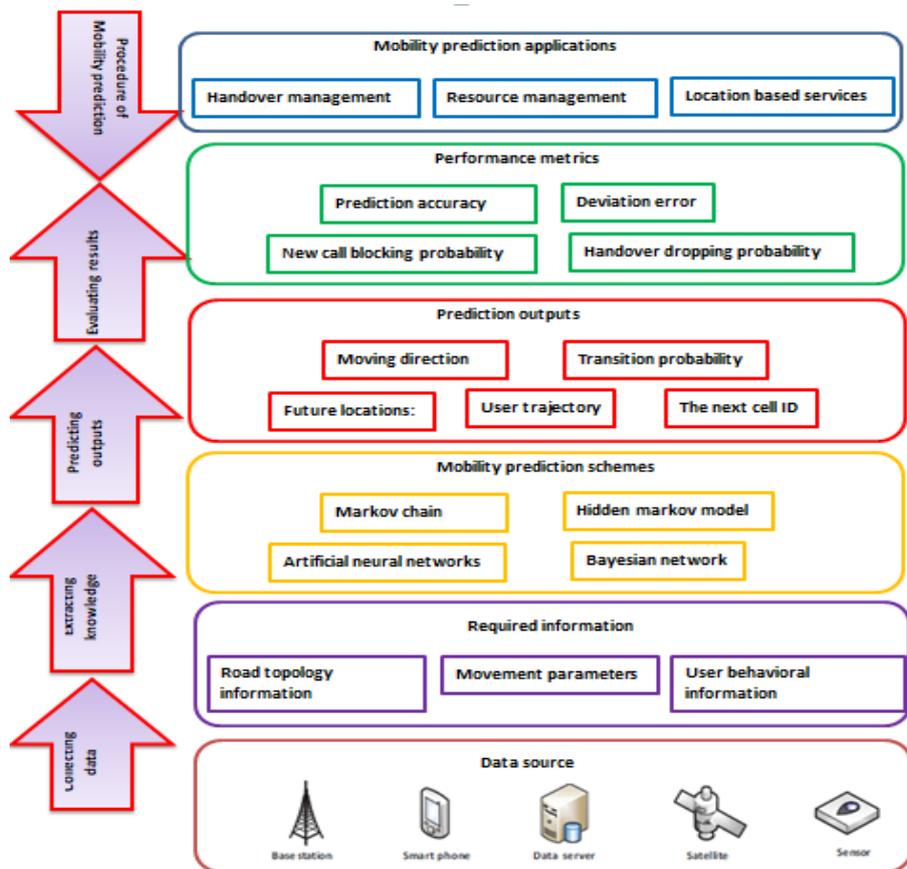
Figure 1: Mobility prediction architecture

Mobility prediction applications

Applications of mobility prediction in mobile networks can generally be classified into three main categories: location-based services, resource management, and handover management. The handover outcome illustrates the typical issues encountered by mobile networks. When a user exits a coverage area and struggles to connect to a new cell, they may experience unacceptable delays or call drops, highlighting the need for an effective handover management system [2]. Furthermore, employing a passive resource reservation method is the sole means of

guaranteeing continuous services without the necessity of reserving a substantial number of resources across the entire networ. By pre-reserving a designated amount of bandwidth in the coverage areas that customers are likely to use, this approach prevents the unnecessary allocation of bandwidth across the entire system, thereby conserving significant resources. Additionally, envision a network capable of accurately forecasting users' future actions. The network will be capable of predicting traffic conditions and developing more efficient strategies, enabling the dissemination of specific location-based information that is closely tied to users. As a result, mobility prediction emerges as a strong candidate for future networks, and the following provides a comprehensive overview of its application classification.

Handover management: In order to improve service during the handover process in mobile networks, customers unplug from their existing Transmission Points (TPs) and connect to alternative TPs. The handover scenario usually occurs in accordance with the Reference Signal Received Power (RSRP). Because of the RSRP volatility, which causes the services to fluctuate, handover may occur frequently. Long-term user movements can be aided by mobility prediction, which can greatly optimize a variety of handover KPIs, including the frequency of handovers, handover failure rate, and handover latency. By using precise positioning and sophisticated data processing capabilities built into user devices, Khan and Sha's handover ordering and reduction strategy in [2] demonstrated an effective decrease in the number of handovers and new call blockage. Additionally, the authors in [3] considered mobility prediction for managing predictive changeovers within a control/data plane separation architecture. The results indicated that this method can significantly reduce the number of handover failures and provide considerable savings in signal overhead.

Resource management: In some mobile networks, a lack of resources to handle new call requests can frequently lead to a decline in Quality of Service (QoS) or even call disconnections. In order to increase resource employment, we should build an efficient resource reservation mechanism because bandwidth is a very valuable resource [19]. In order to reduce resource crashes and manage user interference, TPs can reserve resources for users who plan to attach ahead of time based on the mobility prediction data. For example, in [19] and [20], Soh et al. created a dynamic resource reserve model that relies on precise placement to achieve considerable resource performance and appropriate and efficient connections admission control. Mobility prediction is also helpful for estimating how long users will stay in particular TPs. When users' duration of stay in specific TPs is known ahead of time, more intelligent resource management can increase the systems' energy efficiency.

Location-based service pre-configuration and network planning: By transmitting appropriate adverts, local traffic information, instant conversation with nearby users, and vendor suggestions, LBS seeks to give customers improved wireless services and information relevant to their geographic locations. Despite these benefits, real-time geographic location-based services can be outdated, which leads us to predict users' future locations. In-service pre-configuration leverages mobility prediction, enabling customers to identify their future destinations [21]. Additionally, by forecasting traffic, mobility prediction provides the network with a more intelligent and efficient method for managing users.

Data source

Like "Oil in the 18th Century", data is essential and impactful in the twenty-first century. Much pertinent data may be gathered, mostly for following the footprints of moving objects, thanks to the quick advancements in data extraction and storage technology. For example, localization systems like the Global Positioning System (GPS), Radio Frequency Identification (RFID), and smartphone sensors can provide location data. In addition, the Base Station (BS) and server house a significant number of user records. The data needed to forecast mobility varies depending on the context. Crowd movement features are more noticeable in other contexts, such shopping malls, yet cell transition histories are more worthy of our attention in barren areas. Most existing methods and strategies employ context and movement history information to predict user mobility. This research falls into two major categories based on the types of mobility history data [21]. The initial class focuses on personal movement histories. This type of prediction is applied to individual users, not to all users. This kind has the benefit of completely revealing pertinent data hidden within a user's movement history to predict his future movements. However, because it lacks information on a user's own movement profile, it performs badly when anticipating a new user or movements on unfamiliar paths. Popular movement histories provide the basis of the second category [22]. Instead of eliminating random movements from the overall profiles of mobile users, mobility prediction in this case focuses on clusters of similar mobility patterns. The noise of random movements would cause the clustering process' quality to deteriorate. It's interesting to note that various works try to resolve these issues and overcome the shortcomings of two classes. In order to address the lack of individual information and prevent some noise from random movements by certain users, the authors in [23] coupled the mobility rules generated from several similar users with clustering and sequential pattern mining. When the control center of mobile communication systems has access to user mobility data (such as movement, cell transition history, behavior patterns, and environment information), the necessary actions (like reserving resources in advance) can be taken to ensure quality of service (QoS) for users for the duration of their connection. The type of data, the setting, and the desired results must all be taken into consideration while selecting the appropriate procedures. In previous publications, mobility prediction has been extensively utilized, taking into account various contexts and methods. These techniques can't be immediately applied to other situations and settings; they have their own areas of application.

Required information

The data needed to forecast mobility varies depending on the context. The majority of current algorithms and schemes employ context information and movement history data to forecast user mobility. These data can be divided into three categories: mobility patterns, user behavior data, and road topography data.

Road topology information: The fact that the cell border is typically hazy and unevenly shaped because of the features of the terrain and the presence of obstructions that obstruct radio wave propagation has been overlooked in numerous research. The cell shape is disregarded for several mobility prediction techniques under analysis, which could lead to inaccurate results. Road topology data could be incorporated into the prediction algorithm to improve accuracy

because Mobile Terminals (MTs) travel on roads. We will provide a number of prediction algorithms based on road topology in this post. In [19], Soh et al. created a mobility prediction model based on road topology, which necessitates that the serving BS be updated with the user's position on a frequent basis. A database of the roads in its coverage area had to be kept up to date by each BS. Mobility prediction can be used to increase the efficiency of handover. The popularity of digital road maps, which were first created for navigational devices, served as inspiration for the proposal of a dynamic bandwidth reservation method in [20] that used a similar road topology-based prediction scheme. These studies, however, were predicated on the supposition that users possessed a sufficiently accurate locating capability, which may not be the case in practice. Furthermore, a result that is erroneous would be the consequence of placement problems. A graphical model based on learning, which allows for precise predictions of mobile users' movements and speeds within a cell, was presented by Li et al. in [24]. They learned the travel patterns and user type transitional probability after initially classifying mobile users into different user groups based on velocities. Two steps, the anticipation step and the maximization step, were used to predict the transitional probability of the future path and user type at the node. In [25], a novel mobility prediction framework was suggested that uses spatial conceptual maps to accurately forecast the moving route and destination. Additionally, information about users' goals, preferences, and assessed spatial data were used without making any assumptions on the accessibility of users' movement histories. This model's uniqueness is in its integration of geographical conceptual maps and user context into the prediction process. A warm-up phase is not necessary for the suggested approach to gather the history of previously visited locales. As a result, the system predicts users' motions with the same level of accuracy regardless of time. In [2], two more instances of handover management utilizing mobility prediction based on road topology were shown. These evaluations reveal that the real road topology is one of the key tenets of mobility prediction methods. Thankfully, a GPS map can provide the road topology map. Certain models can be trained using a road topology map once it has been acquired. This requirement to initially obtain a road topology map limits the road topology's application, particularly in unfamiliar environments or when obtaining the map is challenging.

User behavioral information: The most crucial information for forecasting a user's future movements is user behavioral data, which can be obtained by breaking down time into several segments and examining users' past behaviors [5]. Typically, almost all literature characterizes user behavior according to the same criteria, which include location, group dynamics, time of day, and length of stay in a particular area. In [26], Wanalertlak et al. created a mobility prediction technique based on behavior, which considers factors such as location, group affiliation, time of day, and duration to effectively forecast next-cell behavior by capturing the short-term and periodic patterns of mobile users. By keeping track of each user's handoff history inside the network and tracking their travel direction in relation to the topological positioning of cells, the next point-of-attachments were predicted. Users with similar mobility patterns are grouped together, and the next cell predictions were specifically focused on the frequency of occurrences rather than signal quality. In [26], Duong et al. investigated a number of mobility prediction algorithms to identify their shortcomings. It acted as a catalyst for the suggestion of a mobility strategy that integrated sequential pattern mining and clustering. From

the movement records of every mobile user inside the service area, regular mobility patterns were identified using sequential pattern mining algorithms. Additionally, to tackle the lack of individual mobility behavior, clustering methods were employed to identify similar mobility patterns among several users, as mobile users on a global scale tended to act as groups due to geographic, social, and friendship constraints. A future location prediction mechanism based on three levels of position detection—intra-sector, sector-to-sector, and cell-to-cell—was put into practice by Zahhad et al. in [25]. There were four primary mechanisms that comprised these three prediction levels. They included moving sequence construction, frequent movement pattern extraction, movement rule discovery, and user movement prediction. Frequent-1 and frequent-k patterns were discovered in the second procedure. Then, movement rules and movement prediction were examined based on the recognized frequent patterns. By modeling users' habitual location behavior using the hierarchical Dirichlet process, McInerney et al. [27] were able to predict users' future whereabouts. One of the main contributions was that the Nokia Lausanne dataset, which included in-depth mobility observations of 38 individuals over a one-year period, demonstrated that the suggested model could solve the issue of cold beginnings. Research indicates that clustering and pattern mining are the predominant algorithms employed in mobility prediction methods based on user behavior. The clustering algorithm can decrease algorithmic complexity by aggregating individuals exhibiting similar behaviors into cohesive groups, owing to the influence of the group effect. Additionally, pattern mining methods are used to find recurring patterns in mobility.

Movement parameters: The core concept of this technique involves predicting the user's movement by analyzing historical factors such as distance, direction of movement, and prior movements. Simple algorithms like the MC and mathematical formulas can take advantage of this information's low dimension. In order to forecast cell transitions in mobile networks, Kuruvatti et al. in [27] utilized data on user movements and contextual information related to their daily activities. The user group direction was likely to be zero in other directions and only two in two other directions according to a diurnal mobility model. Therefore, depending on the user's direction of movement, a user group can move from its current cell into one of the two adjacent cells. The direction of the user's movement, the distance to adjacent cells, and their interaction were considered. Although this plan is simple to execute, it is dependent on the known direction. Additionally, it necessitates a homogeneous cell distribution because a random cell distribution would make it challenging to achieve the cell transition. The plan put forward in [1] was expanded to the Ultra-Dense Network (UDN) by Li et al. in [28], who were motivated by [30] and used the position point procedure to deploy small cells. In contrast to [30], however, the user's travel path was adjusted through Lagrange's interpolation, allowing the slope of the trajectory at any specific location to be utilized in ascertaining the user's direction of movement. Li et al. broadened the scope of the area of interest and took into account smaller cells surrounding a user. Additionally, the effects of user speed, the radius of the area of interest, and small cell density on the suggested schemes' performance were examined.

Prediction output

Prediction outputs can generally be divided into four categories: users' future position, the likelihood that they will transition to a nearby TP later on, the TP they are expected to connect to, and their trajectory of movement.

Moving direction: Given that direction is the most essential element of mobility, it is the primary consideration when anticipating users' movements. Since direction is the most crucial component of mobility, it would be the first thing that springs to mind when predicting users' movements. Commuter mobility is direction-oriented rather than random, and knowing the user's route ahead of time will make it easier to determine where they are going. Kuruvatti et al., for instance, anticipated user cell transitions in [27] by estimating the user group; additionally, [28] forecasted moving direction to determine the place that the user is most likely to visit.

Transition probability: The probability of moving from one state to another is referred to as the transition probability. A user's position information can be described by these states, which can be cells, locations, or other attributes. Prediction outcomes generally encompass several states, each characterized by a distinct transition probability within a specific context [29]. Consequently, the states exhibiting the highest transition probabilities may be selected as the predicted results.

Future locations: How to deliver extremely accurate and dependable localization information anywhere and at any time [30] is an intriguing subject that needs to be solved today because it directly affects how well resource management works. Consequently, the most crucial result of mobility prediction is location data. The current position is generally represented by its longitude and latitude, or Cartesian coordinates, to specify its location. Depending on the techniques used, the prediction results may be a location sequence or a pair of coordinates for the following site. To reduce scanning time and handover latency, the authors in [31] allocate resources in advance of the actual handover within a Long-Term Evolution (LTE) femtocell network by predicting the users' destinations.

User trajectory: The course a moving item takes through space as a function of time is called a trajectory. It can be formally expressed as the object's position over time or as the geometry of the path. The latter is the one that is used the most. Additionally, the first derivative of the trajectory polynomial can be employed to determine the user's direction of movement, provided that the trajectory exists within a two-dimensional plane and can be described by a polynomial. Enhancing communication efficiency can be achieved considerably if the network proactively predicts the needs of mobile users and allocates radio resources at various cells along their travel path to the intended destination [25].

The next cell ID: The next cell ID signifies the identification of the cell that a specific user is likely to visit shortly. In general, as users move around, they may connect to multiple cells, leading to the handover process. Radio resources, such as bandwidth, can be preloaded into the designated cell or cells prior to the user's arrival, provided that we have knowledge of which specific cell or cells will be accessed next. The trajectory attributes of eNodeB, along with a

sequence of prior service eNodeBs, are utilized to predict which service eNodeB will be most beneficial for resource conservation and future network planning.

Performance metrics

Under the literature now under publication, various methods and performance indicators are employed to forecast user mobility. The definition of the KPIs used to measure the effectiveness of the suggested methods and schemes is the main topic of this section. Metrics come in two varieties: application metrics and observable metrics. Application metrics, including handover dropping probability and new call-blocking probability, serve to evaluate the effectiveness of the application of prediction results. In contrast, observable metrics like prediction accuracy and deviation error can be directly obtained from the prediction outcomes.

Prediction accuracy: The primary metric utilized for evaluating forecast accuracy and validating the employed methodology is prediction accuracy. The proportion of correct forecasts to the overall number of predictions is a widely accepted definition of prediction accuracy, as illustrated below. [58].

$$\text{prediciton accuracy} = \frac{\text{number of correct prediction}}{\text{total number of prediction}} \tag{1}$$

Because different writers may have different definitions of prediction accuracy, it should be noted that the criterion of correct predictions vary throughout diverse publications. Contrary to the most widely used definition, for instance, a prediction is considered successful if the user's real state and the expected state (cell, position, etc.) match. But in [30], the authors defined prediction accuracy as the ratio of the length of points in a trajectory sequence to the hit rate (the latter is set to 1 if the Euclidean distance between a predicted point and the corresponding practical point is smaller than a threshold).

Deviation error: Determining the accuracy of a prediction can be challenging at times. To measure the average magnitude of errors within a set of predictions, irrespective of their direction, the concept of deviation error is introduced. For example, Qiao et al. used the following definition of deviation in [32–33].

$$Deviation = \frac{1}{N}\sum_{i=1}^{N} d_i \tag{2}$$

$$d_i = \sqrt{\left(x_{i,\text{pre}} - x_{i,\text{pra}}\right)^2 - \left(y_{i,\text{pre}} - y_{i,\text{pra}}\right)^2} \tag{3}$$

where the practical position (xi, pra; yi, pra) and the projected location (xi, pre; yi, pre) are separated by the Euclidean distance d, di. Additionally, N is the total number of forecasts. With k representing the discrete time in the mobility tracking issue, which was described as (in m): Similarly, Root Mean Square Error (RMSE) of position and speed was developed in [2] to evaluate the proximity of the estimated trajectory (^xk; ^yk) to a specified trajectory (xk; yk).

$$RMSE = \sqrt{\frac{1}{N_{mc}} \sum_{m=1}^{N_{mc}} (\hat{x}_k - x_k)^2 - (\hat{y}_k - y_k)^2} \qquad (4)$$

where the number of Monte Carlo runs is indicated by Nmc. The estimated speed (x_ k; y_ k) can be substituted into (5) to create the speed RMSE (in m/s).

Handover dropping probability (HDP): The proportion of handover drops relative to the total number of handovers is referred to as the (HDP). The outcome of the prediction functions as an indirect evaluation of the prediction methodologies, as it is employed in the management of handovers. Handover dropping occurs in the present LTE communication system if the destination cell's Signal-to-Interference-plus-Noise-Ratio (SINR) falls below a threshold, significantly impairing user comfort. In order to effectively lower the handover dropping probability, Soh et al. presented a predictive bandwidth reservation strategy in [20]. The HDP is most frequently defined as follows [32].

$$HDP = \frac{\text{number of dropping handover}}{\text{total number of handover}} \qquad (5)$$

New call blocking probability: A resource reservation approach that relies on predicting user mobility establishes a new probability of call blocking. Dropping a call that is in progress is clearly more annoying to the user than blocking a new call [20]. The call-blocking probability would unavoidably rise in resource reserve schemes in order to reach the HDP aim. A reduced probability of blocking new calls in mobility prediction signifies an enhanced performance in mobility forecasting, given the same HDP. Individuals attempting to access this cell may find it advantageous to implement a resource reservation strategy that relies solely on upcoming handover predictions, thereby reducing unnecessary blocking of new calls.

Mobility prediction schemes

This section provides detailed summaries of the most advanced algorithms that examine and evaluate historical mobile user movements in order to anticipate mobility. Since they are the most widely used and significant approach for mobility prediction systems, we have described the MC, HMM, ANN, and Bayesian Network (BN) as prediction methodologies. These strategies aim to enhance network performance by reducing user latency, minimizing blocking and dropping probabilities, and simultaneously improving prediction accuracy.

Markov chain: A stochastic process that possesses the "memoryless" Markov property is called a Markov chain. What is known as a "Markov chain" is the series of random variables that such a process passes through. Only the dependence of neighboring periods (as in a "chain") defines serial states according to the Markov property. As a result, it can be applied to describe systems that consist of a series of interconnected events, where the system's current state or its previous states determine what happens next. The Markov chain has been utilized for an extended period to predict user movement within wireless networks, owing to the characteristics of the Markov

property. Both [34] and [35] used the original Markov chain technique to forecast user movement, but the former investigated other types of user movement, such as linear, stationary, random, and patterned. The user's movement would begin at a specific location based on the Markovian characteristic, and they might go to any other cells or positions or remain in their current position. The current state, not the past states, determines the transition probability between states. Figure 2 depicts a three-cell configuration within a cellular network. In this context, the three cells, designated as A, B, and C, correspond to the three states of a Markov chain. It is important to emphasize that these states are not confined to cells alone; they may also encompass various locations or movements.
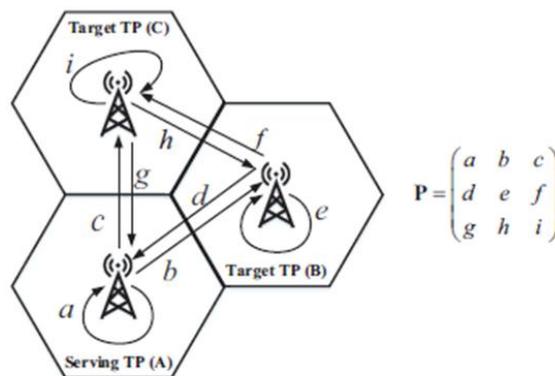
Figure 2. Transition probability [31]

In this case, the transition probability matrix's value is preset. Assuming, for example, that a user is attached to cell A, it may remain there throughout the subsequent time slot and move to either cell B or cell C, with probabilities of a, b, and c, respectively. Additionally, cell B and cell C have transition probabilities of d, e, f, and g, h, i, respectively. A, B, and C must all have a total probability of 1.

The transition probability matrix is denoted by P. Formula 6 [31] is used to determine the subsequent transition probability (Pn) from the serving cell to the target cell:

$$\mathbf{P}_n = [\mathbf{P}_{n-1}] * [\mathbf{P}] \tag{6}$$

Where P_{n-1} is denoted as the current transition probability matrix and n is the number of transition states. Similar to this, a Markov chain-based mobility prediction technique was put out in [31] to forecast user movement during femtocell deployment in order to achieve the objective of a smooth and quick handover. The history of the user's movements is stored in a log file that includes the user ID, time, and location. The location denotes the base station ID to which the user is connected at a particular moment, while the time signifies the date and time of the user's connection to that specific BS. A transactional database is created from the log file to establish a connection between the source and destination base stations. The database may be used to determine which base station is visited the most frequently and to extract the transition matrix. By using embedded association rules, such as the temporal rule and the universal behavior rule, the authors of [36] suggested improved Markov chain algorithms. First, the historical database's user trajectories are taken out. The transition data employed in the online mode is used to anticipate users' subsequent actions during the training phase. When it

comes to unknown users during the training phase, the performance of the suggested augmented Markov chain is subpar; however, the system's performance can be raised by adding additional forecasting algorithms, such as particle or Kalman filters. Gambs et al., however, pointed out in [21] that the usual Markov chains have neglected previously visited places, which would have a detrimental effect on prediction ability. A unique Markov chain called n-MMC, which takes into account the n previously visited places, was offered as a solution to this constraint. However, the transition probability matrix would become increasingly complex due to the consideration of the n-previous locations in the extended version of standard Markov chains. Therefore, this method is not suitable for a network with a lot of BSs, such as a femtocells network. The findings from the previous research suggest that the straightforward nature of Markov chain-based methods renders them easy to understand and implement. However, the performance of these schemes is predominantly influenced by the transition probability matrix, making its acquisition the most critical element of the Markov chain-based approach. The analysis indicated a significant correlation between the performance of the Markov chain-based scheme and the duration required for data collection. Furthermore, when the state space is vast, it is exceedingly difficult to obtain the transition matrix due to the low extensibility of Markov chain-based techniques.

HMM: The sole parameters in smaller Markov models (such as a Markov chain) are the state transition probabilities as the observer can see the state directly. Another set of stochastic processes that generate the series of observed symbols is the only way to view HMM, a double stochastic process with an underlying stochastic process that is hidden and not observable [36]. Hidden states and observable states represent the two categories of states present in HMM. Generally, HMM demonstrates superior performance compared to Markov chains in forecasting user mobility, as observable states offer additional insights into the hidden states.

Artificial Neural Networks (ANN): ANNs are a family of models used in machine learning and cognitive science that are modeled after biological neural networks, specifically the brain and central nervous systems of animals. These models serve to estimate or approximate functions that are generally unknown and may depend on a substantial number of inputs [37]. ANNs are typically depicted as networks of interconnected neurons that communicate with one another, as seen in Fig. 3. Because the connections' numerical weights can be adjusted according to experience, neural nets are input-adaptive and learning.
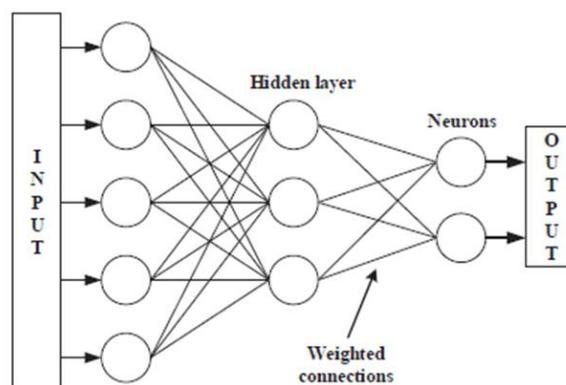
Figure 3: Structure of ANN with a hidden layer [37]

Parija et al. introduced a multilayer neural network model in [38] that leverages previously predicted data to anticipate the future locations of subscribers. The movement patterns were characterized by a pair of values representing distance and direction (dsj; drj). In this context, drj indicates the possible direction a user might take during a specific time interval, while dsj (measured in meters) denotes the distance traveled by the user during that period. Only users with consistent and regular movement patterns may be predicted by this investigation due to the definition of movement patterns' limitations. The user's location, velocity, acceleration, and orientation all logically affect the movement, which may seem random. Since continuous user positions can be used to determine speed, acceleration, and direction, Liou et al. developed a neural location predictor in [38] that uses three location updates from the closest past to predict a future location. Cartesian coordinates were used to define user positions in [39]. Higher computer efficiency was achieved by treating (x; y) of a coordinate individually, in contrast to the majority of previous efforts. Furthermore, a dual-layer cell architecture was established, and the mobility prediction processes commenced solely when users approached the cell boundary. This could enhance bandwidth reservation performance. The authors of [38] gave us two neural network models—the Multilayer Perceptron Back Propagation Neural Network (MLP-BP) and the Functional Link Neural Network (FLNN)—by examining their primary distinctions in order to address the crucial problem of mobility management in mobile networks. To predict the position of the mobile host, the movement pattern data was utilized to train these two models. FLNN consists of a single-layer configuration that enhances the input pattern through nonlinear functional expansion. In contrast, MLP-BP operates as a feed-forward network utilizing backpropagation, which necessitates considerable computational time for the learning process of the network. Furthermore, it was demonstrated that the more straightforward FLNN performed better than MLP-BP in terms of classification accuracy and performance error. Likewise, [39] presented comparisons between MLP-BP and Polynomial Perceptron Network (PPN). On the basis of the Weierstrass approximation theorem, PNN was constructed. It claimed that a polynomial could uniformly estimate any function that is continuous in a closed interval over that interval within any specified tolerance. While PNN's computational cost was superior to MLP-BP's due to its reduced computational complexity and fewer biases and weights, MLP-BP was shown to perform better than PNN. Furthermore, an ANN model takes into account a lot of other parameters. A basic road model with a preset likelihood of users moving was constructed by Wickramasuriya et al. in [40]. They introduced a technique that uses Recurrent Neural Networks (RNN) to accurately forecast which BS a user will most likely connect to next. Its training technique reached an accuracy of over 98% by taking RSS distribution into account rather than user positions. Backpropagation is the most popular method in artificial neural networks (ANNs), which are well-researched algorithms renowned for their adaptable and self-organizing properties. The user position is identified as the most critical parameter in artificial neural network (ANN)-based mobility prediction systems, serving as both an input and an output, as noted in references [38] to [40]. The ultra-dense deployment of small cells within the system is emerging as a definitive trend for the future as mobile communication continues to evolve. This trend relates to the processes of obtaining user positions, such as the fingerprint scheme, even though it would not directly affect ANN-based

schemes. Furthermore, ANNs are renowned for their astounding computational complexity, particularly when they have several hidden layers. This type of network is known as a deep neural network, and it takes a lot of training time to adjust the weights of the neurons.

BN: A directed acyclic graph serves as the foundation for Bayesian Networks, which are probabilistic graphical models that illustrate a set of random variables along with their conditional dependencies [40]. In a complex setting, it offers significant benefits for resolving uncertainty and determining the relationship between various factors. The BN has consequently been extensively used in numerous fields. For example, the BN can be used to calculate the likelihood that a user would be present at a certain location based on prior user movement data. In [41], Zhang et al. presented a BN-based position prediction model. Numerous constraining criteria, such as cell topology information, road typology information, and user movement information, are taken into consideration, whereas the majority of known prediction models only include a portion of the predictive factors. The following procedure is used to create the BN structure, as seen in Fig. 4: 1) Cell and Road Topology (CRT) construction based on cell environment, 2) Road State Transition (RST) construction based on CRT, 3) Probability Distribution Network (PDN) construction based on RST, and 4) BN construction based on PDN. Additionally, the distribution of predicting parameters is examined. It is assumed that the mobile user's state transition chain is conditionally independent of every state. The vector of predictive factors associated with each node in the transition path can be utilized to calculate the probability of a state transition chain in accordance with BN theory. Furthermore, the present location, anticipated position, and current speed are utilized to forecast the duration of stay within the target cell.
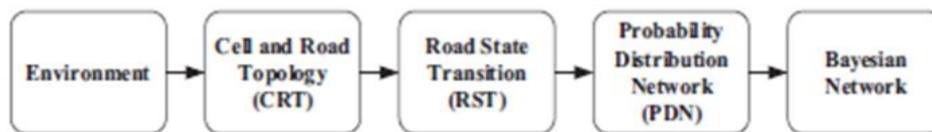
Figure 4: Construction of BN

Liu et al. presented a unique method based on BN for predicting the future location of a moving object under uncertainty. The methodology encompasses space-partitioning techniques, widely utilized for region extraction, the transformation of trajectory and region sequences, frequent sequential pattern mining, and the creation of BN, paralleling the approach outlined in [41]. Popular zones were considered random variables and were utilized to approximate the trajectory sequences of moving objects. The BN was constructed using the frequent region patterns, and the arcs linking the BN's nodes were constructed using the traversal paths of regions. Additionally, a number of techniques were put out to convert the trajectory data into the BN structure. Dynamic Bayesian Networks (DBNs) are BNs that model the array of variables. In order to anticipate the future location, DBN models that took into account locations, time of day, day of the week, and their combinations were proposed in [42]. To accommodate various circumstances, three combination models were created utilizing ensemble, highest probability, and least entropy. Two separate mobility data sets, namely call detail records data and Nokia mobile data, were utilized to conduct a comparative analysis of these three models. To enhance its performance, Bayesian Networks can be integrated with other models, such as neural

networks. A distinctive hybrid Bayesian neural network model for position prediction within cellular networks was introduced in [42]. The efficacy of the proposed model was evaluated against several widely used neural network methodologies, such as back-propagation, Elman, robust, Levenberg-Marquardt, and one-step secant models. Additionally, different parallel implementation approaches on mobile devices were investigated. It was demonstrated that the suggested hybrid model performed better than other tried-and-tested standard models. As a result, we may discover that BN is a kind of probabilistic graphical model that computes probabilities using Bayesian inference. By using edges in a directed graph to express conditional dependency, BN seeks to model conditional dependence and, consequently, causality. By using these correlations, one can effectively employ factors to deduce information about the graph's random variables. However, precise inference is highly computationally demanding in big networks, thus techniques to minimize computation are required. Furthermore, the synthesis of BN depends on the cell environment. The cell environment would become more complex as a result of the dense deployment of tiny cells, making BN construction more challenging. A thorough explanation of other methods, including Kalman Filter (KF), [43], Support Vector Machine (SVM), time series model, biomedical algorithms, and so forth, will not be expanded due to the article's length constraints.

Evaluation of mobility prediction schemes

ANN, BN, HMM, and MC are a few of the mobility prediction techniques that have been put forth in several studies and documents. These methods seek to increase network performance in terms of device latency and dropping probability while also improving prediction accuracy. Every one of them has benefits and drawbacks. Although the ANN approach performs well, it typically has a high training complexity that makes it difficult to understand and susceptible to over-fitting issues. Through the use of factors, the BN model can be made resistant to inference on random variables in graphs. The method can be integrated with additional techniques, such as ANN, to improve predictive accuracy. However, creating a manageable approximation for probabilistic inference is highly intricate, and the complexity increases significantly when dealing with a complicated cellular environment. HMM approaches are categorized into various states according to their characteristics. Since its fundamental idea requires that more information about hidden states be gleaned from observable states, this model is more sophisticated than the Markov model. Dependencies between two states may cause problems for the mobility prediction process, particularly if the system must handle a large volume of user data. The Markov chain technique was chosen as our primary method for forecasting user mobility in mobile network systems. Particularly for discrete sequence modeling, MC is a stochastic process with an easy-to-implement structure. The short memory principle is another benefit of this strategy. It implies that the mobility forecast of the subsequent state can be calculated by examining one, two, or three prior states. In a variety of settings with high user movement, such as big cities and small cities, this element might be quite important.

CONCLUSION

One of the most important and required strategies for resolving some major issues and constraints, such handover in mobile network systems and applications like LBS, is mobility

prediction. We made an effort to clarify the mobility prediction structure, going into detail about each stage, such as the data source and application, separately. The performance of several mobility prediction models and schemes was assessed and analyzed using a variety of criteria, such as HDP and prediction accuracy. This study examines and presents the most important and widely used algorithms and techniques, including as Markov, BN, and HHM. In this paper, mobility prediction systems were evaluated and examined.

REFERENCES

[1] P. Fazio, F. D. Rango, and M. Tropea, "Prediction and QoS enhancement in new generation cellular networks with mobile hosts: a survey on different protocols and conventional/unconventional approaches," IEEE Commun. Surveys Tuts., vol. 19, no. 3, thirdquarter 2017, pp. 1822–1841.

[2] A. N. Khan and S. X. Jun, "A new handoff ordering and reduction scheme based on road topology information," in International Conference on Wireless Communications, Networking and Mobile Computing, Sep. 2006, pp. 1–4.

[3] A. Mohamed, O. Onireti, S. A. Hoseinitabatabaei, M. Imran, A. Imran, and R. Tafazolli, "Mobility prediction for handover management in cellular networks with control/data separation," in IEEE International Conference on Communications (ICC), Jun. 2015, pp. 3939–3944.

[4] H. Zhang and L. Dai, "Mobility Prediction: A Survey on State-of-the-Art Schemes and Future Applications," in IEEE Access, vol. 7, pp. 802-822, 2019.

[5] J. Jiang, Ch. Pan, H. Liu and G. Yang, "Predicting Human Mobility Based on Location Data Modeled by Markov Chains," 2016 Fourth International Conference on Ubiquitous Positioning, Indoor Navigation and Location Based Services (UPINLBS), pp. 145-151, 2016.

[6] V. Ayumi, I. Nurhaida, "Prediction Using Markov for Determining Location of Human Mobility", International Journal of Information Science & Technology, Vol. 4, pp. 1-6, 2020.

[7] L. Menz, R. Herberth, C. Luo, F. Gauterin, A. Gerlicher and Q. Wang, "An Improved Method for Mobility Prediction Using a Markov Model and Density Estimation," 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, pp. 1-6, 2018.

[8] K. -L. Yap and Y. -W. Chong, "Optimized Access Point Selection with Mobility Prediction Using Hidden Markov Model for Wireless Network," 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), pp. 38-42, 2017.

[9] G. A. Montoya and Y. Donoso, "A Prediction Algorithm Based on Markov Chains for Finding the Minimum Cost Path in A Mobile Wireless Sensor Network," 2018 7th International Conference on Computers Communications and Control (ICCCC), pp. 169-175, 2018.

[10] J. Ding, H. Liu, L. T. Yang, T. Yao and W. Zuo, "Multiuser Multivariate Multi-order Markov-Based Multimodal User Mobility Pattern Prediction," in IEEE Internet of Things Journal, vol. 7, no. 5, pp. 4519-4531, 2020.

[11] J. F. Zaki, A. Ali-Eldin, Sh. E. Hussein, S. F. Saraya, and F. Areed, "Traffic Congestion Prediction Based on Hidden Markov Models and Contrast Measure", Ain Shams Engineering Journal, vol. 11, pp. 535-551, 2020.

[12] X. Zheng et al., "Endurance Prediction Based on Hidden Markov Model and Programming Optimization for 28nm 1Mbit Resistive Random-Access Memory Chip," in IEEE Electron Device Letters, vol. 44, no. 6, pp. 919-922, 2023.

[13] S. Han, "Prediction of College Students' Autonomous Learning Ability Based on Markov Chain Prediction Algorithm," IEEE 5th Eurasian Conference on Educational Innovation (ECEI), pp. 216-220, 2020.

[14] A. Ladekar, B. Mohol, A. Gaikwad, S. Shingade, A. Kulkarni and Y. Naval, "PULM: Prediction of User's Location using Machine Learning with Markov Model," 6th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1195-1199, 2022.

[15] H. Shi, Y. Li, X. Zhou, Ch. Zhang, and V. Kostakos, "Semantics-Aware Hidden Markov Model for Human Mobility" IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 3, pp. 1-8, 2021.

[16] H. Wang, S. Zeng, Y. Li, P. Zhang and D. Jin, "Human Mobility Prediction Using Sparse Trajectory Data," in IEEE Transactions on Vehicular Technology, vol. 69, no. 9, pp. 10155-10166, 2020.

[17] R. Kudo, M. Cochrane, K. Takahashi, T. Inoue And K. Mizuno, "Experimental Validation of Link Quality Prediction Using Exact Self-Status of Mobility Robots in Wireless LAN Systems", IEICE Transactions on Communications, vol. 103 no.12, pp.1385-1393, 2020.

[18] G. Yuan, T. Pan and X. Wu, "Joint Adaptive Mobility Prediction and Signal Strength Prediction Based Cell Selection Algorithm in Ultra-Dense Networks," IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 1244-1249, 2022.

[19] W.-S. Soh and H. S. Kim, "QoS provisioning in cellular networks based on mobility prediction techniques," IEEE Commun. Mag., vol. 41, no. 1, Jan. 2003, pp. 86–92.

[20] W. S. Soh and H. S. Kim, "A predictive bandwidth reservation scheme using mobile positioning and road topology information," IEEE/ACM Trans. Netw., vol. 14, no. 5, Oct. 2006, pp. 1078–1091.

[21] G. Xu, S. Gao, M. Daneshmand, C. Wang, and Y. Liu, "A survey for mobility big data analytics for geolocation prediction," IEEE Wireless Commun., vol. 24, no. 1, February 2017, pp. 111–119.

[22] Y. Jia, Y. Wang, X. Jin, and X. Cheng, "Tsbm: the temporalspatial bayesian model for location prediction in social networks," in IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), vol. 2, Aug. 2014, pp. 194–201.

[23] T. V. T. Duong and D. Q. Tran, "A fusion of data mining techniques for predicting movement of mobile users," J. Commun. Networks, vol. 17, no. 6, Dec. 2015, pp. 568–581.

[24] H. Li and G. Ascheid, "Mobility prediction based on graphical model learning," in IEEE Vehicular Technology Conference (VTC Fall), Sept 2012, pp. 1–5.

[25] A. Nadembega, A. Hafid, and T. Taleb, "A destination and mobility path prediction scheme for mobile networks," IEEE Trans. Veh. Technol., vol. 64, no. 6, Jun. 2015, pp. 2577–2590.

[26] T. V. T. Duong and D. Q. Tran, "Mobility prediction based on collective movement behaviors in public wlans," in Science and Information Conference (SAI), Jul. 2015, pp. 1003–1010.

[27] N. P. Kuruvatti, A. Klein, J. Schneider, and H. D. Schotten, "Exploiting diurnal user mobility for predicting cell transitions," in IEEE Globecom Workshops (GC Wkshps), Dec. 2013, pp. 293–297.

[28] B. Li, H. Zhang, and H. Lu, "User mobility prediction based on lagrange's interpolation in ultra-dense networks," in IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Sep. 2016, pp. 1–6.

[29] A. Ulvan, M. Ulvan, and R. Bestak, "The enhancement of handover strategy by mobility prediction in broadband wireless access," in Proceedings of the networking and electronic commerce research conference (NAEC 2009), American Telecommunications Systems Management Association Inc., 2009, pp. 266–276.

[30] A. Boukerche, H. A. B. F. Oliveira, E. F. Nakamura, and A. A. F. Loureiro, "Vehicular Ad Hoc networks: a new challenge for localizationbased systems," Comput. Commun., vol. 31, no. 12, 2008, pp. 2838– 2849.

[31] S. H. Ariffin, N. Abd, N. E. Ghazali et al., "Mobility prediction via Markov model in LTE femtocell," International Journal of Computer Applications, vol. 65, no. 18, 2013.

[32] H. Zhang and L. Dai, "Mobility Prediction: A Survey on State-of-the-Art Schemes and Future Applications," in IEEE Access, vol. 7, pp. 802-822, 2019.

[33] J. Jiang, Ch. Pan, H. Liu and G. Yang, "Predicting Human Mobility Based on Location Data Modeled by Markov Chains," 2016 Fourth International Conference on Ubiquitous Positioning, Indoor Navigation and Location Based Services (UPINLBS), pp. 145-151, 2016

[34]   A. Ulvan, M. Ulvan, and R. Bestak, "The enhancement of handover strategy by mobility prediction in broadband wireless access," in Proceedings of the networking and electronic commerce research conference (NAEC 2009), American Telecommunications Systems Management Association Inc., 2009, pp. 266–276.

[35]   S. H. Ariffin, N. Abd, N. E. Ghazali et al., "Mobility prediction via Markov model in LTE femtocell," International Journal of Computer Applications, vol. 65, no. 18, 2013.

[36]   A. Hadachi, O. Batrashev, A. Lind, G. Singer, and E. Vainikko, "Cell phone subscribers mobility prediction using enhanced Markov chain algorithm," in Proc. IEEE Intelligent Vehicles Symp, Jun. 2014, pp. 1049–1054.

[37]   J. J. Hopfield, "Artificial neural networks," IEEE Circuits Devices Mag., vol. 4, no. 5, 1988, pp. 3–10.

[38]   S. Parija, P. K. Sahu, S. K. Nanda, and S. S. Singh, "A functional link artificial neural network for location management in cellular network," in Proc. Int. Conf. Information Communication and Embedded Systems (ICICES), Feb. 2013, pp. 1160–1164.

[39]   S. Parija, S. Nanda, P. K. Sahu, and S. S. Singh, "Novel intelligent soft computing techniques for location prediction in mobility management," in Students Conference on Engineering and Systems (SCES), Apr. 2013, pp. 1–4.

[40]   D. S.Wickramasuriya, C. A. Perumalla, K. Davaslioglu, and R. D. Gitlin, "Base station prediction and proactive mobility management in virtual cells using recurrent neural networks," in IEEE Wireless and Microwave Technology Conference (WAMICON), Apr. 2017, pp. 1–

[41]   Y. Zhang, J. Hu, J. Dong, Y. Yuan, J. Zhou, and J. Shi, "Location prediction model based on Bayesian network theory," in IEEE Global Telecommunications Conference (GLOBECOM), Nov. 2009, pp. 1–6

[42]   M. Dash, K. K. Koo, J. B. Gomes, and S. P. Krishnaswamy, "Next place prediction by understanding mobility patterns," in IEEE International Conference on Pervasive Computing and Communication Workshops, 2015, pp. 469–474

[43]   H. Feng, C. Liu, Y. Shu, and O.W. Yang, "Location prediction of vehicles in VANETs using a kalman filter," Wirel. Pers. Commun., vol. 80, no. 2, Jan. 2015, pp. 543–559.

RESUME

Shiva Derogar KHACHEKINI

Shiva Derogar received a BSc degree in computer engineering (software) from the Islamic Azad University of Lahijan (Iran) 2012. In October 2013, he joined the Department of Computer Engineering, University of Guilan, Iran, as a full-time MSc. Student. He received the MSc degree in 2015. She was a Ph.D student under advisor Dr. Gürcü Öz in the Computer Engineering Department of EMU University. Her research interests include wireless Networks, Security of Networks, cryptography algorithms, and mobile communications.

Assoc. Prof. Dr. Gürcü ÖZ

Gürcü Öz received the B.S., and M.S. degrees in Electrical and Electronics Engineering from Eastern Mediterranean University, in 1993 and 1996, respectively. She received a Ph.D. degree in Computer Engineering from Eastern Mediterranean University, in 2001. From 1994 to 1996, she worked as a Research Assistant and Lecturer in the Department of Electrical and Electronic Engineering, at Eastern Mediterranean University. From 1996 to 2001, she worked as a Research Assistant and Lecturer in the Department of Computer Engineering, at Eastern Mediterranean University. From 2001 to 2016, she worked as an Assistant Professor, in the Department of Computer Engineering, at Eastern Mediterranean University. From 2016 to the present, she worked as an Associate Professor in the Department of Electrical and Electronic Engineering, at Eastern Mediterranean University. Her current research interests are Computer Networks, Distributed Systems, Advanced System Simulations, Wireless Networks, Cloud Computing, Network Security, and Information Security.

Prof. Dr. Ali Hakan ULUSOY

Ali Hakan Ulusoy was born in Eskişehir, Turkey, on June 3, 1974. He graduated from the double major program of the Department of Electrical and Electronic Engineering (EEE) and the Department of Physics at Eastern Mediterranean University (EMU) in 1996. He received his M.S. and Ph.D. degrees in EEE, and EMU in 1998 and 2004, respectively. Then, he joined the Department of Information Technology, EMU, in 2004 as an instructor where he became a visiting assistant professor in 2005, an assistant professor in 2007, an associate professor in 2013, and a professor in 2019. He worked as the assistant director from January 2012 to August 2017 and the acting director from August 2017 to December 2019 at the Institute of Graduate Studies and Research (IGSR), EMU. He is currently working as the director of IGSR since January 2020. His research interests include wireless communications, receiver design, channel estimation, fuzzy systems, wireless networks, cloud computing, millimeter wave communications, and healthcare system

development.

A COMPREHENSIVE STUDY AND INVESTIGATION OF VANET ROUTING PROTOCOLS IN ITS

Mojtaba Ayoubi MOBARHAN[1], Muhammad SALAMAH[2]

INTRODUCTION

The Internet of Things (IoT) is an advanced technology that facilitates the exchange of information between machines and humans by incorporating devices such as infrared sensors, global positioning systems, laser scanners, and various other sensing technologies into physical equipment [1]. IoT has become a new network paradigm platform that enables connections between different physical items. In addition to being a communication network, the IoT combines technology for processing, communication, and sensing. The three primary levels of an IoT system are the application, network technology, and sensing layers. The first layer, referred to as the sensing layer, is in charge of gathering data and information; the second layer, known as the network layer, manages the network connection, exchanges data with the application layer, and does so [1] [2]. To enhance the security and network performance of IoT systems, numerous experts and researchers in computer, electronic, and information technology fields have recently focused on mobile systems and technologies.

IoT systems may be in charge of gathering, sharing, keeping an eye on, and managing data in a variety of contexts, including factories, buildings, and natural ecosystems. Beyond merely sensing, this technology can also be responsible for carrying out a number of tasks that are necessary for user convenience and security. IoT systems can be broadly divided into a number of sectors, including manufacturing, energy management, medical and healthcare, and transportation, based on a variety of features and aspects. Vehicle transportation regulation and monitoring is one of the most important components of a smart city [3]. In this post, our primary IoT application is a smart transportation system. In order to monitor and manage the transportation network, Information Technology (IT) or transportation cyber-physical systems integrate communication and computing. ITS seeks to improve the transportation infrastructure's availability, efficiency, safety, and dependability. Fleet management systems, smart parking systems, traffic assessment and management, in-vehicle and on-road safety management, and emergency management are just a few of the numerous subsystems and applications that make up ITS. The goal of ITS implementation is urban populations and their transportation. One of the key elements of this system is thought to be connected cars. Automakers such as Tesla and Volkswagen want to introduce automobiles that are outfitted with a gadget that facilitates communication between vehicles. To link mobile vehicles and network

[1] Eastern Mediterranean University, North Cyprus, via Mersin 10, Turkey, Orcid: https://orcid.org/0009-0005-9161-1521, mojtaba.ayoubi@yahoo.com
[2] Eastern Mediterranean University, North Cyprus, via Mersin 10, Turkey, Orcid: https://orcid.org/0000-0001-9677-1624, muhammed.salamah@emu.edu.tr

access points, ITS and its many applications use a variety of network and communication technologies [4]. VANET is among the most significant and useful varieties of these technologies. Vehicles act as nodes in VANETs, a form of ad hoc network that facilitates wireless communication among vehicles. There are three primary classes of this technology. First, cars share information regarding their location, speed, stability, braking, and heading when they communicate with one another. This kind of communication is named Vehicle to Vehicle (V2V). The foundation of V2V technology is specialized short-range communication, which is nearly identical to Wi-Fi. In VANET simulation, a lot of ad-hoc network protocols are employed. A vehicle may be completely automated in VANET, or it may have numerous capabilities that allow it to interact with other components of the intelligent transportation system. In order to promote road safety, communication vehicles can also talk with pedestrians and traffic poles using Vehicle to Infrastructure (V2I) communication. Through a network, the V2I technology enables automobiles to communicate with road infrastructure. The following components should be included in the minimal V2I system: Road Side Units or Equipment (RSU or RSE), vehicle on-board units or equipment (OBU or OBE), and a secure communication channel. The car is equipped with OBUs in this type of ITS. From a practical standpoint, this is the unit that connects the car to the infrastructure. Another cutting-edge technology is Vehicle to Everything (V2X), which uses short-range wireless signals to connect cars to the traffic and everything else in their immediate vicinity. Two fundamental elements of the V2X system are V2V and V2I. V2X is intended to improve safety and avoid collisions. Additional information is provided by this system, including local accidents, road conditions, and weather. There are a number of other uses for VANET, such as smart parking, location positioning, traffic management, payment services, and entertainment services, which can assist ITS in reducing vehicle traffic, even though the main goal of VANET deployment is to prevent the rising number of road accident deaths, which claim millions of lives worldwide [5].

There are a number of areas within VANET where research is ongoing, including manufacturing, security, traffic management, congestion control, and routing protocols. The main challenge facing VANETs currently is the frequent breakage of paths, which occurs due to the highly dynamic topology created by moving vehicles. The movement of intermediate or endpoint nodes can lead to interference, necessitating routing protocols that are aware of constant changes in topology. Rapid changes in topology and frequent disconnections of nodes. make it difficult to build an appropriate routing system for data exchange between cars. V2V traffic information exchange is a major challenge for VANETs. In certain applications, information is taken on-demand (pull model), while in others, it is proactively distributed through broadcast (push model) [6]. Despite sharing many technical characteristics and requirements with a Mobile Ad Hoc Network (MANET), VANET faces a various challenges and issues. VANETs face challenges related to traffic, safety, and user applications, each of which has particular design specifications. When efforts are made to establish VANET in an effort to enhance driver behavior and lower the number of fatalities brought on by auto accidents, many problems occur. The following are the primary issues with VANET and the main challenges from a technological standpoint:

Signal fading: One of the primary issues that can lower VANET performance is objects positioned as barriers between two communicating vehicles. These barriers may consist of a

single car or a number of structures spaced out along a route. The lack of a central coordinator that oversees communications between nodes and is in charge of controlling contention and capacity is another significant issue with the VANET.

Connectivity: The time needed to extend the life of the connection communication should be as long as feasible because of the rapid mobility and swift alterations in topology that cause networks to fragment often. By raising the transmission power, this task can be completed, but throughput degradation may result. Since a VANET's limited effective network width results in poor connectivity in node-to-node communication, connectivity is therefore seen as a crucial issue in VANETs.

Security and privacy: One of the primary issues with VANET is maintaining a suitable balance between security and privacy. For the recipient, receiving reliable information from its source is crucial.

Routing protocol: Designing an efficient routing protocol that can deliver packets in the shortest time possible while minimizing packet loss is a major challenge in VANETs, primarily due to the mobility of devices and the rapid changes in network topology [7].

Our primary contributions

The key contributions of our paper are outlined below:

- ✓ The important applications and structure of IoT are presented
- ✓ We provide a thorough summary of VANET routing protocols.
- ✓ We review the main objectives of the ITS applications.
- ✓ The key challenges of the VANET technology especially in the ITS network are mentioned

Paper organization

The structure of the paper is as follows: Section 2 presents the most significant related works. Section 3 presents some of the most significant IoT applications. The features of ITS and the communication technologies are detailed in section 4, while the VANET routing schemes are discussed in section 5. Classifications of routing model are presented and some important routing protocols are explained briefly. The primary challenges of VANET are outlined in section 6. In conclusion, we present our final thoughts in section 7.

RELATED WORKS

Several researchers and businesses study and operate with VANET architectures, applications, and schemes. One of the most important and alluring study areas in the context of computer technology and industry is VANET routing protocol. [8] examined several VANET routing techniques. Network Simulator 3 (NS3) was used in this work to analyze routing performance, while Simulation of Urban Mobility (SUMO) was utilized to generate traffic mobility. This paper explored the effectiveness of topology-based routing protocols in VANETs, focusing on transmission power, MAC PHY overhead, packet delivery, and end-to-end delay. The findings indicate that selecting the most suitable algorithm relies on various critical factors pertinent to

VANET networks and applications. Based on several real-world traffic circumstances, the authors investigated how the most popular realistic mobility models affected the performance of the VANET [9]. The performance of various routing protocols was assessed using QoS metrics such as average goodput, average overhead, and packet delivery ratio. In this experiment, NS3 and SUMO were employed.

A Stable Multi-Hop Clustering protocol (SDSMC) for VANETs was designed for successful delivery in [10]. This technique featured multi-hop communication between the cars, a cluster architecture, and an appropriate clustering formation and cluster head selection mechanism. The quality of the user's communication is maximized when these processes are present. Researchers used NS3 simulation to create this VANET model, while SUMO software was used to produce vehicle mobility. For SDSMC-VANETs, key metrics such as energy efficiency, packet delivery ratio, end-to-end delay, and routing overhead were evaluated. The outcomes were contrasted with the previous models. Analysis indicated that this model is more efficient and achieves a higher delivery ratio compared to other methods. Multiple VANET routing protocols, including AODV, Ad-hoc Distance Vector (ADV), and DSDV, were the main emphasis of the authors in [11]. The study evaluated the performance of these routing protocols across several parameters, such as average throughput, packet delivery ratio, latency, and total energy consumption. Its findings led to the suggestion of the best routing technique for places with extremely high traffic volumes. Numerous routing protocols that can maximize the needs of VANET systems were investigated by researchers [12]. In this article, they also contrasted cluster-based and position-based routing methods. Cluster-based routing is more practical and dependable than position-based routing, according to the cluster head selection. This study also clarified which protocol is more appropriate for VANET. In order to create the ideal VANET architecture, Arthurs researched and concentrated on delay and non-delay tolerant routing protocols as well as the categorization of different routing models [13]. To identify the most effective routing method based on various performance metrics, including packet data rate, a comparison and evaluation of several position-based routing systems were conducted. The researchers analyzed the performance of two key topology-based routing protocols utilizing the V2V and V2I communication models as mentioned in [14]. These techniques include the proactive Fisheye State Routing (FSR) protocol and the reactive routing protocol. They used different scenarios where a number of input parameters are randomly changed, such as the quantity of nodes, vehicle speed, and size of message packet. Based on throughput and End-to-End (E2E) delay metrics, the results showed that reactive protocols perform better than FSR in V2V and V2I communication. In [15], a qualified comparison of AODV and Dynamic Source Routing (DSR), as two reactive routing systems, was described. based on a number of performance indicators, including E2E delay, overhead transmission, Packet Loss Ratio (PLR), throughput, , and Packet Delivery Ratio (PDR). A VANET system was used to study these reactive protocols. varied approaches produced varied outcomes in different situations. The results of the final analysis showed that DSR performs better than AODV. Since DSR is more reliable and stable than AODV, it was determined to be the most favored protocol. A comparison of the most crucial routing protocols used in VANETs was examined by [16]. The advantages and disadvantages of routing protocols were clearly discussed in this work. It provided a comprehensive examination of various active routing approaches and techniques. These

assessments were given in relation to a number of criteria and merits. The speed, conveyance, PDR, and viability of each of the current routing protocols were compared in this article as well.

IoT applications and importance

Applications based on computer networks are increasingly prevalent in workplaces, households, factories, and manufacturing settings. In the realm of technology, they offer a variety of benefits and applications. These areas of the computer and electronic world implement a variety of standards, protocols, applications, and technologies. One of these systems is the IoT, which is a brand-new, innovative, and useful technology in our modern society [17]. All of the aforementioned advantages are also realized by these technologies, which provide secrecy, safety, and security. IoT, according to some experts, is the way of the future for smart home systems. Smart agriculture, building management, healthcare, energy, and transportation are just a few of the industries that now use IoT technology in practical ways. at order to optimize IoT systems and address their issues, several programs are created, organized, and put into place at university labs and computer and electronics departments of various businesses. These are a few (but not all) of the benefits of IoT. The following is a brief explanation of some significant advantages of IoT systems [18]:

Data Management: Through IoT systems, consumers can access a vast amount of data and information about their mechanical and electronic gadgets in the office or at home. These resources can be used to administer, oversee, and regulate a variety of processes and activities that improve people's quality of life. Even though IoT devices interchange a variety of data types, sizes, and formats over the network, all of this data can be helpful to people and function management when used with the right controlling systems.

Time Control: One of the most important and nonrenewable resources in our world today is time. By regulating and controlling mechanic and electronic device operations, IoT systems can help users save time. Instead of going back to the building to check their electronic or mechanical gadgets, which takes more time, especially in a big city with plenty of traffic cars, users can check their smart devices, such as smartphones and tablets, in just one minute.

Cost Optimization: The most compelling argument for persuading the IoT market is financial. This technology has the potential to replace those responsible for monitoring and managing supplies. IoT solutions can help people and businesses cut and optimize expenses like electricity, staff, additional transportation, and other resources.

Better Decision Making: A system can make more intelligent choices if it is able to examine broader patterns in empirical data. This eliminates the need for assumptions and provides you with data-supported insight into every facet of your company. Furthermore, the ability to see how the system behaves can provide fresh perspectives and ideas that can help you run your company like never before. Because IoT systems can access various types of data and information from IoT nodes that are sent to the core of IoT systems quickly to process and provide results for users, they are able to manage and optimize a lot of functions and make better and more intelligent decisions.

IoT has applications in almost every industry since it can network embedded hardware with constrained processor, memory, and power resources. IoT systems may be in charge of

gathering, sharing, keeping an eye on, and managing data in a variety of contexts, including factories, buildings, and natural ecosystems. In addition to sensing things, IoT systems may also be responsible for carrying out a number of tasks that are necessary for user convenience and security. Applications involving heat, electricity, and energy management provide several instances of detecting and actuating. IoT applications, however, are not limited to these domains. There might be further specific IoT use cases that are more beneficial and important in people's everyday lives. This section gives a summary of some of the most significant and useful application areas. IoT systems can be generally categorized into a wide range of industries, including manufacturing, smart homes, energy management, healthcare and medicine, and transportation, depending on a number of features and aspects. Here, we provide a quick explanation of some significant IoT applications [1–7].

Security & Surveillance: Solutions for the evolving security and surveillance sector go far beyond simple alarm monitoring. Organizations seeking to protect their personnel, property, and structures are now more visible thanks to live video monitoring and other remote security capabilities. By allowing private and public organizations to securely and remotely monitor facilities and public spaces in real time with advanced security and surveillance systems, IoT is playing a significant role in fostering safer cities, homes, and businesses. IoT-based environmental monitoring systems often use sensors to help safeguard the environment by keeping an eye on measurable natural components like soil, air, water, and atmospheric conditions.

Infrastructure Management: One important use of the IoT system is the monitoring and administration of urban and suburban development systems, like railroad tracks, and waste management systems. IoT technologies can be used to change structural circumstances or monitor any events that could raise dangers and concerns. By systemizing chores between different service providers and users, it can also be utilized to effectively schedule maintenance and repair work. The use of RFID tags, sensors, actuators, and IoT systems such as Wireless Sensor Networks (WSN) for infrastructure monitoring, control, and operation is expected to enhance emergency response coordination and incident management, as well as improve quality of service, uptime, and operational costs across all platform-related domains.

Manufacturing: The IoT is utilized in smart manufacturing and industrial applications by managing and controlling networks of manufacturing equipment, overseeing assets and situations, or controlling manufacturing processes. Rapid product development, dynamic product demand response, and real-time supply chain or industrial production network optimization are all made possible by IoT smart systems. The IoT encompasses digital control systems that automate process controls, operator tools, and service information systems to enhance security and performance. Computerizing robotic equipment to do production operations with little human intervention is known as industrial automation. Based on four components; transportation, processing, sensing, and communication, it enables a collection of machines to generate goods more rapidly and precisely.

Energy Management: Generally speaking, combining Internet-connected sensor and actuation systems should maximize energy use. To effectively balance power generation and energy consumption, it is expected that IoT systems and devices will be incorporated into various

energy-consuming equipment, such as switches, power outlets, lightbulbs, and televisions. These devices will also have the capability to communicate with the utility supply company. Along with enabling sophisticated features like scheduling, these devices would also provide customers the option to centrally manage their gadgets using a cloud-based interface or control them remotely. For instance, users have the ability to control ovens remotely, manage heating systems, and modify lighting settings.

Medical and Healthcare: Remote health monitoring, emergency alerting system control, and management are made possible by IoT technologies. These health monitoring gadgets might be anything from sophisticated equipment that can track specific implants to monitors for blood pressure and heart rate. "Smart beds" that can detect when they are occupied and when a patient attempts to get out have been implemented in some hospitals. These beds can adjust themselves to ensure that patients receive the appropriate pressure and support without needing manual assistance from nurses and doctors. Furthermore, specialized sensors can be installed in living spaces to monitor the health and overall well-being of elderly individuals, aiding them in regaining lost mobility through therapy.

Building and Home Automation: The mechanical, electrical, and electronic systems present in various types of buildings; whether public or private, commercial, institutional, or residential, can be monitored and controlled by IoT devices using home automation and building automation systems. Since many businesses in the civil and architectural sectors view smart houses and Building Management Systems (BMS) as their future, this application is among the most significant, well-liked, and frequent examples in the IoT world. Although smart homes have been a dream for many years, widespread adoption and large-scale deployments have frequently been hampered by the absence of engaging user experiences and useful technologies. Homeowners can also utilize IoT technology to monitor and manage their water and energy consumption in their homes. The home must become smarter in order to benefit from energy saving and reduction measures as utility systems become "smarter."

Transportation: The Internet of Things (IoT) can enhance the integration of information processing, control, and communication across various transportation systems. Each component of these systems is supported by a range of IoT solutions. The dynamic interaction among different elements enables inter- and intra-vehicular communication, smart parking, intelligent traffic management, electronic toll collection, fleet and logistics coordination, vehicle control, safety measures, and road assistance.

Various IoT applications present prospects for computer, mechanical, and electrical businesses and markets, but they also come with certain requirements, particularly in network domains. Because the majority of IoT system applications need different types of networks in order to provide customer service and enhance system performance. In various IoT applications, wireless networks like Wi-Fi and mobile networks play a vital role alongside other networks. Below is a general overview of the typical requirements for IoT communication use cases:

- Broad specifications for downlink and uplink data speeds.
- The quantity of reports or readings required from the IoT device for the corresponding application; the acceptable limits for latency responses necessary for the application; and the operational speed of the IoT device in which the application is employed.

- The type and extent of security required to safeguard the data and communications with the IoT device.
- Battery specifications for IoT devices that are required for a specific application.

A few IoT applications, such as transportation, smart home and building, healthcare, and security, are chosen for analysis based on their primary network requirements (Table 1) [3].

Table 1: IoT application requirements

Application	Data Rate	Relative speed	Latency	Duty Cycle	Range	Battery	Security
Transportation	Up to 100s of kbps UL	High-speed 10-150km/h	Low (Seconds)	1 report/hour~ 1 report/day	few km	3 months	High
Building and home automation	50-500kbps UL	No: Fixed Position	High (Hours)	1 report/hour~ 1 report/day	few km	10 years	High
Security & Surveillance	05.-8Mbps UL	No: Fixed Position	Zero (Milliseconds)	Real-time UL stream	few km	2months	High
Medical and healthcare	50-500kbps UL	Pedestrian: <5km/h	Low (Seconds)	1 report/hour~ 1 report/day +ad-hoc emergency	< 10s of meters	2 Years	High

Intelligent transportation system

The most crucial aspect of a smart city is infrastructure management. One of the main IoT use cases is in smart cities. According to the aforementioned McKinsey analysis, cities are the second or third most significant target area for IoT, with an estimated $1 trillion to $1.6 trillion in economic impact by 2025. There are several pilots: there are around 300 smart city initiatives in China and India alone. With 83 distinct projects falling under One of the twelve focus areas includes: environment, information and communications technology, mobility, water, energy, materials (waste), nature, built environment, public space, open government, information flows, and services. The project has resulted in reduced costs and increased revenues, positively affecting Barcelona's budget by $100 million and reportedly creating nearly 50,000 new jobs. As part of a strategic initiative led by the Prime Minister's Office, Singapore is making significant investments to establish itself as a "Smart Nation." This includes providing open access to a diverse array of government data from various sources, with much of it available in real-time [19]. Urban traffic network congestion has a negative impact on people's quality of life and poses a serious threat to the environment and any economy. In order to alleviate traffic congestion, extreme measures must be taken to cut down on travel times, line lengths, and dangerous gas emissions, particularly when people are waiting at traffic signals. The fundamental cause of the numerous drawbacks of traditional traffic management systems is the inadequate communication between cars and between vehicles and other vehicles. A variety of Vehicle Automation and Communication Systems (VACS) have been created in recent years to improve traffic flow and safety [20]. For this reason, there have also been significant advancements in the field of intelligent autonomous vehicles [21]. Due to the aforementioned

advancements, communication technology is now a critical component of traffic management systems, or ITS in general. Safety can be greatly increased and traffic and gas emissions can be minimized with the use of efficient communication. Since the 1980s, advanced control systems have been in place to address traffic congestion [17]. Our transportation systems have been revolutionized by recent developments in electronics, communications, controls, computers, and sensors. We now have a new field of study called ITS as a result of the extensive research that has been done in this area [18]. The technology mentioned earlier is integrated into Intelligent Transportation Systems (ITS) to improve the overall quality of road transportation, which includes increasing average speed, enhancing safety, and reducing traffic congestion, idle time, and harmful emissions. To implement such systems effectively, it is essential to have a comprehensive understanding of traffic dynamics at both local and global levels, as well as the effects of related phenomena and disturbances, such as the creation and propagation of shock waves and the initiation of congestion.

ITS Communication Technologies: A significant factor in the efficiency of ITS is communication technology. In ITS applications, a diverse range of communication technologies are currently being utilized. Examples of these technologies include RFID, Bluetooth, Wi-Fi, GPS/GLONASS/Galileo positioning, Wimax, wired connections, ZigBee, Dedicated Short-Range Communications (DSRC)/Wireless Access for Vehicular Environments (WAVE), 2G/3G/4G cellular networks, and Instrumentation Scientific Medical Radio Frequencies (ISM-RF). Numerous media control and routing algorithms have been created specifically for these systems. V2I communication is used by RFID, which has been helpful in parking and vehicle identification for automated toll collection [22]. RFID technology has numerous advantages for ITS, despite its restricted communication range. Attaching RFID tags to traffic signs is one use case for RFID that hasn't been extensively used yet. Vehicles traveling in the vicinity can detect these. Bluetooth is predominantly used to collect the MAC addresses of mobile devices as they move past the cars, which aids in the creation of origin-destination matrices [23]. Applications of ITS can use this data to forecast future traffic patterns. Bluetooth technology is employed to monitor the traffic density of specific network segments in real-time. Bluetooth has a limited communication range, just like RFID. The most popular communication technology in ITS research projects is Wi-Fi, particularly the 802.11a version that operates at 5.8 GHz. The availability of the ad hoc mode and its communication range of approximately 50 meters are the reasons for this. The equipment used in actual experiments is inexpensive and easily accessible. 5.8GHz is license-free and sufficiently close to the 5.9GHz suggested standard for VANETs. It can be utilized for location-based services, intra-vehicular information distribution, online connectivity, V2V, and V2I communication. For VANETs, the suggested standard is WAVE/802.11p DSRC [24]. It operates at 5.9GHz and is half as fast as the IEEE802.11a standard. The communications range is increased to about 1 kilometer by this minor adjustment (as well as a few others). 802.11p is the new MAC standard. The higher-level layers of the communication protocol, known as WAVE, control the security and application functionality. It is anticipated to exist at the infrastructure level as RSU and in every vehicle as an OBU. These standards are expected to be used for all V2V and V2I communication. For many years, automobiles have extensively utilized 2G, 3G, 4G, and 5G mobile communication technologies to access location-based services, such as interactive mapping and internet connectivity. These

technologies facilitate the connection between vehicles, personal devices, and infrastructure. Additionally, for several decades, cars have relied significantly on GPS, GLONASS, and Galileo systems to provide increasingly precise positioning information. Without it, location-based services like navigational data cannot be provided. For V2V and V2I to function correctly, the majority, if not all, rely on position information. WiMAX was first anticipated to be a part of VANETS because of its 1-kilometer range. Commercial adoption was extremely slow, nonetheless, and several communications businesses even stopped using it. Because of this problem, there is a lack of equipment that can use it, and most experiments have stayed at the simulation level. ZigBee is frequently employed in Wireless Sensor Networks (WSN) and is also being explored in Vehicle Ad-hoc Network (VANET) research due to its affordability and the availability of simulation models in various simulators. However, its communication at the 2.4GHz frequency, which is densely packed with devices, and its limited communication range of 20 meters can be easily disrupted. In contrast to RFID and Bluetooth, ZigBee supports duplex communication, allowing for the simulation of complex protocols and scenarios. The ISM RF operates as a license-free basic duplex wireless communication system, with a range extending from 20 meters to 2 kilometers, depending on the antenna and the available radio frequency power [26]. Since there is no set protocol, researchers are free to simulate and test whatever wireless protocol they like. The lack of a standard methodology makes it challenging to combine parts from various manufacturers. Nonetheless, researchers are quite skilled at managing them because they were created for self-management. They utilize a wide range of frequencies, including 2.4GHz, 868MHz, and 915MHz.

VANET routing schemes and protocols

The dynamic, self-organizing, and decentralized architecture of VANETs is one of the distinctive features shared by both conventional ad-hoc networks and traffic networks. Furthermore, the VANETs have special structural characteristics, can manage high mobility scenarios, and perform tailored traffic applications. Routing and message/packet forwarding have always been major issues for VANETs. In a highly dynamic vehicle network, finding, building, and choosing an appropriate route can be challenging, particularly when bandwidth, computational overhead, packet delay, and stability are taken into account [9] [5].

Significant efforts in research and development have been dedicated to the formulation of routing protocols and forwarding mechanisms for Vehicular Ad Hoc Networks (VANETs). The existing routing process for VANETs can be summarized by examining the essential information required for routing protocols, which encompasses topology-based protocols, position-based protocols, map-based protocols, and road-based protocols [7]. Topology-based protocols facilitate routing and forwarding by utilizing the configuration of road segments, regardless of whether a global routing table, such as DSDV, or a local one, like DSR and AODV, is employed. In contrast, position-based protocols, including Greedy Perimeter Coordinator Routing (GPCR) and Intersection-based Geographical Routing Protocol (IGRP), primarily concentrate on the locations of vehicles. Map-based protocols are designed to consider varying conditions of road segments. While Geographic Source Routing (GSR) takes junctions into account, Shortest Path-based Traffic Light Aware Routing (STAR) emphasizes the influence of traffic signals. Additionally, road-based protocols that facilitate communication among different road segments include Vehicle Assisted Data Delivery (VADD).

The majority of the aforementioned research primarily concentrates on the processes involved in distributed routing discovery and selection [9] [11]. The challenges associated with coordination among the nodes significantly hinder the efficiency of data transmission. In the context of vehicular ad hoc networks (VANETs), which are characterized by a highly dynamic topology, broad coverage, and limited robustness, the routing protocol plays a vital role. An effective routing system is essential for maintaining high transmission quality within VANETs [26]. One of the most efficient and practical classes of VANET routing protocols is topology-based, which is based on multiple implementation paradigms. They fall into three primary groups [15]. The next sections include descriptions of the classification, which is shown in Fig. 1.
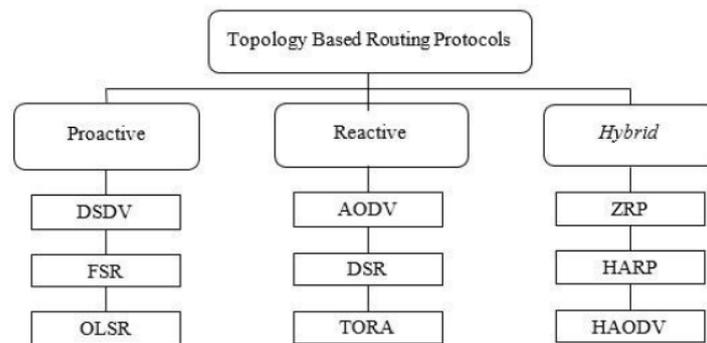
Figure 1: Classifications of topology-based routing protocols

Topology-based protocols can be categorized into three primary subclasses: proactive protocols, such as OLSR, DSDV, or table-driven protocols; reactive or on-demand protocols, exemplified by AODV; and hybrid protocols, represented by the Zone Routing Protocol (ZRP).

Proactive routing protocols: In these proactive routing protocol-using networks, nodes maintain many tables that display the overall network topology. These protocols are based on algorithms based on the shortest path. Whenever there is a change in the network topology, it is necessary for each node to refresh its routing table. The OLSR and DSDV protocols are based on this assumption [8]. For these protocols, we developed the following description:

DSDV: This is one of the best routing strategies for VANETs. The standard method used by the DSDV algorithm vector is the Distributed Bellman-Ford (DBF) algorithm [27]. This protocol, which uses a distance vector technique, uses successive numbers that originate from and are regenerated by the point to solve looping issues brought on by out-of-date routing information [15].

OLSR: This proactive routing system leverages link-state algorithm optimization and receiving traffic management messages through multipoint relay technology [28] [29]. Each node, as per the OLSR paradigm, is capable of retaining information regarding both its neighboring routers and its own node. As a result, there is significantly less time delay when a route is necessary, and a route is predicted when information needs to be provided [7].

Reactive routing protocols: Reactive routing protocols, which are classified as on-demand methods, modify the routing table in response to failures of nodes and links or changes in the

network structure. These protocols are designed to lessen the overhead that proactive routing techniques produce. For these categories of routing protocols, a singular route is established as needed [11]. The following are two popular and practical reactive routing protocols:

AODV: This protocol is used for dynamism, auto-starting, and multi-hop routing between mobile nodes. AODV is implemented in line with the applicable basis when it is necessary for the network [6]. Route maintenance and discovery activities are also performed as required in the AODV, even when self-communication requires only two nodes [9] [30].

DSR: This routing protocol is founded on the essential principles of routing methodology. In this model, a route cache is monitored by the dynamic source routing nodes. When a new node is found, the route cache is updated [15]. Nodes may find and maintain source routes to any destination in the network because to two essential DSR components: route maintenance and route discovery.

Hybrid routing protocol: These protocols employ both proactive and reactive routing strategies [8]. In order to determine the nodes' paths within the zone and their neighbors, proactive routing protocols complete reactive routing procedures, which is how it works well with zone routing protocols [15]. The most recognized instance of a hybrid routing protocol is ZRP. The prevalent challenges in an ad hoc mobile network primarily arise between adjacent nodes [30]. The zone that the ZRP shows in its k-neighborhood contains almost all of the nodes. In ZRP, a node's routing zone owns any node that is k hops away from another node. The source's proactively cached routing table can be used to initiate a route to a place within the local environment. The packet will be sent immediately since the source and the destination are in the same place at the moment [8].

Main challenges of VANET

Researchers and developers face many obstacles when developing VANET applications, protocols, and simulation tools. A few researchers have examined the networking and communication aspects of VANET, as well as the security and privacy issues. Others focus on the VANET routing protocols' requirements and how they could reduce network bandwidth usage while enhancing communication time. 1. The quality of vehicular communications is primarily influenced by vehicle density and speed. In networks with high vehicle density, it becomes difficult for vehicles to effectively share the available network resources, leading to potential congestion issues [31]. Additionally, the fast speeds of the vehicles make it harder to maintain the connection between the nodes. The differences in vehicle density and speed by environment (urban, suburban, and highway) are shown in Table 2.

Table 2: The velocity and the concentration of vehicles within various environments.

Scenario	Vehicle density (vehicles/km^2)	Velocity (km^2)	Communication Range (m)
Urban	1000-3000	0-100	50-100
Suburban	500-1000	0-200	100-200
Highway	100-500	0-500	200-1000

Even when they are not stable, many IoT devices in the majority of ITS applications want to send a variety of data to a central control unit or to one another. It means they are moving faster than normal. Because of this, ITS faces two main challenges: first, the enormous amount of data that automobiles send to the network, and second, the frequent modifications to the network design brought on by the fast speed of automobiles. Router tables must also be revised because of the significant mobility of ITS nodes and the regular alterations in network topologies [32]. These upgrades need to happen fast because ITS applications are delayed. When delays surpass a specific threshold, many ITS domains, like smart parking, have mechanisms that reject packets and expire data [2] [3]. IoT systems need to improve their routing mechanism in order to avoid crashing their routers and networks. The most important VANET challenges can be classified into six primary categories: scalability, quality of service (QoS), security and privacy, limitations in transmission capacity, stability, efficient routing, and social and economic issues [15]. The following is a detailed explanation of each of these issues [5]:

Scalability: Scalability presents a significant challenge in VANET topologies, as routing protocols must maintain full coverage regardless of the network's size. Assessing the quantity of vehicles within a designated area is essential, given that nodes in a Vehicle Ad Hoc Network (VANET) may frequently connect and disconnect. Additionally, the number of cars changes with the time of day. Fewer vehicles are present on the roads at night compared to daytime hours [33]. Therefore, if there is only one car within a 10-kilometer radius, it is very difficult for the node to link with the others. The limited bandwidth poses a significant challenge when numerous nodes within an extensive network seek to communicate simultaneously. The growing demand can also result in higher interference and network congestion in VANET systems. [6] [11].

Quality of Services (QoS): It can be challenging to obtain the appropriate QoS in VANET applications. The average PDR, throughput, and E2E delay make up this statistic [31]. A functional routing system requires a low average E2E delay, a limited packet drops rate or maximum PDR, and a high throughput. An ideal routing protocol should demonstrate no normalized control or routing overhead, achieve 100% throughput, have a 0% packet drop rate or a 100% packet data rate, and maintain an average end-to-end latency of 0 milliseconds. [7] [11].

Security and privacy: One of the primary and most pressing issues associated with all wireless networks, including VANETs, pertains to privacy and security. Each node within a VANET possesses the capability to enter and exit the network at will, while also engaging in communication with numerous other ITS nodes. Furthermore, this type of wireless network comprises nodes that may remain concealed. [34]. Thus, malicious nodes that are part of the network can produce wormholes, rushing attacks, black holes, and eavesdropping [5]. The routing protocols should be designed to protect the network from such security vulnerabilities.

Transmission bandwidth limitation: The VANET system is allocated a bandwidth of 75 MHz within the 5.9 GHz frequency spectrum, in accordance with the IEEE WAVE standard [35]. The system provides a peak data rate of 27 MHz and functions efficiently within a limited range of 1000 meters, employing an orthogonal modulation technique. Nevertheless, an increase in the

number of mobile nodes leads to a higher degree of congestion [36]. In periods of elevated demand, effective communication may become significantly difficult due to the constraints of limited bandwidth. An optimally designed routing protocol ought to reduce normalized control or routing overhead in order to improve bandwidth efficiency and mitigate this challenge [11].

Stability: Stability presents a significant challenge within mobile networks. When all vehicles are traveling in the same direction, it becomes simpler to uphold a consistent network topology [37] [33]. Should the vehicles continue to travel in the same direction, the stability of the network topology will diminish over time as both time and speed escalate. The impact of vehicle speed on the stability of the network topology is less significant compared to the influence of driving direction. Conversely, as the density of vehicles rises, the topology exhibits greater stability, akin to the conditions observed in more densely populated urban areas [5] [7].

Efficient routing: A routing technique that increases system capacity, minimizes latency, and requires less computational complexity is considered effective in a VANET. The development of a feasible and useful algorithm that satisfies all of the above specified criteria and can function in a range of network topologies is a major field of technological research. The ever-changing characteristics of mobile nodes in Vehicular Ad Hoc Networks (VANETs) pose significant challenges in identifying and maintaining routes [38] [39]. Most VANET routing methods demand a unique address for each participating vehicle. This suggests that a technique that can be utilized to set unique addresses for nodes is desirable, even though these models do not prohibit duplicate allocation addresses in VANET systems [6].

Social and economic challenges: In addition to the technical challenges of implementing the VANET, social and economic concerns may need to be taken into account. Because customers may object to this type of surveillance, it can be difficult to persuade technology companies to install a traffic light violation notification system [11] [28]. To enhance the reliability and security of the information system, it is essential to eliminate harmful data. To guarantee smooth network communication, routing or traffic management using the optimal model is required [5][6].

CONCLUSION

One of the primary IoT applications in the modern era is smart cities. ITS is essential to the development of smart cities. ITS makes use of a variety of technologies and networks. The most widely used network technology for ITS communication in smart cities is VANET. Numerous issues, including security, stability, and routing effectiveness, confront VANET. In order to enhance VANET network performance, this article concentrated on routing techniques and protocols. First, we outlined the key components and attributes of IoT systems. They described some of the most important and well-liked IoT applications, such as smart health care and smart buildings. The primary categorization of VANET routing techniques and protocols was then explained. The main issues with VANET routing protocols are assessed and described.

REFERENCES

[1] Q. Fu, E. Sun, K. Meng, M. Li and Y. Zhang, "Deep Q-Learning for Routing Schemes in SDN-Based Data Center Networks," in IEEE Access, vol. 8, pp. 103491-103499, 2020.

[2] Y. Huang, M. Lee, T. Fan-Chiang, X. Huang and C. Hsu, "Minimizing Flow Initialization Latency in Software Defined Networks," 17th Asia-Pacific Network Operations and Management Symposium (APNOMS), pp. 303-308, 2015.

[3] M. U. Rehman, M. A. Shah, M. Khan and S. Ahmad, "A VANET based Smart Car Parking System to Minimize Searching Time, Fuel Consumption and CO2 Emission," 24th International Conference on Automation and Computing (ICAC), pp. 1-6, 2018.

[4] B. Mukhopadhyay and T. Samanta, "A Smart Parking-lot Occupancy Model in 5G V2V and V2I Wireless Communication," IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 1552-1557, 2021.

[5] A. Ahamed and H. Vakilzadian, "Issues and Challenges in VANET Routing Protocols," IEEE International Conference on Electro/Information Technology (ICEIT), pp. 0723-0728, 2018.

[6] K. Mehta1, L. G. Malik, and P. Bajaj, "VANET: Challenges, Issues and Solutions," Sixth International Conference on Emerging Trends in Engineering and Technology (SICETET) pp. 4799-4785, 2013.

[7] P. Sharma and S. Jain, "Review of VANET Challenges and Protocol for Architecture Design and Intelligent Traffic System," 2nd International Conference on Data, Engineering and Applications (IDEA), pp. 1-4, 2020.

[8] M. Rajhi, H. Madkhali and I. Daghriri, "Comparison and Analysis Performance in Topology-Based Routing Protocols in Vehicular Ad-hoc Network (VANET)," IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), pp. 1139-1146, 2021.

[9] K. Sundari and A. Senthil Thilak, "Impact of Realistic Mobility Models on the Performance of VANET Routing Protocols," International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (ICSCEPT), pp. 1-6, 2023.

[10] M. M. Hamdi, S. A. Jassim and B. S. Abdulhakeem, "Successful Delivery Using Stable Multi-Hop Clustering Protocol for Energy Efficient Highway VANETs," 7th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), pp. 1-6, 2023.

[11] E. Khoza, C. Tu and P. A. Owolawi, "Comparative Study on Routing Protocols for Vehicular Ad-Hoc Networks (VANETs)," International Conference on Advances in Big Data, Computing and Data Communication Systems (ICABCD), pp. 1-6, 2018.

[12] A. Srivastava, N. Bagga and M. Rakhra, "Analysis of Cluster-Based and Position-based Routing Protocol in VANET," 9th International Conference on Reliability, Information Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1-5, 2021.

[13] D. S. Kumar, R. Thenmozhi and K. D. S, "Delay and Non Delay Tolerant Hybrid Routing Protocols in VANET," 1st International Conference on Computational Science and Technology (ICCST), pp. 666-670, 2022.

[14] R. Z. Akbar, Istikmal and Sussi, "Performance Analysis FSR and DSR Routing Protocol in VANET with V2V and V2I Models," 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), pp. 158-163, 2020.

[15] K. P. Sampoornam, S. Saranya, S. Vigneshwaran, P. Sofiarani, S. Sarmitha and N. Sarumathi, "A Comparative Study on Reactive Routing Protocols in VANET," 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 726-731, 2020.

[16]   S. Yadav, N. K. Rajput, A. K. Sagar and D. Maheshwari, "Secure and Reliable Routing Protocols for VANETs,"   4th International Conference on Computing Communication and Automation (ICCCA), pp. 1-5, 2018.

[17]   T. Anagnostopoulos, A. Zaslavsky, K. Kolomvatsos, A. Medvedev, P. Amirian, J. Morley, and S. Hadjieftymiades, "Challenges and Opportunities of Waste Management in IoT-Enabled Smart Cities: A Survey", IEEE Transactions On Sustainable Computing, vol. 2, no. 3, pp. 275-289, 2017.

[18]   W. Cheng, X. Cheng, M. Song, B. Chen, and W. W. Zhao, "On the Design and Deployment of RFID Assisted Navigation Systems for VANETs," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 7, pp. 1267– 1274, 2012.

[19]   H. Mamdouhi, S. Khatun, and J. Zarrin, "Bluetooth Wireless Monitoring, Managing and Control for Inter Vehicle in Vehicular Ad-Hoc Networks," Journal of Computer Science, vol. 5, no. 12, pp. 922–929, 2009.

[20]   C. Diakaki, M. Papageorgiou, V. Dinopoulou, I. Papamichail, and G. Malandraki, "State-of-the-Art and Practice Review of Public Transport Priority Strategies," IET Intel. Transport Syst., vol. 9, pp. 391–406, 2015.

[21]   V. Butakov and P. Ioannou, "Driving Autopilot with Personalization Feature for Improved Safety and Comfort," in IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, vol. 5, pp. 387–393, 2015.

[22]   W. Cheng, X. Cheng, M. Song, B. Chen, and W. W. Zhao, "On the Design and Deployment of RFID Assisted Navigation Systems for VANETs," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 7, pp. 1267– 1274, 2012.

[23]   H. Mamdouhi, S. Khatun, and J. Zarrin, "Bluetooth Wireless Monitoring, Managing and Control for Inter Vehicle in Vehicular Ad-Hoc Networks," Journal of Computer Science, vol. 5, no. 12, pp. 922–929, 2009.

[24]   X. Ma And X. Chen, "Performance Analysis Of IEEE 802.11 Broadcast Scheme In Ad Hoc Wireless Lans," IEEE Trans. Veh. Technol., vol. 57, no. 6, pp. 3757–3768, 2008.

[25]   C. Chen, Q. Pei, and X. Li, "A GTS Allocation Scheme to Improve Multiple-Access Performance in Vehicular Sensor Networks," IEEE Trans. Veh. Technol., vol. 65, no. 3, pp. 1549– 1563, 2016.

[26]   Y. Wang, J. Hu, Y. Zhang, and C. Xu, "Reliability Evaluation of IEEE 802.11p-Based Vehicle-to-Vehicle Communication in an Urban Expressway," Tsinghua Sci. Technol., vol. 20, no. 4, pp. 417–428, 2015.

[27]   H. Gao, C. Liu, Y. Li and X. Yang, "V2VR: Reliable Hybrid-Network-Oriented V2V Data Transmission and Routing Considering RSUs and Connectivity Probability," in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 6, pp. 3533-3546, 2021.

[28]   G. Sun, L. Song, H. Yu, V. Chang, X. Du and M. Guizani, "V2V Routing in a VANET Based on the Autoregressive Integrated Moving Average Model," in IEEE Transactions on Vehicular Technology, vol. 68, no. 1, pp. 908-922, 2019.

[29]   L. L. Cárdenas, A. M. Mezher, P. A. Barbecho Bautista, J. P. Astudillo León and M. A. Igartua, "A Multimetric Predictive ANN-Based Routing Protocol for Vehicular Ad Hoc Networks," in IEEE Access, vol. 9, pp. 86037-86053, 2021.

[30]   C. Liu, G. Zhang, W. Guo and R. He, "Kalman Prediction-Based Neighbor Discovery and its Effect on Routing Protocol in Vehicular Ad Hoc Networks," in IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 1, pp. 159-169,. 2020.

[31]   C. Chen, L. Liu, T. Qiu, D. O. Wu and Z. Ren, "Delay-Aware Grid-Based Geographic Routing in Urban VANETs: A Backbone Approach," in IEEE/ACM Transactions on Networking, vol. 27, no. 6, pp. 2324-2337, 2019.

[32]   L. Liu, C. Chen, B. Wang, Y. Zhou and Q. Pei, "An Efficient and Reliable QoF Routing for Urban VANETs With Backbone Nodes," in IEEE Access, vol. 7, pp. 38273-38286, 2019.

[33]   J. Shen, C. Wang, A. Castiglione, D. Liu and C. Esposito, "Trustworthiness Evaluation-Based Routing Protocol for Incompletely Predictable Vehicular Ad Hoc Networks," in IEEE Transactions on Big Data, vol. 8, no. 1, pp. 48-59, 2022.

[34]   R. Han, Q. Guan, F. R. Yu, J. Shi and F. Ji, "Congestion and Position Aware Dynamic Routing for the Internet of Vehicles," in IEEE Transactions on Vehicular Technology, vol. 69, no. 12, pp. 16082-16094, 2020.

[35]   B. Amina, and E. Mohamed, "Performance Evaluation of VANETs Routing Protocols Using SUMO and NS3," IEEE 5th International Congress on Information Science and Technology (ICIST), pp. 525-530, 2018.

[36]   H. Liu, T. Qiu, X. Zhou, C. Chen and N. Chen, "Parking-Area-Assisted Spider-Web Routing Protocol for Emergency Data in Urban VANET," in IEEE Transactions on Vehicular Technology, vol. 69, no. 1, pp. 971-982,. 2020.

[37]   S. A. Rashid, M. Alhartomi, L. Audah and M. M. Hamdi, "Reliability-Aware Multi-Objective Optimization-Based Routing Protocol for VANETs Using Enhanced Gaussian Mutation Harmony Searching," in IEEE Access, vol. 10, pp. 26613-26627, 2022.

[38]   F. H. Kumbhar and S. Y. Shin, "VAR: Novel Vehicular Ad-Hoc Reliable Routing Approach for Compatible and Trustworthy Paradigm," in IEEE Communications Letters, vol. 25, no. 2, pp. 670-674, 2021.

[39]   R. Han, Q. Guan, F. R. Yu, J. Shi and F. Ji, "Congestion and Position Aware Dynamic Routing for the Internet of Vehicles," in IEEE Transactions on Vehicular Technology, vol. 69, no. 12, pp. 16082-16094, 2020.

RESUME

Mojtaba Ayoubi MOBARHAN

Mojtaba Ayoubi Mobarhan received a BSc degree in electrical engineering (electronic) from the Islamic Azad University of Lahijan (Iran) in 2008. In October 2009, he joined the Department of Computer Engineering, University of Guilan, Iran, as a full-time MSc. Student. He received the MSc degree in 2011. He is a Ph.D. student under the guidance of Dr. Muhammed Salamah in the EMU University Computer Engineering Department. His research interests include wireless Networks, the Internet of Things, the Security of Networks, cryptography algorithms, and mobile communications.

Prof. Dr Mohammed SALAMAH

M. Salamah received a B.S., M.S., and PhD. degrees in Electrical and Electronics Engineering from Middle East Technical University in 1988, 1990, and 1995, respectively. From 1988 to 1996, he worked as a Research Assistant and Lecturer at METU. In 1995, he worked as an invited researcher in the Networks Department of Ecole Nationale Superieure des Telecommunications in Paris, France. Since 1996, he has been with the Department of Computer Engineering at Eastern Mediterranean University, TRNC, Turkey, where he is a Professor. He served as the Assistant Chairman of the Computer Engineering Department of EMU from 1997 to 2004; The chair of the Student Project and Design Center at EMU from 2005 to 2008; Rectorate Coordinator for Student Affairs and Informatics from 2008 to 2010; Head of the Computer Engineering Department of EMU from 2010 to 2013; and Rector's Advisor for the Middle East at EMU from 2012-2014. His current research interests are computer networks, wireless networks with an emphasis on resource allocation and management, performance evaluation, multiprocessing, and hardware-oriented algorithms.

EXPLAINABLE DEEP LEARNING FOR SKIN CANCER DETECTION USING DENSENET121 AND GRAD-CAM

Havva Hazel ARAS[1], Nurettin DOĞAN[2]

INTRODUCTION

Skin cancer is among the earliest documented diseases in medicine, with references dating back to Ancient Egypt and Greece. The advancement of microscopy and histopathology techniques in the late 19th and early 20th centuries enabled the detailed examination and classification of various skin cancer types under the microscope [1]. The discovery of aggressive types like melanoma further emphasized the importance of early diagnosis. From the mid-20th century onwards, treatment methods such as radiotherapy and chemotherapy were developed and began to be used alongside surgical interventions [2]. However, the side effects and limited success rates of these methods necessitated the exploration of new treatment and diagnostic approaches. Since the late 1980s, the advancement of digital imaging technologies and computer-aided diagnostic systems has marked a new era in skin cancer diagnosis. In particular, the widespread adoption of dermatoscopy devices has allowed for more detailed examination of skin lesions [3]. In the early 21st century, significant advancements in artificial intelligence and machine learning brought transformative breakthroughs to the field of medical imaging. In the beginning, conventional machine learning techniques like neural networks and support vector machines (SVM) were utilized. In recent years, deep learning models have garnered significant attention in the diagnosis of skin cancer. [4]. Convolutional neural networks (CNNs) have played a particularly prominent role in this area. These models, trained on extensive datasets of images, have proven highly successful in classifying skin lesions and recognizing malignant features. However, the use of these technologies has raised some important questions. The inner mechanisms of deep learning models are frequently viewed as a "black box," making it challenging to interpret how they arrive at their decisions. This issue poses challenges to the reliability and acceptability of these models, particularly in healthcare [5].

Recent progress in artificial intelligence and deep learning are offering groundbreaking innovations in the medical field. Deep learning models, trained on extensive datasets, are capable of detecting intricate patterns and abnormalities much faster and more accurately than the human eye [6]. These technologies hold immense potential, particularly in the domain of medical imaging. The potential is especially significant in the diagnosis of skin cancer. However, comprehending the decision-making process of AI models is crucial for the acceptance and reliability of these technologies. Explainable AI (XAI) methods enable us to comprehend the inner workings of these models, thereby enhancing the trust of both healthcare

[1] Selçuk University, Konya/Türkiye, Orcid: https://orcid.org/0000-0002-4179-3188, hazelaras42@gmail.com
[2] Selçuk University, Konya/Türkiye, Orcid: https://orcid.org/0000-0002-8267-8469, nurettin.dogan@selcuk.edu.tr

professionals and patients in these systems [7]. This is especially important in the medical field, where decisions can be sensitive and have life-altering consequences. In this context, the advancement of deep learning models for diagnosing skin cancer and the effort to make these models more interpretable will offer significant contributions to the medical community. This study will not only support the widespread adoption of AI in healthcare but also facilitate patients' access to accurate and timely treatments. Such efforts will help healthcare professionals make more informed and reliable decisions in the sector [8].

In recent years, there has been a substantial increase in research focused on diagnosing skin cancer, particularly malignant melanoma. These studies focus on the development of early detection and treatment methods, converging at the intersection of biomedical and computer engineering disciplines.

Zhang et al. introduced an optimized convolutional neural network (CNN) approach aimed at improving the early diagnosis of skin cancer. In skin cancer diagnosis, convolutional neural networks have achieved high accuracy rates by utilizing image processing and computer vision techniques. In the proposed method, the efficiency of the CNN is enhanced through the use of an improved whale optimization algorithm (WOA). This algorithm mathematically models the whale hunting behavior to optimize the weights and biases of the CNN. The approach was evaluated using two distinct datasets, Dermquest and DermIS.The simulation results indicate that the proposed method surpasses other existing approaches in performance. Notably, high values were achieved in using performance metrics like accuracy, sensitivity, and specificity. The optimized CNN achieved an accuracy of 97%, with a sensitivity of 95% and a specificity of 98% [9].

Ogudo et al. present an optimized artificial intelligence model, OSSAE-BPNN (Optimized Stacked Sparse Autoencoder and Backpropagation Neural Network), for identifying and classifying skin lesions in dermoscopic images. The model incorporates multi-level threshold-based image segmentation, OSSAE-based feature extraction, and parameter tuning using the Seagull Optimization Algorithm (SGO). The OSSAE-BPNN model automatically detects and classifies skin lesions from dermoscopic images. The study conducted comprehensive experimental analyses on skin lesions from seven distinct classes: Angioma, Nevus, Lentigo NOS, Solar Lentigo, Melanoma, Seborrheic Keratosis, and Basal Cell Carcinoma. The experimental results show that the OSSAE-BPNN model surpasses other existing strategies. The model attained an average accuracy of 96.2%, sensitivity of 85.3%, specificity of 97.8%, precision of 85.4%, and an F-score of 84.9% [10].

Rukhsar et al. explored cancer classification using deep learning approaches on RNA-Seq gene expression data. The work analyzes data from five types of cancer obtained from the Mendeley data repository: LUAD, BRCA, KIRC, LUSC, and UCEC. In the initial step, RNA-Seq values were converted into two dimension images using normalization and zero-padding. In the second step, deep learning-based feature extraction and selection were performed. In the final step, classification was carried out using eight distinct deep learning algorithms. The best results were achieved with a 70% training and 30% test data split. The highest performance was obtained utilizing a Convolutional Neural Network (CNN), achieving an accuracy of 97% [11].

Orhan and Yavşan proposed an AI-assisted detection model using deep learning techniques for melanoma diagnosis. In the study, two different datasets were combined to create a comprehensive dataset consisting of 8,598 images. During model development, deep learning algorithms like ResNet, AlexNet, VGG16, MobileNet and VGG19 were employed. The performance of these models was evaluated during the training, validation, and testing phases. The MobileNet model demonstrated the highest performance with an accuracy rate of 84.94% [12].

In Table 1, the conducted studies are compared in detail.

Table 1: Comparison of Previous Studies

Study	Dataset	Originality	Method
Zhang et al. [9]	Dermquest and DermIS	The use of the Improved Whale Optimization Algorithm (WOA) and CNN	WOA + CNN
Ogudo et al. [10]	Dermoscopy Images Benchmark Dataset	This study introduces an automated skin lesion detection and classification technique, named OSSAE-BPNN, which leverages a backpropagation neural network (BPNN) in conjunction with an optimized stacked sparse autoencoder (OSSAE)-based feature extractor.	OSSAE-BPNN
Rukhsar et al. [11]	RNA-Seq gene expression data of BRCA, KIRC, LUAD, LUSC and UCEC	The 1D RNA-Seq data were transformed into 2D images. Eight different deep learning models were used for classification.	VGG16, VGG19, CNN, ResNet50, ResNet101, ResNet152, GoogleNet, AlexNet
Orhan and Yavşan [12]	Nevus classifier and Skin lesions datasets	A unified dataset approach to enhance data diversity and realism for melanoma detection.	MobileNet, AlexNet, ResNet, VGG16, VGG19

When examining the studies, it is evident that the limited diversity and size of datasets negatively impact model performance. This also reduces the model's generalization capability. Additionally, class imbalance issues make accurate diagnosis more challenging, particularly for rare classes like melanoma. Most studies have not been adequately tested in clinical settings, leaving their adaptability to real-world applications and integration with physicians uncertain. The absence of explainability in deep learning models presents a challenge for reliability within the medical field.

This study seeks to enhance interpretability in artificial intelligence models for skin cancer diagnosis by employing the DenseNet121 deep learning model alongside the Grad-CAM method. By enhancing the explainabilityof the internal processes within deep learning models, transparency and reliability in medical diagnostic processes will be improved. As a result, healthcare professionals can make more informed decisions, and patients will feel greater confidence in their treatment processes. This approach has the potential to promote the adoption

and widespread application of AI in skin cancer diagnosis, thus enhancing the effectiveness of early detection and treatment.

METHOD

This study employed a deep learning-based method for the classification of skin diseases. A classification model was trained using the DenseNet121 architecture on dermoscopic images from the HAM10000 dataset. The Grad-CAM method was utilized to enhance the interpretability of the model's decisions and to highlight the regions it focused on during classification [13]. A thorough analysis was performed during the training and evaluation phases, employing performance metrics such as accuracy, precision, sensitivity, AUC (Area Under the Curve) and F1 score. Figure 1 summarizes the method's overall workflow and the steps used in the process.
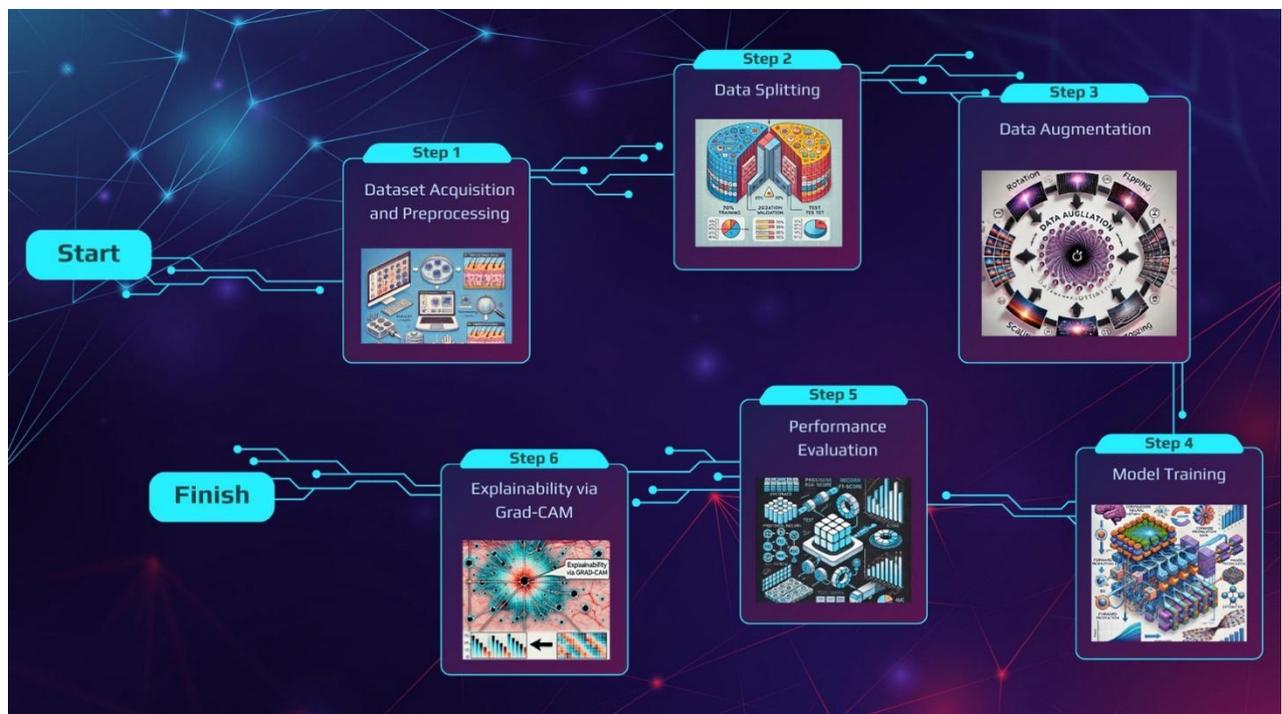
Figure 1: Deep Learning Workflow for Skin Lesion Classification

Dataset

The HAM10000 dataset is a large and diverse image collection used for the classification of skin diseases. It contains a total of 10,015 high-resolution dermoscopic images, gathered by dermatologists, and covers seven different types of skin diseases. The disease classes in the dataset are defined as melanoma, seborrheic keratosis, actinic keratosis, benign keratosis-like lesions, basal cell carcinoma, dermatofibroma, and vascular lesions [14]. An example image for each class is shown in Figure 2. This dataset serves as a comprehensive resource for accurately classifying skin lesions and diagnosing skin diseases.
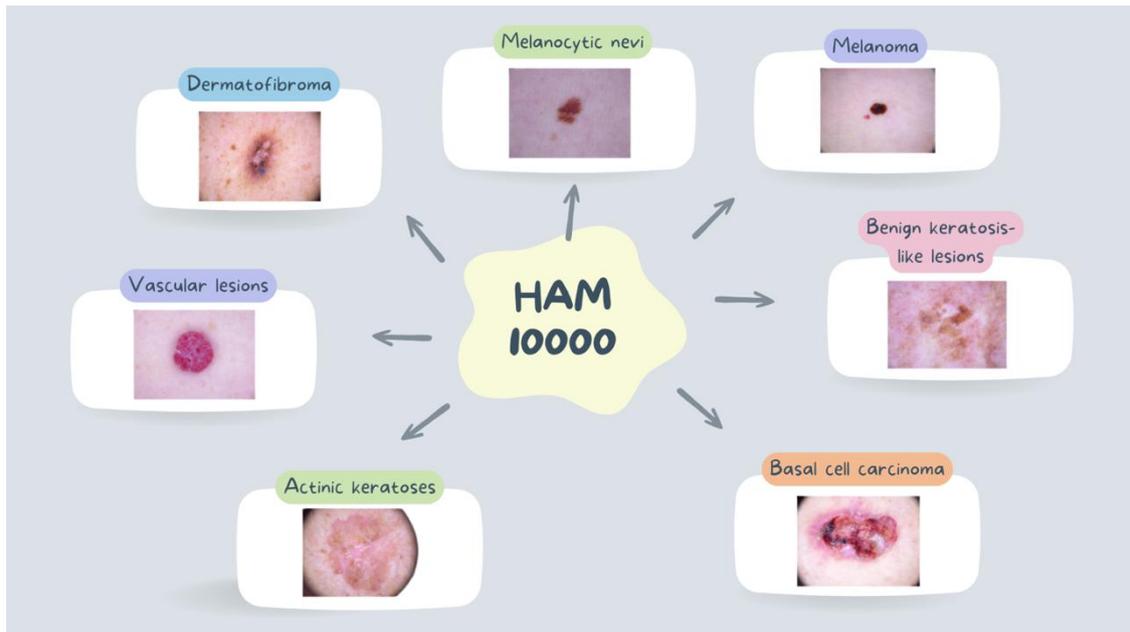
Figure 2: Skin Lesion Categories in the HAM10000 Dataset

DenseNet121

DenseNet121 is a widely used neural network architecture in deep learning models, known for its dense connectivity between layers. In this model, each layer receives the outputs of all previous layers, allowing the creation of richer feature maps. This structure accelerates the learning process while preventing an unnecessary increase in the number of parameters. With a total of 121 layers, DenseNet121 effectively addresses the vanishing gradient problem that often arises in deeper networks, enabling efficient learning. Commonly used in tasks such as image classification, this model has the capacity to achieve high accuracy even on low-dimensional data. The advantages provided by DenseNet121 make it an ideal solution for challenging tasks like the classification of skin diseases [15].

Performance Metrics

Metrics like accuracy, precision, recall, F1 score, and AUC were utilized to assess the model's performance. Accuracy assesses the overall correctness of the model's predictions, whereas precision reflects the proportion of true positive predictions among all positive predictions. Recall evaluates the model's effectiveness in correctly identifying actual positives, while the F1 score provides a harmonic mean that balances precision and recall. The AUC value reflects the overall classification performance of the model.Additionally, a confusion matrix was employed to analyze the model's correct and incorrect classifications in detail. To further analyze the model's performance, Mean Squared Error (MSE), variance, and bias were also considered [16].

Grad-Cam

Grad-CAM is a technique employed to visualize and interpret the decision-making process of deep learning models. This method highlights the regions of an image that the model focuses on while making a classification. Grad-CAM creates an activation map by leveraging the gradients of the classification layer's output with respect to the final convolutional layer. This

allows for a visual representation of the features the model relies on when making predictions. Grad-CAM is widely used, particularly in medical image analysis, to make the model's decisions more interpretable and understandable [17].

Experimental Results

To balance the dataset's uneven structure, class augmentation and data augmentation techniques were employed. Specifically, image augmentation techniques were applied to increase the number of examples in underrepresented classes. The data augmentation methods used include horizontal flipping, rotation, scaling, cropping, shearing, Gaussian blur, additive Gaussian noise, and contrast adjustments. These transformations helped to multiply the images in each class, contributing to a balanced dataset. Additionally, all images were resized to a specified dimension to meet the model's training requirements. As a result, as shown in Figure 3, the dataset was expanded and balanced, ensuring each class contained an equal number of examples.
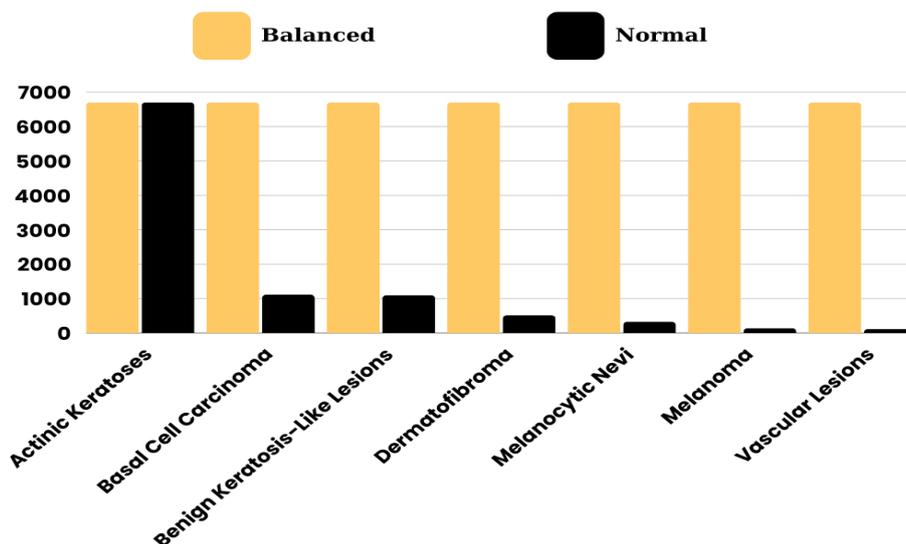
Figure 3: Distribution chart of examples for different classes in the dataset used in the study. Each bar represents the number of samples for the respective class.

The dataset was divided into three subsets to assess the model's performance and generalization capability: 70% was allocated for training, 20% for validation, and 10% for testing. The training set used the model to learn its parameters, while the validation set used to evaluate the model's generalization ability and to identify the best-performing model. The test set was set aside to independently evaluate the performance of the final model. Moreover, during the validation process, the Grad-CAM method was used to evaluate the model's ability to correctly highlight malignant skin lesions. This process was employed to observe both the classification accuracy and the interpretability of the model's decisions.

In this study, the DenseNet121 architecture was utilized as the base model for classifying skin diseases. DenseNet121 was initialized with pre-trained weights from the ImageNet dataset, and additional layers were incorporated to tailor the model for classification. All layers in the base

model were frozen to ensure that training was performed only on the newly added layers. At the top of the model, a GlobalAveragePooling2D layer was applied to summarize the feature maps. A fully connected layer with 256 units and a ReLU activation function was then added. To prevent overfitting, L2 regularization and a Dropout layer were incorporated into this layer. Finally, a Dense layer with a softmax activation function for 7 classes was added as the classification layer, making the model suitable for skin disease classification.

During the training of the model, various methods were employed to enhance performance and prevent overfitting. The training process was set to a total of 15 epochs, but early stopping was implemented to maximize the model's generalization ability. By monitoring validation loss, training was halted at the 9th epoch when no further improvement was observed. This approach prevented unnecessary training iterations and reduced the risk of overfitting. Additionally, the learning rate was dynamically adjusted using the ReduceLROnPlateau method. When no improvement in validation loss was detected, the learning rate was reduced by a factor of 0.1, allowing the training to proceed with smaller steps.Upon completing the training process, the accuracy and loss values for both the training and validation sets were analyzed to evaluate the model's performance. To illustrate these results, Figure 4 displays the progression of training and validation accuracy, while Figure 5 depicts the changes in training and validation loss. These graphs clearly depict how the model improved during the training process and highlight the point where early stopping occurred. The convergence of training and validation accuracy at a certain point, along with the decrease in loss, indicates that the model's generalization ability was successfully optimized.
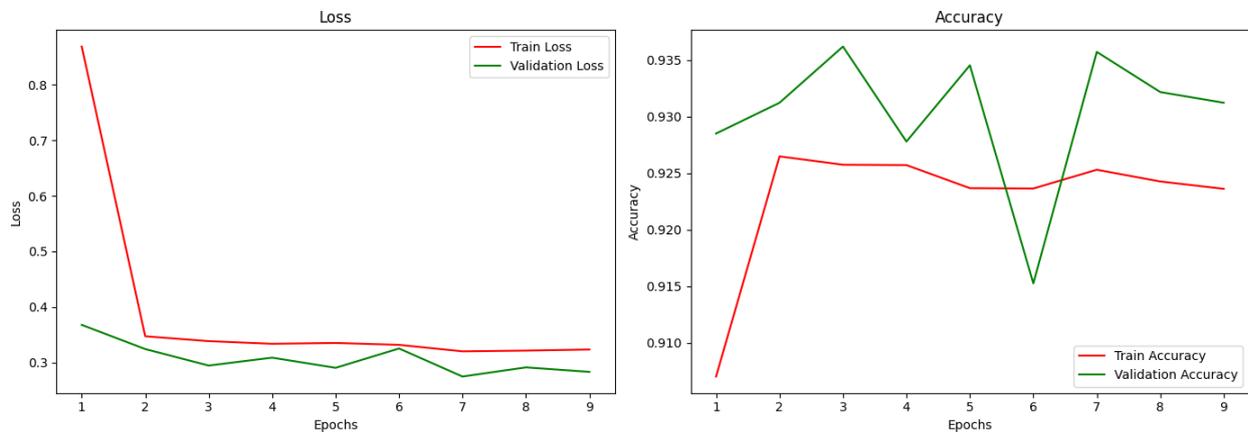
Figure 4: Model Loss and Accuracy over Training Epochs

The evaluation results on the test dataset, highlighting the model's overall performance, are presented in Figure 4.

Accuracy	Sensitivity	Specificity	Precision	F1-Score	AUC
%93.43	%100	%99.39	%94.51	%93.61	%99.59

Mean Squared Error	Variance	Bias
0.3379	3.8638	0.0018
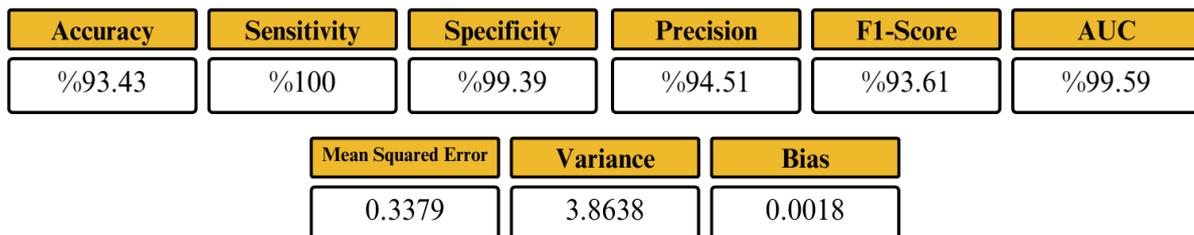
Figure 5: Model Performance Metrics

According to the test results, the model's performance is quite high. The accuracy is 93.4%, indicating that the majority of the classifications on the test data were correct. The sensitivity value is 1.0, meaning the model correctly identified all true positive examples, with no false negatives, which is crucial for critical classification tasks like skin disease detection. The specificity is calculated at 99.39%, showing that the model has a very high ability to correctly distinguish negative classes. The precision is 94.5%, indicating that the majority of the positive predictions were correct, with a low number of false positives. The F1 score reached 93.61%, indicating a well-achieved balance between precision and sensitivity. The AUC value is 99.59%, showcasing the model's outstanding capability to differentiate between positive and negative classes. The Mean Squared Error (MSE) is 0.337878, suggesting that the model's predictions are closely aligned with the true values, with minimal deviation. The variance is 3.863760, showing a wide distribution of predictions. The bias, calculated as 0.001779, indicates that there is no systematic bias in the model's predictions. These results demonstrate that the model achieved balanced learning and overall strong classification performance.

The classification metrics for each class are presented in Figure 6.
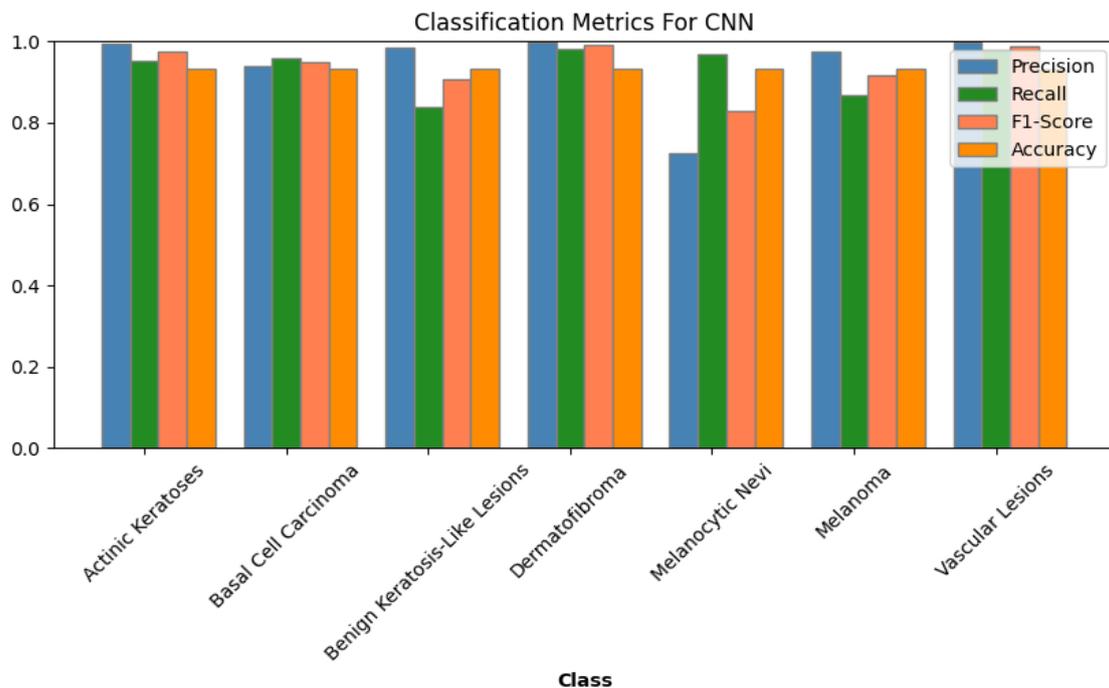
Figure 6: Classification Metrics for CNN Across Skin Lesion Classes

Figure 7 presents the confusion matrix, offering a detailed breakdown of the model's performance across each class. The matrix highlights the correct and incorrect classifications for each class, allowing for a clear visualization of how well the model distinguishes between different categories.
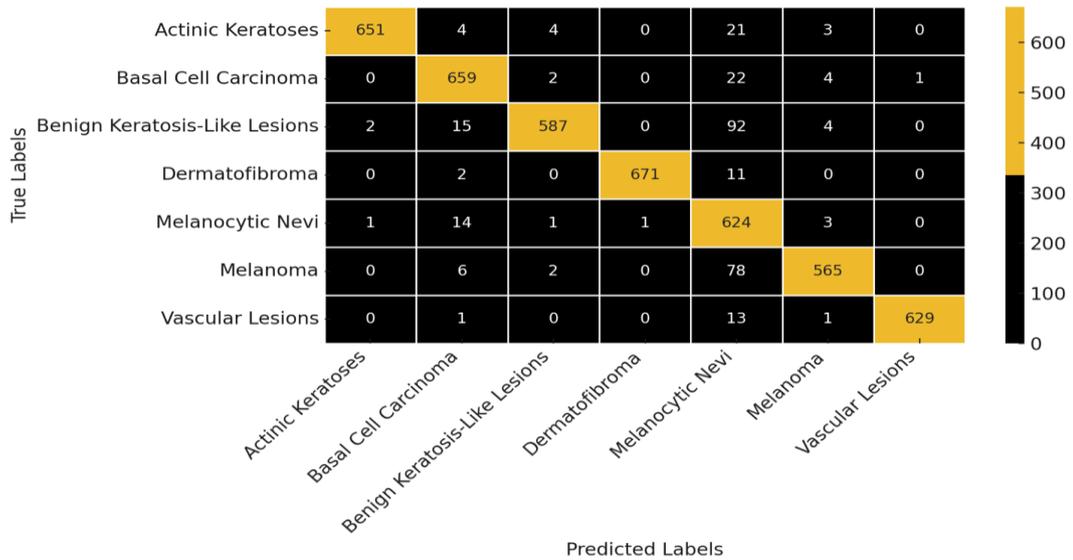
Figure 7: Confusion Matrix for Skin Lesion Classification

Upon reviewing the confusion matrix was analyzed to evaluate the model's performance for each individual class. For the Actinic Keratoses class, 651 correct positive predictions were made, with 32 misclassifications. Among these, 21 instances were misclassified as Melanocytic Nevi. For the Basal Cell Carcinoma class, the model made 659 correct predictions and 29 misclassifications. Notably, misclassifications between the Actinic Keratoses and Melanocytic Nevi classes stood out. In the Benign Keratosis-Like Lesions class, 587 correct predictions were made, while 98 examples were misclassified, mostly into the Actinic Keratoses and Melanocytic Nevi classes. The Dermatifibroma class achieved strong results, with only 13 misclassifications, predominantly in the Actinic Keratoses and Melanocytic Nevi classes. For the Melanocytic Nevi class, the model made 624 correct predictions but misclassified 19 examples, mostly as Benign Keratosis-Like Lesions. In the Melanoma class, 565 correct predictions were made, with 78 misclassifications, most of which were misclassified as Melanocytic Nevi. Lastly, in the Vascular Lesions class, 629 correct predictions were made with only 15 misclassifications. These results indicate that while the model generally performed with high accuracy, there was more confusion between certain classes, particularly between Benign Keratosis-Like Lesions and Melanocytic Nevi.

In Figure 8, the results obtained using the Grad-CAM method, which was employed to visualize the model's decision-making process, are presented. The figure displays the original image, the heatmap generated for the corresponding classification, and the overlay of the heatmap on the original image. This visualization highlights the regions of the image that the model focused on during the classification process, providing insight into its decision-making.
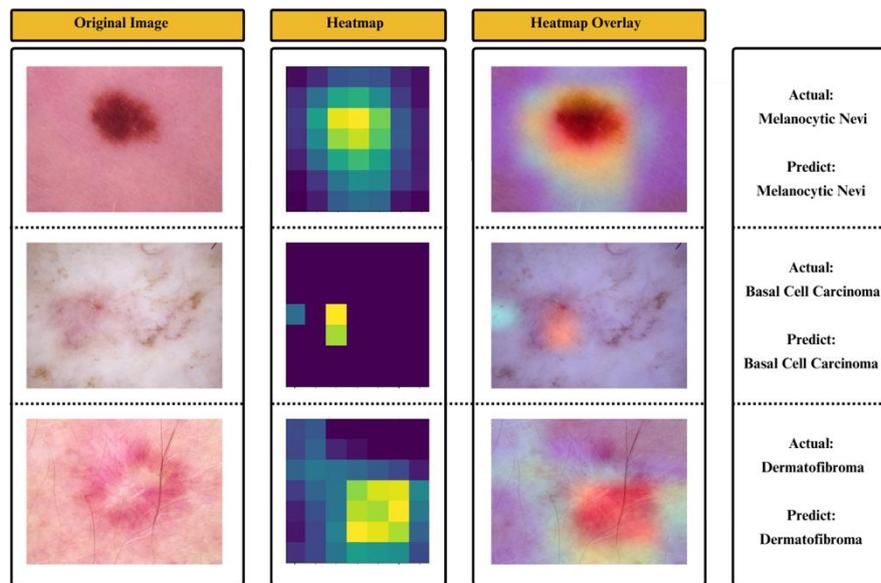
Figure 8: Grad-CAM Heatmaps for Skin Lesion Classification

In the first row, there is an image of a skin lesion that the model correctly classified as Melanocytic Nevi. Upon examining the heatmap, it is evident that the model focused primarily on the central areas of the lesion, extracting information from these regions. This indicates that the classification decision is both accurate and explainable.

In the second row, the image represents a lesion correctly classified as Basal Cell Carcinoma. The heatmap and overlay image reveal that the model concentrated its attention on the central region of the lesion, supporting the correct classification. The area of focus highlighted in the heatmap demonstrates that the model concentrated on critical regions for classification.

In the third row, a lesion classified as Dermatifibroma is shown. The model correctly focused on the lower part of the lesion to make its classification. The heatmap illustrates that the areas the model concentrated on were crucial for the classification, leading to a correct prediction.

Discussion and Future Works

This study employed a deep learning-based approach utilizing the DenseNet121 architecture, achieving effective results in skin disease classification. The experiments conducted on the HAM10000 dataset demonstrated that the model achieved high accuracy, sensitivity, and precision values [18]. However, some confusion occurred between certain classes, particularly between Benign Keratosis-Like Lesions and Melanocytic Nevi. This suggests that the model's capacity to differentiate between classes with similar visual features needs to be enhanced. The visualizations generated using the Grad-CAM method provided valuable insights into the regions from which the model extracted information, increasing the interpretability of its decisions. However, in some cases, it was observed that the model focused only on specific regions of the lesion. This indicates that the model may not be extracting sufficient information from the entire image, which could contribute to classification errors.

In future studies, more advanced data augmentation techniques could be used to improve the model's performance and minimize confusion between classes. Specifically, enhancing data augmentation methods for imbalanced datasets could lead to better learning of underrepresented classes. Additionally, transfer learning strategies could be tested on different architectures to explore the success of various neural network models in this problem domain. Working with larger and more diverse datasets could further improve the model's generalization ability. Lastly, the development of explainable AI methods like Grad-CAM could provide deeper insights into the model's decision-making processes. Such visualization techniques will play a crucial role in increasing the reliability of decision support systems in the medical field and expanding their use in clinical applications.

REFERENCES

[1] I. G. Ferreira, M. B. Weber, and R. R. Bonamigo, "History of dermatology: the study of skin diseases over the centuries," Anais brasileiros de dermatologia, vol. 96, no. 3, pp. 332-345, 2021.

[2] M. M. BERTAGNOLLI, "GASTROINTESTINAL CANCER: SURGICAL ONCOLOGY," Principles and Practice of Gastrointestinal Oncology, p. 55, 2008.

[3] M. Nasir et al., "Melanoma detection and classification using computerized analysis of dermoscopic systems: a review," Current Medical Imaging, vol. 16, no. 7, pp. 794-822, 2020.

[4] M. Sahu, R. Gupta, R. K. Ambasta, and P. Kumar, "Artificial intelligence and machine learning in precision medicine: A paradigm shift in big data analysis," Progress in molecular biology and translational science, vol. 190, no. 1, pp. 57-100, 2022.

[5] A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)," IEEE access, vol. 6, pp. 52138-52160, 2018.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, pp. 436-444, 2015.

[7] R. Dwivedi et al., "Explainable AI (XAI): Core ideas, techniques, and solutions," ACM Computing Surveys, vol. 55, no. 9, pp. 1-33, 2023.

[8] J. Gerlings, M. S. Jensen, and A. Shollo, "Explainable AI, but explainable to whom? An exploratory case study of xAI in healthcare," Handbook of Artificial Intelligence in Healthcare: Vol 2: Practicalities and Prospects, pp. 169-198, 2022.

[9] N. Zhang, Y.-X. Cai, Y.-Y. Wang, Y.-T. Tian, X.-L. Wang, and B. Badami, "Skin cancer diagnosis based on optimized convolutional neural network," Artificial intelligence in medicine, vol. 102, p. 101756, 2020.

[10] K. A. Ogudo, R. Surendran, and O. I. Khalaf, "Optimal Artificial Intelligence Based Automated Skin Lesion Detection and Classification Model," Computer Systems Science & Engineering, vol. 44, no. 1, 2023.

[11] L. Rukhsar, W. H. Bangyal, M. S. Ali Khan, A. A. Ag Ibrahim, K. Nisar, and D. B. Rawat, "Analyzing RNA-seq gene expression data using deep learning approaches for cancer classification," Applied Sciences, vol. 12, no. 4, p. 1850, 2022.

[12] H. Orhan and E. Yavşan, "Artificial intelligence-assisted detection model for melanoma diagnosis using deep learning techniques," Mathematical Modelling and Numerical Simulation with Applications, vol. 3, no. 2, pp. 159-169, 2023.

[13] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-CAM: Why did you say that?," arXiv preprint arXiv:1611.07450, 2016.

[14] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," Scientific data, vol. 5, no. 1, pp. 1-9, 2018.

[15] T. Arulananth, S. W. Prakash, R. K. Ayyasamy, V. Kavitha, P. Kuppusamy, and P. Chinnasamy, "Classification of Paediatric Pneumonia Using Modified DenseNet-121 Deep-Learning Model," IEEE Access, 2024.

[16] M. Vakili, M. Ghamsari, and M. Rezaei, "Performance analysis and comparison of machine and deep learning algorithms for IoT data classification," arXiv preprint arXiv:2001.09636, 2020.

[17] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 618-626.

[18] A. De, N. Mishra, and H.-T. Chang, "An approach to the dermatological classification of histopathological skin images using a hybridized CNN-DenseNet model," PeerJ Computer Science, vol. 10, p. e1884, 2024.

RESUME

Havva Hazel ARAS

Havva Hazel ARAS completed the English Preparatory Program at Necmettin Erbakan University, Konya, in 2014. She pursued her undergraduate studies in Computer Engineering at Selcuk University, Konya, graduating in 2019. Currently, she is undertaking a master's degree in Computer Engineering at Selcuk University, which she began in 2024. ARAS has a diverse professional background, having held positions such as Product Specialist, Business Analyst, and Product Manager in reputable organizations including Pingpong University, Solvix Tech, and Mindbehind-Insider. Her expertise encompasses software development, AI integration, and project management, with significant contributions to chatbot development, agile project execution, and product life cycle management. Additionally, ARAS has authored several academic and technical works in the fields of artificial intelligence and image processing. Her commitment to bridging theoretical knowledge with practical applications continues to drive meaningful advancements in her field.

Prof. Dr. Nurettin DOĞAN

Nurettin DOĞAN is a Professor at the Department of Computer Engineering, Faculty of Technology, Selçuk University. He received his Master's and PhD degrees from Ankara University. He has published papers in various national and international journals on image encryption, approximate analytic serial solutions methods and artificial intelligence.

IMPROVING SYNTHETIC DATA GENERATION IN FINANCE WITH FEATURE SCALING AND OUTLIER REMOVAL

Ahmet Yasin AYTAR[1], Abdurrahman DEMİRLİ[2]

INTRODUCTION

Synthetic data generation has emerged as a practical approach for producing artificial datasets that replicate the statistical properties of real-world data. The need for synthetic data has become increasingly relevant in finance due to the restrictions on using and sharing sensitive financial information. By creating synthetic data, organizations can share information with third parties while preserving privacy, comply with regulatory requirements, and address class imbalances in machine learning applications.

However, generating synthetic tabular data that reliably reflects the complexity of financial datasets remains challenging. Financial data often contain mixed data types, complex feature relationships, and irregular distributions, such as skewness and outliers, all complicating the data generation process. Effective techniques are required to manage these characteristics while maintaining the practical utility of the synthetic data.

This study proposes an evaluation framework and preprocessing methods tailored for Generative Adversarial Network (GAN)-based approaches to synthetic data generation in finance. The framework assesses synthetic data quality along five dimensions—fidelity, privacy, utility, outlier management, and feature relationship preservation. In addition, we introduce a two-step preprocessing process: a scaling selection layer that adapts normalization techniques for each feature and an outlier elimination layer that uses the Interquartile Range (IQR) method to reduce the impact of extreme values. These methods aim to stabilize data input to the GAN, enhancing its ability to capture essential patterns in financial data.

We applied the proposed framework to two financial datasets: an internal bank dataset and a publicly available dataset from Kaggle, using three GAN architectures—standard GAN, Wasserstein GAN with Gradient Penalty (WGAN-GP), and Conditional Tabular GAN (CTGAN) [1]. This choice allows a comprehensive evaluation of the preprocessing methods' impact across architectures. The study's contributions lie in establishing a versatile framework for synthetic data assessment and introducing preprocessing techniques that can improve GAN performance in generating representative financial data.

---

[1] Architecht Bilişim Sistemleri ve Pazarlama Ticaret A.Ş, Istanbul/Turkey, Orcid: https://orcid.org/0009-0006-9644-2592, ahmet.aytar@architecht.com

[2] Architecht Bilişim Sistemleri ve Pazarlama Ticaret A.Ş, Istanbul/Turkey, Orcid: https://orcid.org/0009-0009-9948-3332, abdurrahman.demirli@architecht.com

LITERATURE REVIEW

Generative Adversarial Networks (GANs) have become widely adopted for synthetic data generation across various fields, particularly in finance, where data privacy and regulatory restrictions often limit information sharing. Originally introduced by Goodfellow et al., GANs operate through adversarial training between a generator and a discriminator, which together approximate the distribution of real data [2]. Although primarily developed for image data, the GAN architecture has been adapted for tabular data generation through models like the Conditional Tabular GAN (CTGAN), which is designed to handle both categorical and numerical features in structured data settings [3]. However, recent research highlights that CTGAN may underperform on datasets with a predominance of continuous features or skewed distributions, conditions often encountered in financial datasets [4].

To address these challenges, the Wasserstein GAN with Gradient Penalty (WGAN-GP) has emerged as an alternative architecture, noted for its enhanced stability and fidelity in capturing data distributions through the use of a Wasserstein distance-based loss function [5]. WGAN-GP has shown particular efficacy in modeling complex, heavy-tailed data structures, making it suitable for high-fidelity synthetic data generation within financial contexts [6]. Studies by Eckerli and Osterrieder demonstrate that WGAN-GP can replicate financial datasets with high fidelity, making it an increasingly popular choice in applications where data fidelity and training stability are paramount [4].

Preprocessing techniques such as feature scaling and outlier removal are essential for optimizing GAN performance. Research by Hastie et al. emphasizes the role of scaling methods like MinMaxScaler and RobustScaler in standardizing data, reducing the influence of outliers and thereby enhancing model stability [7]. In one study, it is highlighted that employing the Interquartile Range (IQR) method for outlier removal significantly benefits simpler GAN architectures, like standard GAN and WGAN-GP, by reducing the impact of extreme values on generated data quality [8].

Evaluating synthetic data quality requires metrics that assess fidelity, privacy, and utility. Commonly used metrics include Kullback-Leibler (KL) divergence, Kolmogorov-Smirnov (KS) distance, and the Bhattacharyya coefficient to quantify how closely synthetic data aligns with the original data distributions [9, 10, 11, 12]. In finance, it is essential to preserve statistical and structural features for downstream tasks such as predictive modeling. Metrics such as the count of identical rows between synthetic and original datasets (for privacy) and feature importance consistency (for utility) are necessary to ensure synthetic data applicability [4].

This study builds upon existing research by incorporating preprocessing steps, including feature scaling and outlier removal, within GAN-based synthetic data frameworks. These enhancements led to improved alignment with original data distributions, as evidenced by reductions in KL divergence and KS distance, especially in standard GAN and WGAN-GP models. By utilizing a comprehensive evaluation framework that assesses distributional similarity and correlation preservation, our approach contributes to the growing emphasis on

robust synthetic data quality assessment, particularly in fields requiring high sensitivity to data fidelity, such as finance.

METHODS

Data Description

The evaluation of our synthetic data generation framework is conducted on two distinct datasets: one from an internal bank dataset and one publicly available dataset sourced from Kaggle. The internal bank dataset comprises 18 features and 16,408 rows of data, while the Kaggle dataset contains 64 features and 10,000 rows. All datasets share a common structure, consisting of financial ratios as continuous features and a target variable indicating bankruptcy status (binary classification). The financial ratios include various indicators related to liquidity, profitability, and leverage, while the target variable classifies the entities as either bankrupt or non-bankrupt.

Evaluation Framework

This study evaluates synthetic data quality through key metrics that assess fidelity, privacy, utility, outlier representation, and feature relationship preservation. Fidelity is examined with metrics such as Kullback-Leibler (KL) divergence, Bhattacharyya coefficient, and Kolmogorov-Smirnov (KS) distance. KL divergence measures the probability distribution differences between original and synthetic datasets, with lower values indicating better alignment. The Bhattacharyya coefficient assesses distributional overlap, where values closer to 1 indicate higher fidelity. The KS distance further evaluates the maximum difference in cumulative distributions, with smaller values reflecting closer alignment between the two datasets.

Privacy metrics ensure that synthetic data maintains confidentiality by limiting resemblance to the original data. We use two methods: counting identical rows and calculating a nearest neighbor's privacy score. A minimal or zero count of identical rows indicates strong privacy preservation, as it reduces the risk of replication of original data points. The nearest neighbors score complements this by assessing how closely synthetic data points resemble the nearest points in the original data; a lower score suggests better privacy protection.

Utility is evaluated through feature importance consistency and machine learning model performance to ensure synthetic data's effectiveness in practical applications. By comparing feature importance across models trained on both original and synthetic data, we check for the preservation of key data characteristics. Additionally, model performance metrics such as accuracy and F1 score gauge the synthetic data's predictive validity. When these metrics are closely aligned, it suggests that synthetic data can effectively support downstream tasks.

Outlier representation and feature relationship preservation are critical for retaining the data's structural properties. Outlier analysis confirms that synthetic data does not introduce or overlook significant deviations that could distort analyses. Feature relationship preservation is assessed through correlation analysis, comparing absolute differences in feature correlations between original and synthetic data. Additionally, the log-cluster metric is employed to evaluate how well synthetic data aligns with the clustering structure of the original data. By performing

k-means clustering on the combined original and synthetic datasets, this metric calculates the ratio of original points within each cluster, with a lower log-transformed score indicating closer alignment. Preserving these relationships is crucial for fields where feature interactions underpin predictive tasks. This combined framework thus provides a comprehensive and systematic evaluation of synthetic data quality.

Preprocessing Layers

Our synthetic data generation framework incorporates two critical preprocessing methods—scaling and outlier elimination layers—designed to prepare data for Generative Adversarial Networks (GANs). These layers improve the quality of synthetic tabular data by ensuring that the input to GAN models is well-conditioned, which is essential for stability and accuracy during model training. This preprocessing approach is systematically integrated with three GAN architectures: standard GAN, Wasserstein GAN with Gradient Penalty (WGAN-GP), and Conditional Tabular GAN (CTGAN). By applying these methods, we aim to enhance the GANs' ability to replicate the distributional properties of the original dataset, ultimately improving the fidelity and utility of the generated synthetic data.

Scaling Layer

The scaling layer standardizes numerical features in the dataset, helping to stabilize GAN training by ensuring that features are on compatible scales. Standardization is crucial because unscaled or inconsistently scaled features can disrupt GAN learning, especially when certain features dominate due to larger ranges. To address this, we apply a variety of scaling techniques, including RobustScaler, MinMaxScaler, Log Transformation, and PowerTransformer variants (Box-Cox and Yeo-Johnson). Each technique offers unique advantages based on the distribution characteristics of the data. For example, RobustScaler scales data using the median and interquartile range, which minimizes the impact of outliers, while MinMaxScaler maps data to a defined range, typically 0 to 1, improving feature comparability.

Selecting the optimal scaling method for each feature is guided by the Kolmogorov-Smirnov (KS) test, a non-parametric measure that evaluates how well the transformed data matches the original distribution. This process follows a structured sequence:

• **Step 1:** Evaluate each numerical feature independently.

• **Step 2:** Apply different scaling techniques to each feature (RobustScaler, MinMaxScaler, Log Transformation, PowerTransformer (Box-Cox), PowerTransformer (Yeo-Johnson)).

• **Step 3:** Conduct the Kolmogorov-Smirnov (KS) test to compare the cumulative distribution functions of the original and transformed data.

• **Step 4:** Calculate the KS statistic to quantify the maximum difference between these distributions.

• **Step 5:** Select the scaling technique that results in the lowest KS statistic for each feature, indicating the most effective normalization.

This method of individualized scaling ensures that each feature is transformed in a way that best aligns with its original distribution, providing a consistent input for the GAN models. Using optimally scaled data promotes more stable and efficient convergence, enhancing the GANs' capacity to accurately learn the underlying data distribution.

Outlier Layer

The outlier elimination layer serves to filter out extreme values from the dataset before GAN training. Outliers can distort GAN learning by skewing the representation of the original data distribution, making it essential to address these anomalies in advance. We use the Interquartile Range (IQR) method for outlier detection and removal, as it is effective in identifying extreme values without assuming a normal distribution. This method is applied to each numerical feature independently to provide robust preprocessing across the dataset.

The IQR-based outlier elimination process is carried out as follows:

• **Step 1:** Calculate the IQR as the difference between the third quartile (Q3) and the first quartile (Q1) of the feature's data distribution.

• **Step 2:** Establish outlier thresholds:

o Lower bound: Q1 minus 3 times the IQR ($Q1 - 3 \times IQR$).

o Upper bound: Q3 plus 3 times the IQR ($Q3 + 3 \times IQR$).

• **Step 3:** Identify data points falling outside these thresholds as outliers.

• **Step 4:** Remove the identified outliers from the dataset.

By excluding these extreme values, the GAN models are presented with data that represents the core structure of the original distribution, minimizing distortions that could arise from atypical observations. This step is critical for ensuring that the GAN learns from a clean, representative dataset, which ultimately contributes to producing synthetic data that better reflects the inherent patterns and characteristics of the original data.

With both scaling and outlier elimination layers implemented, our preprocessing framework establishes a well-prepared dataset that facilitates stable GAN training across the standard GAN, WGAN-GP, and CTGAN architectures. Together, these methods reduce noise, stabilize training, and improve the fidelity of the synthetic data generated, ensuring it is both statistically representative and suitable for practical applications.

Results

Evaluation Framework

Our evaluation framework is a comprehensive tool for assessing synthetic data quality generated by various GAN architectures, focusing on essential metrics such as fidelity, privacy, utility, outlier handling, and feature relationship preservation. Developed as a Streamlit application, the framework allows users to upload datasets (up to 200MB), configure GAN parameters, and generate synthetic data through an intuitive interface.

As shown in Figure 1, users can set key hyperparameters on the left sidebar, such as the number of epochs, batch size, learning rate, and the specific GAN model (e.g., CTGAN, WGAN-GP).

After adjusting these settings, users can proceed by clicking "Generate Synthetic Data" below the data summary section. The framework then processes the data according to the chosen GAN model and user-defined parameters, displaying initial data summaries to confirm the dataset's successful upload and integrity.
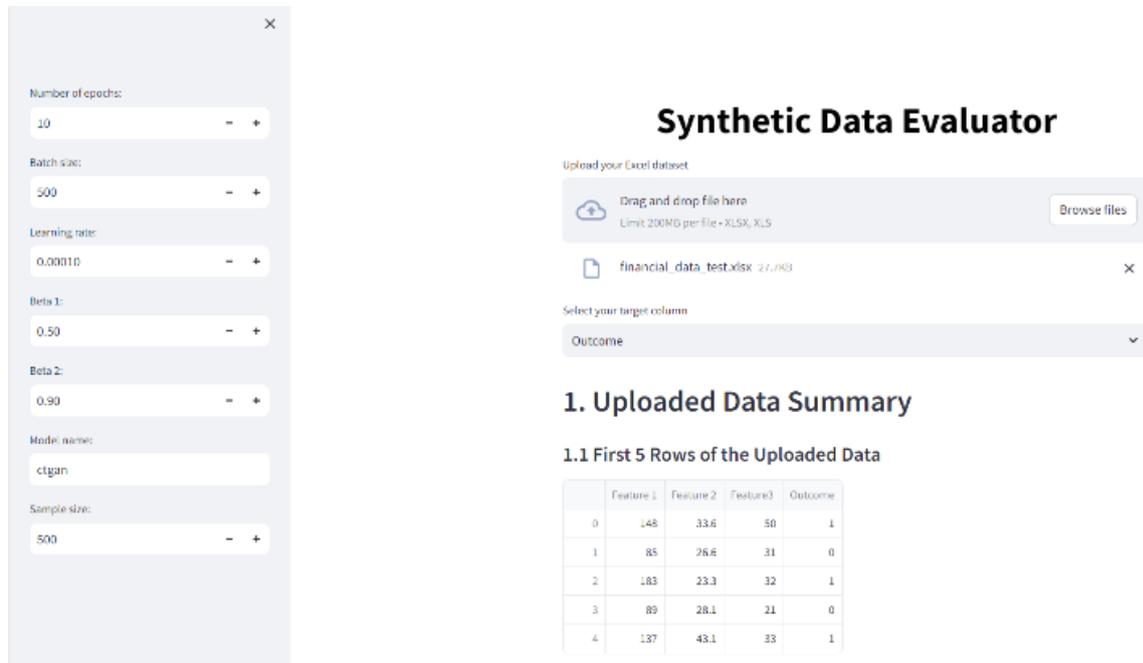
Figure 1: Initial interface of the evaluation framework with data upload, summary display, and GAN parameter settings.

Once synthetic data is generated, the framework evaluates its quality across all primary metrics. Figure 2 illustrates the Feature Relationship Preservation section, which includes both a correlation graph and the log-cluster metric. The correlation graph assesses how well the synthetic data maintains relationships between features, while the log-cluster metric calculates the alignment of clustering behavior between original and synthetic datasets. Together, these metrics help novice users understand the synthetic data's structural alignment with the original data. A lower log-cluster value and close correlation alignment indicate that the synthetic data retains the original dataset's underlying relationships, reflecting high quality.

This evaluation framework thus provides a seamless workflow—from data upload and configuration to an in-depth quality assessment of the generated synthetic data. By offering comprehensive insights into each metric, it guides users in determining the reliability and applicability of synthetic data for real-world financial tasks.

6. Feature Relationship Preservation

6.1 Correlation Analysis
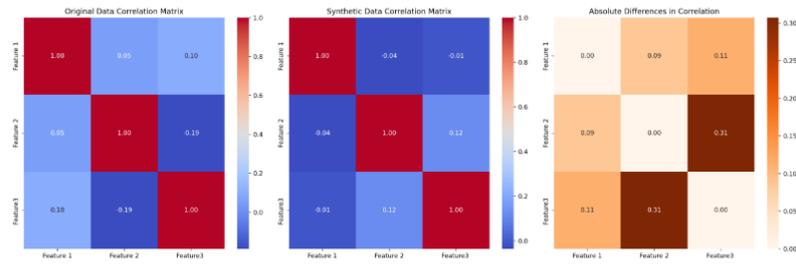
Average of absolute differences in correlations: `0.1125`

6.2 Distinguishability Log-cluster

Log-cluster metric performs k-means clustering on the combined original and synthetic datasets, calculates the ratio of original points in each cluster compared to the expected proportion, and then computes a log-transformed sum of squared ratio differences to form a metric. A lower value of this metric indicates that the synthetic data closely mimics the distribution of the original data within the clusters, which suggests a higher quality of synthetic data that preserves the structure of the original dataset.

Log-cluster metric: `-4.054`

Figure 2: Feature Relationship Preservation section showing correlation graphs and log-cluster metric to assess feature relationships in synthetic data.

Preprocessing Layers

Table 1: Performance Metrics for Synthetic Data Generation Models on the Bank Dataset

Algorithms	KL Div.	KS Dist.	Bh. Coef.	Abs. Diff. In Corr.
GAN	8.7126	0.00033	0.816	0.2966
Scaling Layer GAN	5.924	0.00017	0.938	0.4577
Outlier Layer GAN	4.2579	0.00045	0.7295	0.3128
Fully Preprocessed GAN	11.486	0.00019	0.9525	0.6444
WGANGP	12.4425	0.00016	0.990	0.3519
Scaling Layer WGANGP	8.0099	0.00013	0.9081	0.3164
Outlier Layer WGANGP	11.767	0.00012	0.92788	0.4084
Fully Preprocessed WGANGP	10.9225	0.00024	0.9221	0.7152
CTGAN	0.640	0.00076	0.5036	0.077
Scaling Layer CTGAN	1.380	0.00049	0.6427	0.0734
Outlier Eliminated CTGAN	0.6112	0.00076	0.4993	0.0796
Fully Preprocessed CTGAN	1.4025	0.00061	0.6210	0.0906

In this study, we evaluate synthetic data generation models using various configurations of GAN-based architectures on two datasets—a bank dataset and a Kaggle dataset. Performance is assessed with four key metrics to highlight statistical similarity with the original data, focusing on distributional and relational alignment: KL divergence, Bhattacharyya coefficient, KS distance, and absolute differences in correlation. For simplicity, we use these core metrics to concentrate on statistical similarity as a foundation for the study. All comparisons are made against the original data to provide a baseline reference. To be clear, "Fully Preprocessed" configurations include both scaling and outlier elimination layers, applied sequentially (outlier elimination first) alongside the GAN model.

Table 2: Performance Metrics for Synthetic Data Generation Models on the Kaggle Dataset

Algorithms	KL Div.	KS Dist.	Bh. Coef.	Abs. Diff. In Corr.
GAN	7.855	0.00055	0.6877	0.2607
Scaling Layer GAN	8.144	0.00025	0.8592	0.3362
Outlier Layer GAN	8.7052	0.00051	0.5514	0.2484
Fully Preprocessed GAN	11.400	0.00032	0.8055	0.2571
WGANGP	9.109	0.00052	0.702	0.2451
Scaling Layer WGANGP	10.915	0.00018	0.8873	0.3144
Outlier Layer WGANGP	10.605	0.00046	0.6310	0.2556
Fully Preprocessed WGANGP	14.383	0.00023	0.867	0.2681
CTGAN	1.1824	0.00075	0.5920	0.211
Scaling Layer CTGAN	2.5194	0.00030	0.7643	0.2092
Outlier Eliminated CTGAN	1.428	0.00081	0.3685	0.2162
Fully Preprocessed CTGAN	2.4922	0.00045	0.64815	0.2162

Bank Data Analysis

Results indicate that preprocessing layers impact model performance differently. For the GAN model, the scaling layer reduced KL divergence (from 8.71 to 5.92), suggesting improved distributional alignment, though with a slight increase in the correlation difference. The outlier layer further lowered KL divergence to 4.26 and maintained stable correlation differences, indicating that handling outliers effectively enhances synthetic data quality. Notably, the fully preprocessed GAN configuration did not yield additional gains and even showed some decline in performance, with higher KL divergence (11.49).

For WGAN-GP, the outlier layer marginally improved KS distance and the Bhattacharyya coefficient, though the fully preprocessed configuration showed little added benefit. CTGAN initially demonstrated strong alignment with the original data (KL divergence of 0.64), and

preprocessing did not significantly improve its performance, suggesting that CTGAN may inherently suit this dataset without extensive preprocessing.

Kaggle Data Analysis

On the Kaggle dataset, outlier elimination consistently benefited GAN and CTGAN, lowering KL divergence and improving KS distance, particularly for CTGAN (KL divergence reduced from 1.18 to 1.43). However, combining both preprocessing layers generally did not yield cumulative benefits and, in some cases, increased KL divergence, as seen with the "Fully Preprocessed GAN" and "Fully Preprocessed WGAN-GP" configurations. This pattern suggests that targeted preprocessing, especially outlier handling, is more effective than combining layers indiscriminately.

Overall, the analysis highlights that preprocessing—particularly the outlier elimination layer—can enhance distributional fidelity, though its impact varies depending on the model and dataset. The scaling layer's effectiveness was inconsistent, while fully preprocessed configurations often showed mixed results. These findings emphasize the importance of selective preprocessing tailored to the dataset characteristics and the specific GAN model.

Discussion

The results demonstrate that incorporating preprocessing layers, such as scaling and outlier elimination, enhances data quality in GAN-based synthetic data generation frameworks, especially for standard GAN and WGAN-GP models. Although CTGAN naturally performs well with tabular data, these preprocessing techniques allow simpler GAN models to achieve comparable gains. This highlights the potential of standard GAN and WGAN-GP models when preprocessing is applied, particularly for datasets dominated by continuous features or when simpler models are preferred for efficiency.

The improvements were more notable in the bank dataset, with its complex and skewed distributions, than in the Kaggle dataset. For the GAN model, reductions in KL divergence and KS distance demonstrated a closer alignment to the original data, underscoring the effectiveness of preprocessing in handling real-world data characteristics. For WGAN-GP, while most metrics improved, the outlier elimination layer had a limited impact, suggesting that further refinement of outlier handling could enhance its performance. Overall, preprocessing proved especially beneficial for datasets with heavy tails and non-standard distributions.

This study also emphasizes the flexibility and potential advantages of using standard GAN and WGAN-GP models in scenarios where the dataset is largely continuous or simple in structure. These models offer opportunities for custom loss functions and architectural adjustments, making them practical alternatives when computational efficiency is prioritized.

The comparative analysis between bank and Kaggle datasets underscored the robustness of the preprocessing methods. The bank dataset's irregular and skewed distributions were representative of real-world challenges, and the preprocessing strategies improved data quality in this context, showing that these methods apply to complex financial data modeling.

The evaluation framework, which included multiple metrics for assessing distributional similarity, correlation preservation, and outlier presence, was effective for comparing synthetic

data quality across GAN configurations. This comprehensive framework identified areas where preprocessing could be refined, ensuring that synthetic data maintains statistical and relational properties crucial for downstream applications.

Future research could improve outlier handling with adaptive, feature-specific thresholds and explore advanced preprocessing strategies that consider inter-feature dependencies. Additionally, combining CTGAN's handling of categorical data with WGAN-GP's customizable loss functions could lead to hybrid models better suited for diverse data types.

CONCLUSION

This study examined the effects of adding scaling and outlier elimination preprocessing layers to GAN-based models for generating synthetic financial data, particularly for bankruptcy datasets. Results showed that these preprocessing steps substantially enhanced data quality for standard GAN and WGAN-GP models, especially with the bank dataset, which featured complex and realistic distributions. Improvements were most noticeable in metrics assessing distributional similarity, indicating closer alignment between synthetic and original data. While CTGAN inherently performed well due to its tabular data focus, the preprocessing layers enabled standard GAN and WGAN-GP to become viable alternatives, particularly in cases where categorical handling is less critical or model customization is preferred. The evaluation framework proved robust for assessing synthetic data quality, suggesting it may be broadly applicable. Future work could further refine outlier detection with adaptive thresholds or explore hybrid models that integrate the strengths of multiple GAN architectures to enhance performance on complex datasets.

REFERENCES

[1] https://www.kaggle.com/c/companies-bankruptcy-forecast/data

[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," Communications of the ACM, vol. 63, pp. 139-144, 2014.

[3] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling tabular data using conditional GAN," Advances in Neural Information Processing Systems, vol. 32, 2019.

[4] F. Eckerli and J. Osterrieder, "Generative adversarial networks in finance: An overview," Machine Learning eJournal, 2021.

[5] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs with gradient penalty," Advances in Neural Information Processing Systems, vol. 30, 2017.

[6] F. de Meer Pardo, "Enriching financial datasets with generative adversarial networks," M.S. thesis, Delft University of Technology, The Netherlands, 2019.

[7] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., New York: Springer, 2017.

[8] A. Azimi, B. Boboeva, I.E. Varshavskiy, S. Khalilbekov, A. Nizamitdinov, N. Noyoftova, and S. Shulgin, "zGAN: An Outlier-focused Generative Adversarial Network for Realistic Synthetic Data Generation," 2024.

[9] P. Pilar and N. Wahlstrom, "Probabilistic matching of real and generated data statistics in generative adversarial networks," ArXiv, abs/2306.10943, 2023.

[10] S. Kullback and R.A. Leibler, "On information and sufficiency," Annals of Mathematical Statistics, vol. 22, pp. 79-86, 1951.

[11] N.V. Smirnov, "Table for estimating the goodness of fit of empirical distributions," Annals of Mathematical Statistics, vol. 19, no. 2, pp. 279-281, 1948.

[12] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," Bulletin of the Calcutta Mathematical Society, vol. 35, pp. 99-109, 1943.

RESUME

Ahmet Yasin AYTAR

Ahmet Yasin Aytar is an accomplished data scientist with a solid academic foundation spanning prestigious institutions. He began his undergraduate studies in Electrical and Electronics Engineering at Middle East Technical University (METU) before transferring to Bogazici University, where he earned his degree in Mechanical Engineering. Currently, he is pursuing a Master of Science in Data Science at Sabanci University. His academic journey is marked by a focus on advanced data-driven methodologies, including machine learning, deep learning, and big data processing, equipping him with the skills to address complex analytical challenges. Professionally, Ahmet serves as a Data Scientist at Architecht, where he develops innovative solutions to enhance machine learning operations and optimize customer experience through predictive analytics. His role reflects a commitment to leveraging data science to create impactful business solutions. Ahmet has also contributed to academic research and open-source development, underscoring his dedication to advancing the field through both theoretical and practical innovation.

Abdurrahman DEMİRLİ

Abdurrahman Demirli began his undergraduate studies at Sabancı University in 2016 and graduated in 2021 with a degree in Industrial Engineering. Alongside his major, he completed a minor in Finance, showcasing his interdisciplinary academic background.

Since 2021, Demirli has been working as an AI Engineer at Architecht, where he has contributed to projects focused on CRM systems and Generative AI. His work involves designing and implementing innovative AI solutions that enhance customer relationship management and leverage the power of generative technologies. With a strong foundation in both engineering and finance, Demirli combines technical expertise and strategic insights to develop impactful AI-driven solutions.

APPLICATION OF DEEP LEARNING ALGORITHMS IN CRACK DETECTION AFTER MAGNETIC PARTICLE TESTING

Yasin ULUS[1], Mehmet Akif ŞAHMAN[2]

INTRODUCTION

In today's industrial production and structural safety, material evaluation and defect are of critical importance. The early detection of cracks in metals and similar materials plays a vital role in structural integrity and operational safety. Magnetic particle inspection (MPI), also known as MPT, is a non-destructive testing (NDT) technique used to detect surface and shallow subsurface defects in ferromagnetic materials.[1] When magnetic particles are introduced onto the surface, they are attracted by the magnetic flux leakage MFL, forming visible indications that inspectors can detect and assess.[2] MPT is a widely used non-destructive testing method for detecting surface cracks in materials. However, traditional MPT techniques are often time-consuming and may be limited in accurately detecting underlying detects, as they typically rely on human-based interpretation and manual analysis.[3] The significance of a crack depends on its length, width, depth, and location. Therefore, monitoring the structural health, reliability, and performance is essential for the long-term serviceability of the infrastructure. Visual inspection is the prevalent form of crack detection, which is a slow, labour intensive task. The results of these of these inspections rely on the skill, experience, and subjectivity of the inspector[4]. Deep learning is able to learn the internal laws and presentation levels of the sample data. The information obtained in the learning process is helpful to the interpretation of data. Deep learning has made many achievements in search technology, data mining, machine learning, machine translation, natural language processing, multimedia learning, recommendation and personalization technology, and other related fields. It enables machines to imitate human activities, such as audiovisual and thinking, solving many complex pattern recognition problems, which furthers the significant progress made in artificial intelligence related technologies [5]. In recent years, the advancement of artificial intelligence techniques, such as deep learning models, has offered new possibilities in the field of industrial inspection and defect detection. Deep learning algorithms, especially CNNs, have the ability to automatically learn and recognize patterns from large amounts of data. Due to these capabilities, the use of deep learning models for detecting cracks in MPT images holds significant potential. CNNs are deep learning architectures that have shown successful results in image processing and pattern recognition. CNNs learn features hierarchically from images, enabling them to extract high-level features directly from pixel values. It is expected that CNNs will produce unique and effective results for detecting cracks in MPT images by processing the raw image data. In this study, it is planned to combine different feature extraction methods such as

[1] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0009-0008-3859-8016, yasin.ulus91@gmail.com
[2] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0000-0002-1718-3777, asahman@selcuk.edu.tr.com

Convolutional Neural Networks and Histogram of Oriented Gradients (HOG). CNNs will extract high-level features from the images, and by integrating them with traditional methods like HOG, the crack detection performance in magnetic particle testing images will be enhanced.

MATERIALS AND METHOD

Non-destructive testing are methods to evaluate material integrity for surface or internal flaws or metallurgical condition without interfering in any way with the destruction of the material or its suitability for service [6]. There are varieties of methods to evaluate materials and components as per their state of application. The field of Non-Destructive Evaluation (NDE) or Non-Destructive Testing (NDT) involves the identification and characterization of damages or defect on the surface and interior of materials without cutting apart or otherwise altering the material. In other words, NDT refers to the assessment or evaluation and inspection process of materials or components for characterization or finding defects and flaws in comparison with some standards without altering the original attributes or harming the object being tested. NDT techniques make available or provide a cost effective means of testing of a sample for individual investigation and examination or may be applied on the whole material for checking in a production quality control system [7]. Magnetic Particle Testing, a non-destructive inspection method used for detecting surface defects, aims to locate discontinuities on the material surface. In this technique, a leakage flux is generated on the material. Magnetic particles distributed over the material are drawn to the areas of leakage flux, accumulating around surface discontinuities. This process makes it possible to identify the locations of the discontinuities. Magnetic particle tests can be applied to all ferromagnetic materials. Leakage fluxes on the material are generated on the setup shown in Figure 1.

Figure 1: Magnetic test bench

The material is prepared to interact with a magnetic field, enabling it to gain a magnetic characteristic for inspection. The magnetization process can be carried out using different

methods electromagnetic methods: A magnetic field is created using an electromagnet or electromagnetic coils. Semi-magnetic methods: The material is magnetized using a temporary magnet or a natural magnetic field. The magnetization of the material is shown in Figure 2.

Figure 2: Magnetizing the material

Magnetic particles are applied to the surface or interior of the magnetized material. These particles are typically substances with magnetic properties, such as iron powder or titanium dioxide. Figure 3. shows an image of the material with magnetic particles applied, taken under ultraviolet light.

Figure 3: Material subjected to magnetic particle testing

R, G, B (red, green, blue respectively) values are the fundamental components used to define the colors of a pixel. An image consists of pixels in a digital medium, and each pixel is represented by a specific combination of red, green, and blue color channels. Each color channel (R, G, B) has a value between 0 and 255. The combination of these values determines the color of each pixel. In this study, a Canon EOS 2000D DSLR digital camera was used to capture images. The images were taken in an area isolated from the external environment, where magnetic particle testing was applied and held under ultraviolet light. The resolution of the captured images was set to 1510x1499 pixels, as shown in Figure 4. HOG is a feature extraction technique used to represent an object or scene in an image. It creates a histogram using gradient

information in the image, aiming to describe the object's shape and structure. The steps of HOG feature extraction are as follows: First, the input image is resized to 64x128 pixels. Then, gamma normalization is applied to this new image. The image is divided into cell grids of 8x8 pixels. Next, a sliding window of 16x16 pixels is placed on top of the grid and shifted between cells. The histograms are then normalized and combined to form a one-dimensional feature vector.
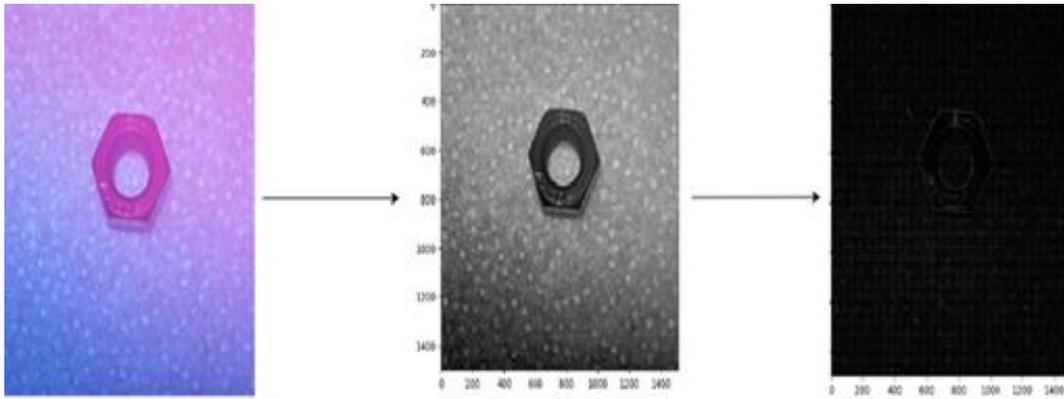
Figure 4: Gradient transform image

Then, the gradients of each pixel in the image are calculated using the formulas in Equations 1 and 2:

$$dx = I(x + 1, y) - I(x, y) \tag{1}$$
$$dy = I(x, y + 1) - I(x, y) \tag{2}$$

where, d_x and d_y are the horizontal and vertical gradients, respectively, and $I(x, y)$ is the pixel value at the position (x, y). The gradient orientation Q is calculated using the formula in Equation 3.

$$Q(x, y) = tan^{-1}(\frac{d_y}{d_x}) \tag{3}$$

A histogram of the gradient orientation is created for each block. This is done by voting the orientation angles of each pixel into a predefined number of histogram bins. Figure 5 shows the transformation of the scaled image into HOG.
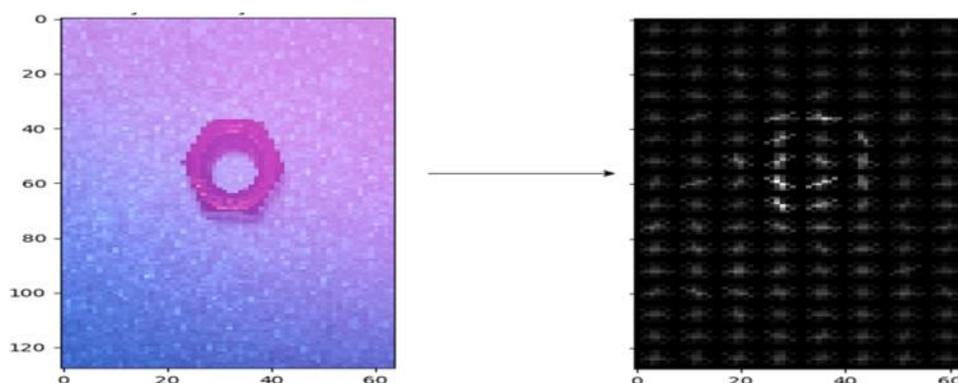
Figure 5: Conversion of scaled image to HOC

Deep Learning

Deep learning (DL) is a machine learning technique based on artificial neural networks, used for learning from complex data. It is employed to recognize patterns, extract features, identify images, and ultimately make predictions within a certain logical framework based on these data from large data pools. Deep learning enables computation models consisting of multiple layers to learn data representations with various levels of abstraction. These methods have significantly advanced state-of-the-art technology in various fields such as speech recognition, visual object recognition, and object detection. Deep learning explores the complex structure of large datasets using the backpropagation algorithm to specify how a machine should adjust its internal parameters to compute the representation at each layer based on the representation from the previous layer. While deep convolutional networks have led to groundbreaking developments in processing images, videos, speech, and sound, recurrent networks have shed light on sequential data such as text and speech [8].
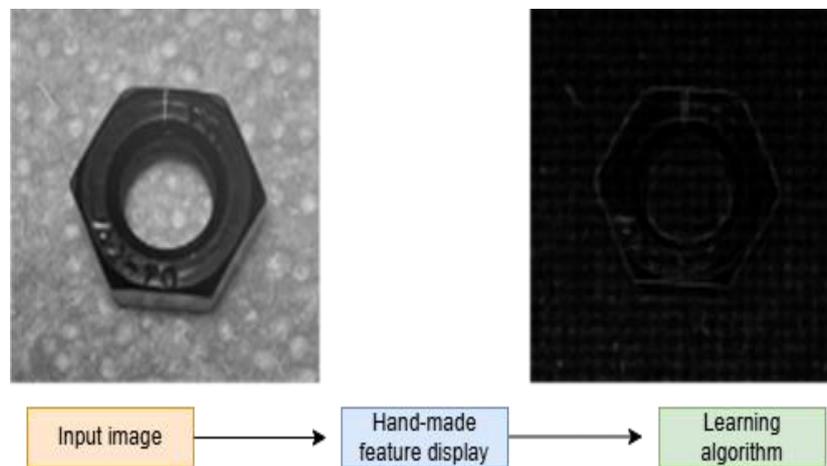
Figure 6: Representation learning

Figure 6 shows the representation learning method using HOG feature extraction, which is one of the traditional pattern recognition methods.

Artifical Neural Networks

The human brain has vast neural networks capable of succeeding in cognitive, perceptual, and control tasks where humans excel. The brain has the ability to perform computationally challenging perceptual actions (e.g., recognizing faces, understanding speech) and control activities (e.g., body movements and bodily functions). The human brain is a collection of over 10 billion interconnected neurons. Each neuron is a cell that uses biochemical reactions to receive, process, and transmit information [9].
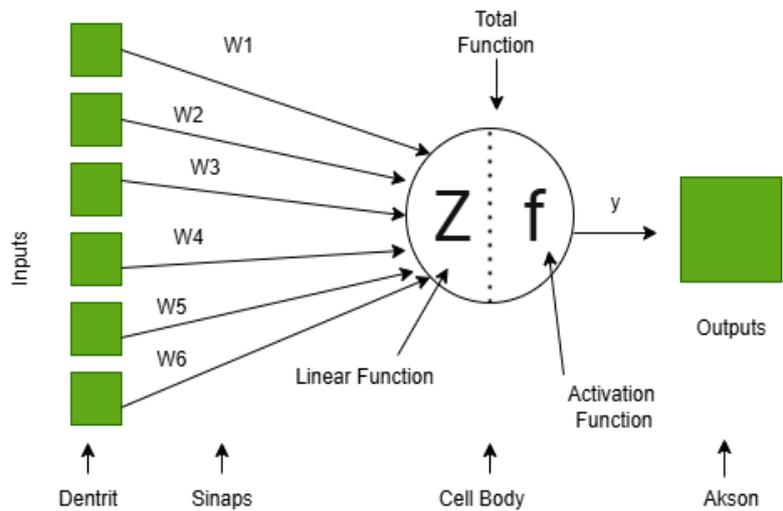
Figure 7: Artificial neuron

The situation in artificial neural networks is similar. Artificial neurons consist of information input, information output, the weight of each input, the summation function, and the activation function. Additionally, there is a value called bias for each neuron. The dendrites correspond to the inputs, the axon corresponds to the output, the synapse corresponds to the weight and bias values, and the cell body corresponds to the summation and activation functions. From this, we can conclude that learning in artificial neural networks occurs through the weights and bias values. Figure 7 shows the artificial neural network model.
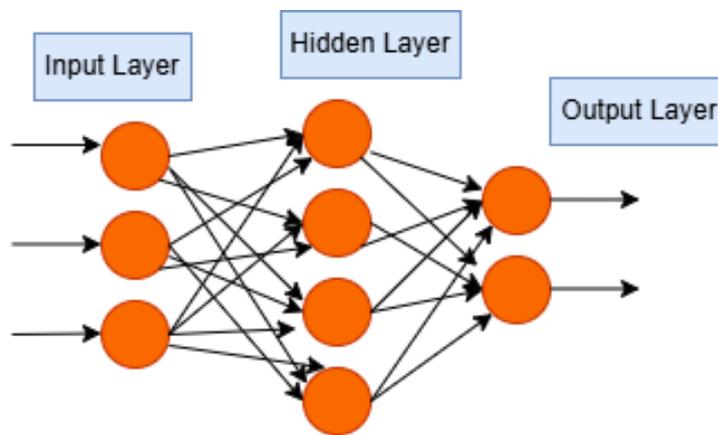
Figure 8: Multilayer artificial neural network

Artificial neural networks have an extraordinary ability to predict, classify, and match data. These networks are divided into different types, with the simplest and most commonly used being the multilayer perceptron. These networks understand the complex relationships between input data by following the patterns of the human brain and mathematical relationships to accurately predict the output with small errors. Figure 8 depicts a multilayer artificial neural network.

Convolutional Neural Networks

Convolutional neural networks (CNNs) are powerful deep learning algorithms capable of handling millions of parameters while saving computational costs by processing a 2-

dimensional image through convolution with filters/kernels and producing output volumes [10]. CNNs are deep learning algorithms that take input images and transform then using filters or kernels to extract features. An NxN image is convolved with an fxf filter, and this convolution operation learns the same feature across the entire image[11], Figure 9 shows the general structure of a convolutional neural network.
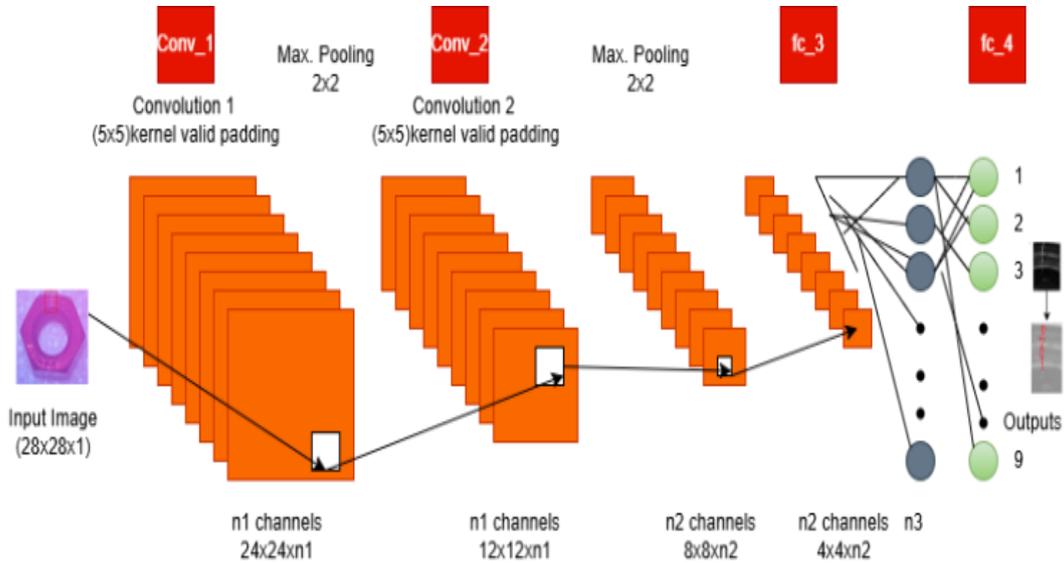
Figure 9: Convolutional neural network structure

The convolutional layer consists of a two-dimensional matrix and filter, and the filter will multiply the image matrix in the upper left image. For the 5×5 image with a filter size of 3×3, the convolved feature of the crack can be identified by multiplying and summing them up [12]. Then, slide the filter matrix over the image and compute the dot product to detect patterns at the stride of 1, specifically as shown in Figure 10.
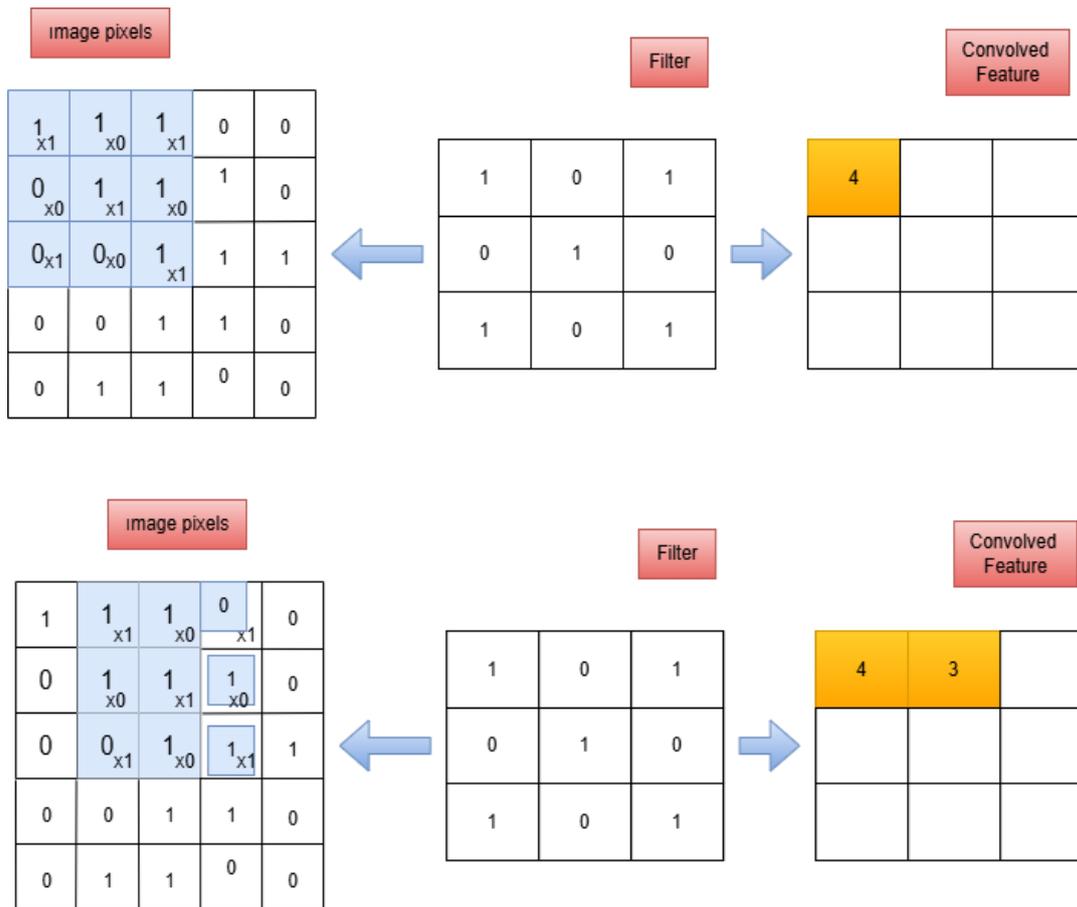
Figure 10: Convolutional example of image (5 × 5) with filter (3 × 3)

A pooling layer follows a convolutional layer to reduce the dimensions of the input data. Convolutional and pooling layers typically form pairs and work together. The convolutional feature map is large, which can increase computation costs and the likelihood of overfitting, making it essential to reduce its size. Achieving this involves analyzing a limited region in the input matrix, known as the pooling region, and extracting its dominant features. This extraction can occur in two ways: average pooling or max pooling. In average pooling, the average of all values in the pooling region is calculated, while in max pooling, the maximum value within the pooling region is considered. The resulting matrix from performing such an operation contains the main features of the input matrix in a smaller size, or in other words, it summarizes the features extracted by the convolutional layer. Figure 11 shows the pooling layer.
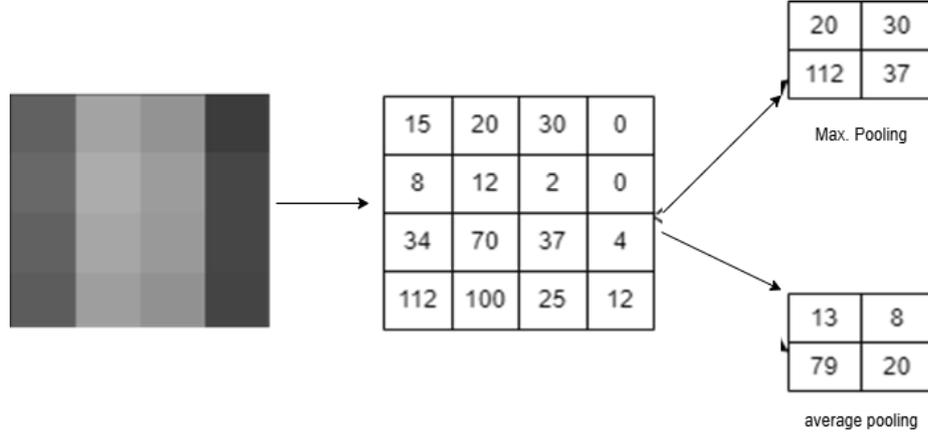
Figure 11: Pooling layer

EXPERIMENT RESULTS

A transfer learning approach with a pre-trained CNN model has been adopted to evaluate the effectiveness of detecting surface cracks in nut images obtained through magnetic particle testing. Firstly a crack image is cropped and resized to 28x28x1 pixels. Tehe dataset, including 500 images is divided into images including cracks (possitive samples) and images without crakcs (negative samples). In order to evaluate the accuracy of the crack detection, intersection over union (IoU), sensitivity (SEN), specificity (SPE), and Dice similarity coefficient (DSC) are used. Here, precision represents the probabilities of positive classification, which turns out to be corrected, as shown in Equation 8. SEN (Recall) indicates the probabilities of actual positive classified correctly, as shown in Equation 5. IoU is used to measure the degree of overlap between the predicted results and the real labels, which can provide a quantitative evaluation of the matching degree between the predicted results and the real labels, as shown in Equation 4. DSC (F1) is an overall indicator of a model's accuracy, which combines both precision and recall, as shown in Equation 7. Specificity (SPE) is a commonly used performance measure to evaluate the ability of binary classification models to recognize negative instances. The SPE value ranges from 0 to 1, with higher values indicating a stronger ability of the model to identify negative cases, as shown in Equation 6. These metrics can be calculated as follows:

$$IoU = \frac{TP}{TP + FP + FN} \tag{4}$$

$$SEN(Recall) = \frac{TP}{TP + FN} \tag{5}$$

$$SPE = \frac{TN}{TN + FP} \tag{6}$$

$$DSC(F1) = \frac{2TP}{2TP + FP + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Where TP denotes true positives, TN denotes true negatives, FP denotes false positives, and FN denotes false negatives.

The experiments were conducted using three datasets of varying sizes generated from images obtained after magnetic particle testing. These datasets consist of images with cracks, images without cracks, and test images. The experiments were carried out on a laptop system with 8 GB RAM and an 11th Gen Intel(R) Core(TM) i5-1135G7 CPU @ 2.40GHz, using Python programming.

Table 1: Experimental results of CNN model

CNN Model		
	Pozitif	**Negatif**
Test Accuracy	83.27 %	83.56 %
Test Loss	0.06512	0.06512
Precision	0.83	0.84
Recall	0.84	0.83
F1 Score	0.84	0.83

Our CNN model was trained over 20 epochs with three different varying datasets. As shown in Table 1, the model's values were determined based on test accuracy, test loss, precision, recall, and F1 score. Additionally, the changes in training accuracy, training loss, validation accuracy, and validation loss according to the number of epochs are presented in Table 2.

Table 2: 1st and 20th Epoch training accuracy/training loss/validation accuracy/validation loss

CNN Model		
	1st epoch	**20st epoch**
Training Accuracy	0.4128	0.8327
Training Loss	0.5482	0.0987
Validation Accuracy	0.6179	0.8389
Validation Loss	0.5940	0.0488

A normalized confusion matrix is shown in Figure 12, which interprets how the label is predicted. The value of diagonal elements represents the high degree of correctly predicted classes.
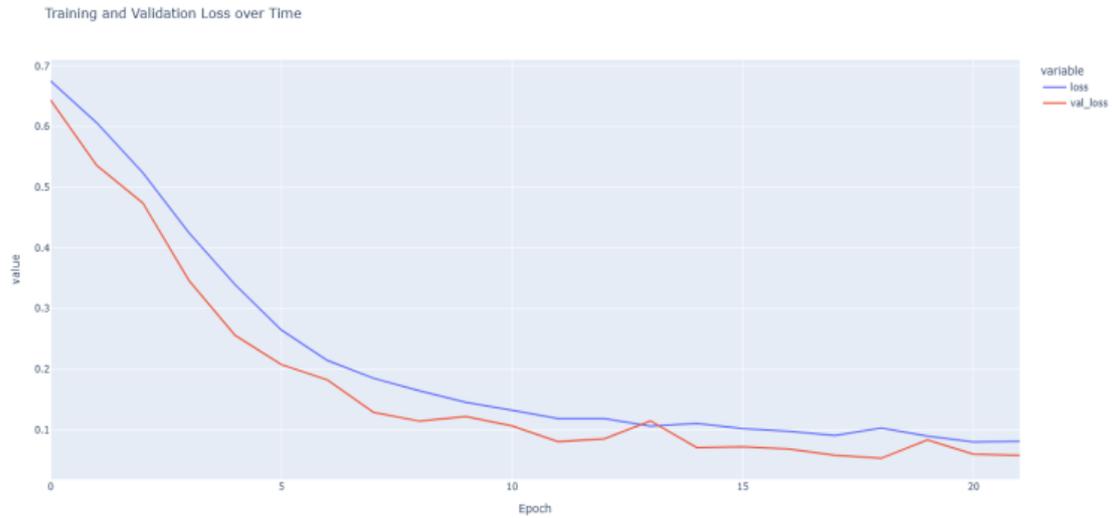
Figure 12: CNN's loss graph over time

Figure13. Normalized confusion matrix of basic CNN model. As shown in Figure 14, it indicates that the classifier correctly separates the crack and non-crack correctly 80% of the time. However, since the AUC is not 1, the model has type 1 and type 2 errors, where a crack is judged as a non-crack. The curve leans towards the top left corner, indicating the classifier is still good at identifying a crack.
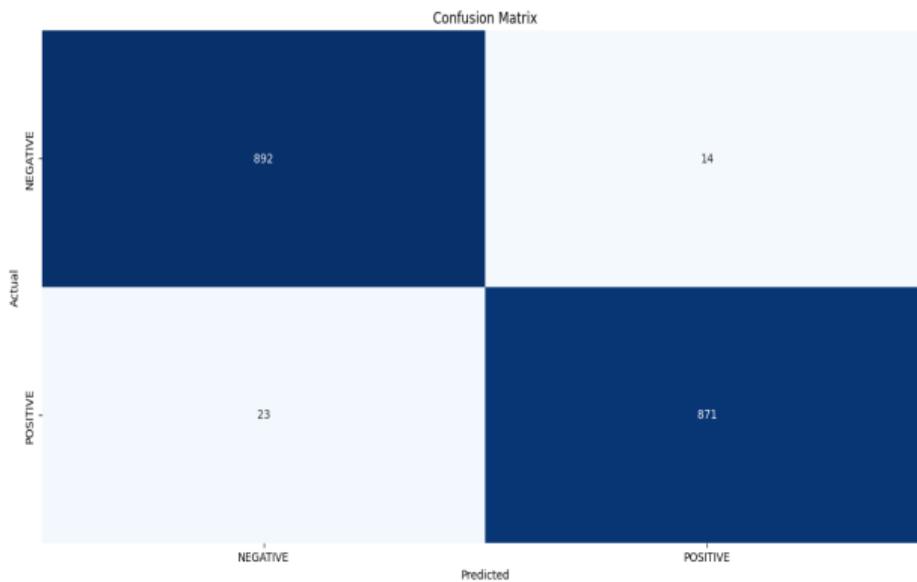
Figure 13: Confusion matrix of basic CNN model
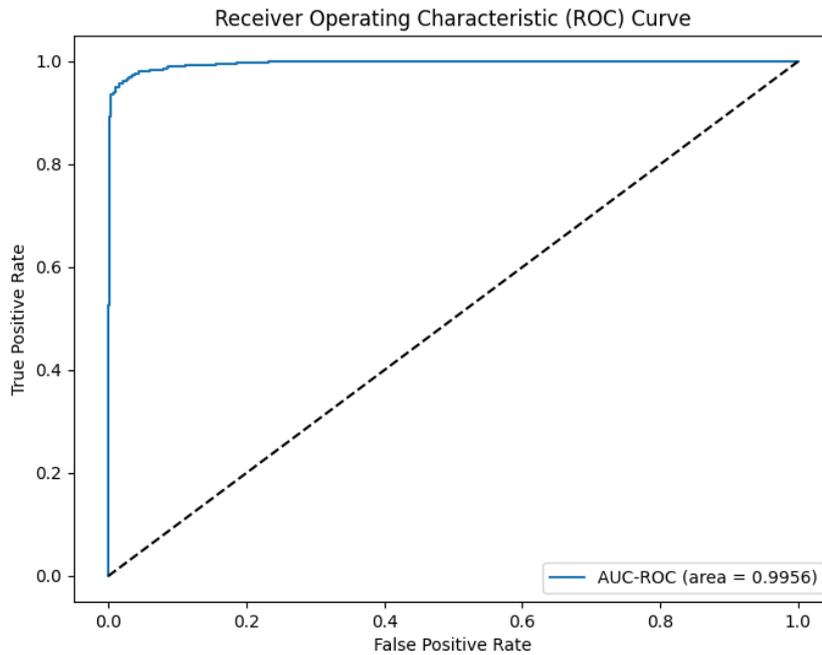
Figure 14: Result of model with basic CNN and model with transfer learning.

CONCLUSION

Due to natural aging and wear, the appearance of cracks in structural materials like nuts is a common issue. If left unaddressed, these cracks can worsen over time and potentially lead to safety hazards. Incorporating deep learning technology, especially for image classification, has shown promising results in detecting cracks in such materials. This study focuses on evaluating the effectiveness of deep learning technology for image classification, particularly CNNs. Although CNNs typically require significant amounts of data to achieve optimal performance, it is also possible to manually collect such images. This study addresses this predicament by employing a transfer learning approach, allowing for successful image classification even when data are limited. The CNN method yielded an accuracy level of approximately 83%.

The main contributions of this paper are summarized in two folds. First, this paper reviews the current defect detection practices for basic metal fasteners like nuts and examines the challenges faced by traditional methods, while exploring the feasibility of applying deep learning techniques for crack detection in this sector. Second, a CNN-based method is proposed for detecting cracks on the surfaces of metal nuts, with a detailed explanation of the implementation of these components.

REFERENCES

[1] Ramírez-Gil, F.J., E.C.N. Silva, and W. Montealegre-Rubio, Through-thickness perforated steel plates optimized for ballistic impact applications. Materials & Design, 2021. 212: p. 110257.

[2] Karthik, M.M., et al., Magnetic flux leakage technique to detect loss in metallic area in external post-tensioning systems. Engineering Structures, 2019. 201: p. 109765.

[3] Motlagh, N.H., T. Taleb, and O. Arouk, Low-altitude unmanned aerial vehicles-based internet of things services: Comprehensive survey and future perspectives. IEEE Internet of Things Journal, 2016. 3(6): p. 899-922.

[4] Valença, J., et al., Assessment of cracks on concrete bridges using image processing supported by laser scanning survey. Construction and Building Materials, 2017. 146: p. 668-678.

[5] Li, S., et al., Deep learning for hyperspectral image classification: An overview. IEEE Transactions on Geoscience and Remote Sensing, 2019. 57(9): p. 6690-6709.

[6] Kumar, S. and D. Mahto, Recent trends in industrial and other engineering applications of non destructive testing: a review. International Journal of Scientific & Engineering Research, 2013. 4(9).

[7] Gholizadeh, S., A review of non-destructive testing methods of composite materials. Procedia structural integrity, 2016. 1: p. 50-57.

[8] Leung, M.K., et al., Deep learning of the tissue-regulated splicing code. Bioinformatics, 2014. 30(12): p. i121-i129.

[9] Abraham, A., Meta learning evolutionary artificial neural networks. Neurocomputing, 2004. 56: p. 1-38.

[10] LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. nature, 2015. 521(7553): p. 436-444.

[11] Zeiler, M. Visualizing and Understanding Convolutional Networks. in European conference on computer vision/arXiv. 2014.

[12] Szegedy, C., et al. Rethinking the inception architecture for computer vision. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

RESUME

Yasin ULUS

Yasin ULUS graduated from the Department of Electrical and Electronics Engineering at Kırıkkale University in 2014. Since 2022, he has been pursuing a master's degree in Electrical and Electronics Engineering at Selçuk University. He has been serving in the Turkish Armed Forces since 2017. He is married and has one child.

Assoc. Prof. Dr. Mehmet Akif ŞAHMAN

Mehmet Akif Şahman, Ph.D. He graduated from Selçuk University, Department of Computer Systems Education in 2005. He received his master's degree from Selçuk University, Department of Electronics and Computer Systems Education in 2008. He completed his Ph.D. in Computer Engineering at Selçuk University in 2016. Since 2022, he has been working as an Associate Professor in the Department of Electrical and Electronics Engineering at Selçuk University. His current research interests include machine learning, artificial intelligence, imbalanced datasets, swarm intelligence, nature-inspired algorithms.

A REVIEW OF THE DEVELOPMENTS AND APPLICATIONS OF ARTIFICIAL HUMMINGBIRD ALGORITHM IN OPTIMIZATION

Esma ÖZEL[1], Onur İNAN[2]

INTRODUCTION

OPTIMIZATION algorithms have been pivotal in addressing complex computational and engineering problems, where the objective is to find optimal solutions within challenging constraints. Among the different bio-inspired algorithms that have been proposed in recent times, Artificial Hummingbird Algorithm has been recognized because of its unique simulation of foraging and flight behaviors of hummingbirds [1]. AHA has some flying patterns that are new and certain foraging strategies-territorial, guided, and migratory-which are novel ways of balancing exploration with exploitation [2].

The metaheuristic methods can avoid sticking in a local optimum point and reach the global solution, thus reducing expenses by efficiently searching through much larger databases more quickly [3]. Since its introduction, AHA has shown significant promise in surpassing conventional metaheuristics such as Genetic Algorithms (GA) and Simulated Annealing (SA) [4-6]. But despite its potential, AHA has drawbacks when it comes to solving high-dimensional issues, namely early convergence and limited exploration capacity [7, 8]. To get over these restrictions, recent developments such as hybrid algorithms and multi-strategy improvements have been suggested [7, 8].

Another research applies the chaotic hummingbird algorithm (CAHA) to a number of mechanical design optimization problems, including the construction of pressure vessels, gear trains, and welded beams. In comparison to conventional algorithms in the field of mechanical design, the study asserts that the employment of z maps such as Gauss/mouse improves performance, offers better optimization solutions, convergence speed, and robustness [10].

In a recent study, the authors created a probabilistic energy management technique to deal with pricing and load uncertainties, as well as the unpredictability of solar and wind energy outputs. To model these uncertainties, they combined the Point Estimation Method (PEM) with a hybrid Artificial Hummingbird Algorithm (AHA). In order to capture the unpredictability in power generation and demand and provide more accurate and dependable system performance, the study used normal and Weibull distribution functions [11].

This review seeks to answer a number of key questions about the growth, effectiveness, and prospects of AHA. Firstly, how does AHA compare to other metaheuristic algorithms in terms of performance and efficiency? This question seeks to contextualize AHA's competitive edge by examining its effectiveness in various benchmark problems. Secondly, what are the key

[1] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0000-0001-9806-3883, esmagungor@selcuk.edu.tr
[2] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0000-0003-4573-7025, oinan@selcuk.edu.tr

challenges and limitations of AHA, and how have recent enhancements addressed these? Understanding these limitations and the modifications that have been introduced, such as hybridization with Lévy flights and quantum-based optimization, will shed light on the trajectory of AHA's evolution [5, 12].

Furthermore, the review explores how AHA has been applied to real-world engineering challenges, including applications in truss topology optimization [5], robotics [13], and energy systems like solar modules [14]. These examples highlight the algorithm's versatility and the breadth of its impact. Finally, we discuss future directions for AHA development, considering the potential for even broader applicability and enhanced performance through further innovations in algorithmic design.

Artificial Hummingbird Algorithm

A new method in computational intelligence named as The Artificial Hummingbird Optimization Algorithm, modeled to mimic the adaptive foraging and flight patterns of hummingbirds (Figure 1). By leveraging these natural mechanisms, the algorithm provides effective solutions to complex optimization challenges in a wide range of fields [1].

Figure 1: Foraging hummingbird

AHA is a nature-inspired optimization algorithm that replicates the flight patterns and foraging habits of hummingbirds [1]. It has attracted considerable attention in recent scholarly works thanks to its ability to solve complex optimization challenges. Algorithm of AHA is organized around three distinct foraging strategies (Figure 2): migratory, territorial, and guided, along with three primary flight maneuvers (Figure 3): omnidirectional, diagonal, and axial. These elements allow the AHA to balance between exploration (wide search of the solution space) and exploitation (fine-tuning solutions), making it effective for various optimization tasks.
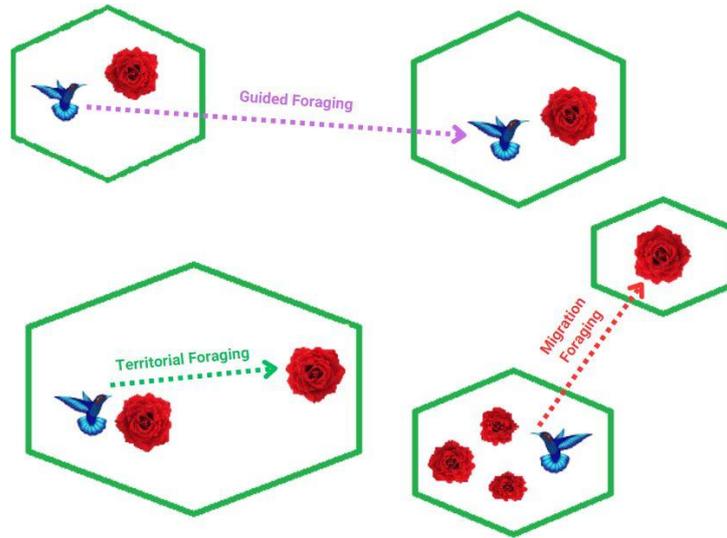
Figure 2: Foraging behaviors of hummingbirds

AHA also incorporates a visit table, simulating the memory of hummingbirds, which helps the algorithm avoid revisiting unpromising areas and accelerates convergence towards optimal solutions.
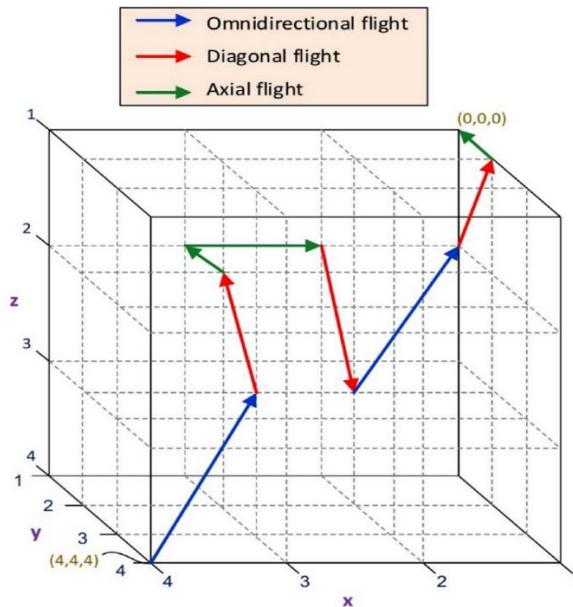
Figure 3: The movement of a hummingbird within a three-dimensional coordinate system [1].

Food Sources

Hummingbirds identify food sources by evaluating nectar quality, refill rate, content and the time since the last visit. In AHA. In AHA, all food sources are considered to have the same type and number of flowers, and the fitness function of a source is based on how quickly its nectar is replenished.

Hummingbirds

The precise location of a hummingbird and its food source remains constant since each hummingbird feeds from the same source. Hummingbirds retain the memory of the location of this food source, the rate at which the nectar is replenished, and share this information with other hummingbirds nearby. Each hummingbird is also able to remember the amount of time that has passed since its last visit to a particular food source.

Visit Table

The visit table records the visit level for each hummingbird at each food source, indicating the time since the hummingbird last visited that specific source. A hummingbird gives priority to visiting the most refresh rated food resource. Among sources with the same highest visit level, it typically chooses the one with the fastest nectar replenishment rate to maximize nectar intake. The visit table helps each hummingbird identify the specific food source it needs, and is frequently updated during each iteration.

In guided foraging each hummingbird searches the most nectar containing resource. Agent aims to find most recently visited resources for foraging. Once a new target is determined hummingbird flies toward it [1]. Algorithm 1 explains the strategy for guided foraging of hummingbirds. The foraging process has also three flight skills. If we think about it as a direction switch vector, it dictates availability of directions to follow. As can be seen in Figure 3 hummingbird agent can fly one of coordinate axis which is Axial flight or one rectangle corner to another which is diagonal fly and omnidirectional flight which is agent can fly along in any direction so all agents can uniquely excel.

For ith hummingbird from 1 to n

   Perform equation (6)

   If $f(v_i(t+1)) < f(x_i(t))$

     $x_i(t+1) = v_i(t+1)$

     For jth food source from 1 to n $(j \neq tar, i)$

       Visit_table(i,j) = Visit_table(i,j) + 1

     End

     Visit_table(i,tar) = 0

     For jth food source from 1 to n

       $\text{Visit\_table}(j,i) = \max_{l \in n \text{ and } l \neq j} (\text{Visit\_table}(j,l)) + 1$

     End

   Else

     For jth food source from 1 to n $(j \neq tar, i)$

       Visit_table(i,j) = Visit_table(i,j) + 1

     End

     Visit_table(i,tar) = 0

   End

End

Algorithm 1: Guided foraging strategy

Once a hummingbird has depleted the nectar at its target food source, it tends to seek out a new source rather than returning to other known ones. Thus, the hummingbird can quickly explore the neighboring area within its territory, where it may locate a new food source that could offer a better solution than the existing one. Algorithm 2 explains the strategy for territorial foraging of hummingbirds [1].

For ith hummingbird from 1 to n
   Perform equation (9)
   If $f(v_i(t+1)) < f(x_i(t))$
     $x_i(t+1) = v_i(t+1)$
     For jth food source from 1 to n $(j \neq i)$
       Visit_table(i,j)=Visit_table(i,j)+1
     End
     For jth food source from 1 to n
       Visit_table$(j,i) = \max\limits_{l \in n \text{ and } l \neq j} (\text{Visit\_table}(j,l)) + 1$
     End
   Else
     For jth food source from 1 to n $(j \neq i)$
       Visit_table(i,j)=Visit_table(i,j)+1
     End
   End
End

Algorithm 2: Territorial foraging strategy

The AHA algorithm introduces a migration coefficient to encourage hummingbirds to migrate to distant food sources when their current source is low on nectar. If the migration coefficient exceeds iterations, the bird stays at the new source [1].

If $\mod(t, 2n) == 0$
   Perform equation (11)
   For jth food source from 1 to n $(j \neq wor)$
     Visit_table(wor,j)=Visit_table(wor,j)+1
   End
   For jth food source from 1 to n
     Visit_table$(j,wor) = \max\limits_{l \in n \text{ and } l \neq j} (\text{Visit\_table}(j,l)) + 1$
   End
End

Algorithm 3: Migration foraging strategy

AHA initializes random solutions and visit table, with 50% chance of guided or territorial foraging. Guided foraging moves hummingbirds towards food sources, while territorial

foraging disrupts local neighborhoods every 2n iterations and migration occurs. All behaviors of foraging use omnidirectional, diagonal, and axial flights [1]. As can be seen in pseudocode of AHA (Algorithm 5) the process continues interactively until the stopping criterion is met, and the food source with the best nectar-refilling rate is returned as an approximation of the global optimum.

```
Input: n, d, f, Max_Iteration, Low, Up
Output: Globalminimum, Globalminimizer
Initialization:
    For ith hummingbird from 1 to n,
        Do xᵢ=Low+r(Up-Low),
        For jth food source from 1 to n, Do
            If i≠j
            Then Visit_tableᵢ,ⱼ=1,
            Else Visit_tableᵢ,ⱼ=null,
            End If
        End For
    End For
While t≤Max_Iteration Do
    For ith hummingbird from 1 to n, Do
        If rand≤0.5 Then
            If r<1/3 Then perform equation (3),
                Else If r>2/3
                    Then perform equation (4),
                    Else perform equation (5),
                End If
        End If
        Perform equation (6),
        If f(vᵢ(t+1))<f(xᵢ(t))
        Then xᵢ(t+1)=vᵢ(t+1),
            For jth food source from 1 to n (j≠tar,i), Do
            Visit_table(i,j) =Visit_table(i,j) +1,
            End For
                Visit_table(i,tar)=0,
            For jth food source from 1 to n, Do
        Visit_table(j,i)= max_{l∈n and l≠j} (Visit_table(j,l)) +1,

            End For
        Else
            For jth food source from 1 to n (j≠tar,i), Do
```

```
            Visit_table(i,j)=Visit_table(i,j)+1,
            End For
                Visit_table(i,tar)=0,
        End
    Else
        Perform equation (9),
        If f(vᵢ(t+1))<f(xᵢ(t))
        Then xᵢ(t+1)= vᵢ(t+1),
            For jth food source from 1 to n (j≠i), Do
            Visit_table(i,j)=Visit_table(i,j)+1,
            End For
            For jth food source from 1 to n, Do
        Visit_table(j,i) = max_{l∈n and l≠j} (Visit_table(j,l)) +1,

            End For
        Else
            For jth food source from 1 to n (j≠i), Do
            Visit_table(i,j)=Visit_table(i,j)+1,
            End For
        End If
    End If
End For
If mod(t,2n)==0,
Then perform equation (11),
    For jth food source from 1 to n (j≠wor), Do
    Visit_table(wor,j)=Visit_table(wor,j)+1,
    End For
    For jth food source from 1 to n, Do
Visit_table(j,wor) = max_{l∈n and l≠j} (Visit_table(j,l)) +1,

    End For
End If
End While
```

Algorithm 4: Pseudo-code of Algorithm

Research Methodology

A survey of current studies on AHA applications is carried out using a review protocol. Following this protocol, the study addresses specific research questions and identifies the primary application areas of the algorithm, and discusses hybridized versions and the advantages or disadvantages of them. Figure 4 represents the review methodology used in this paper.
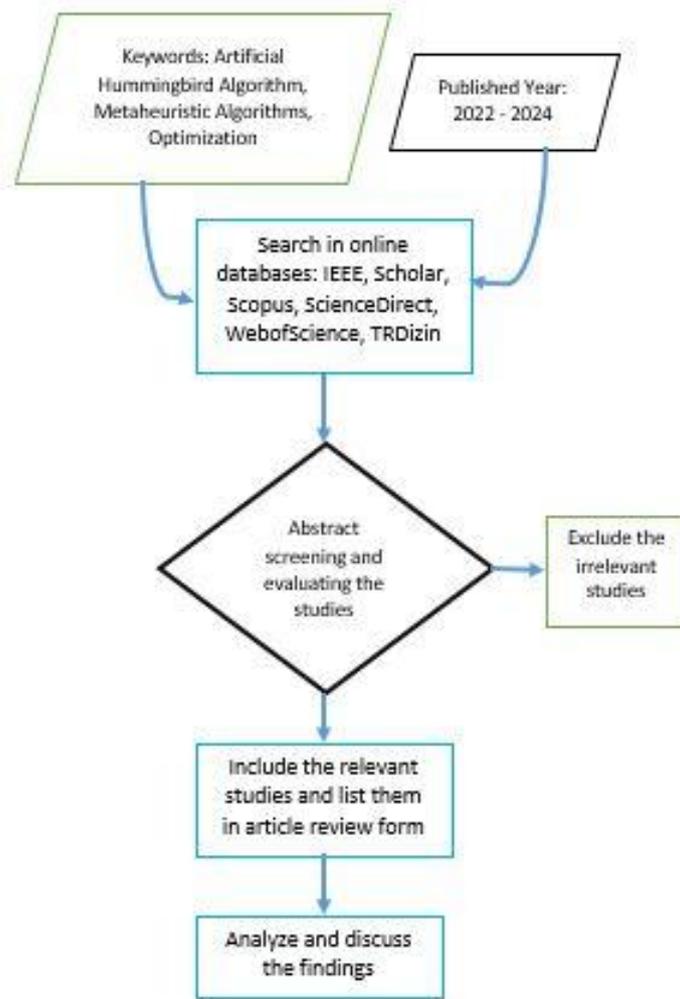
Figure 4: Review Methodology

The Research Questions

In order to achieve a successful review study, determining the research questions before searching is a necessity. This study addressed the following questions:

Q1. How does AHA compare to other metaheuristic algorithms in terms of performance and efficiency?

Q2. What are the key challenges and limitations of AHA, and how have recent enhancements addressed these?

Q3. How AHA has been applied to real-world engineering challenges?

The Research Strategy

IEEE Explore, TRDizin, Google Scholar, Science Direct, Scopus and Web of Science indexing engines were searched to determine studies about AHA. Most recent studies were reached by specifying last two years as date range. In literature search process "metaheuristic algorithms", "optimization", "artificial hummingbird algorithm", "hybrid algorithms" keywords and their

different combinations were used respectively and the results were filtered by fields and repetitive studies eliminated.

Inclusion/Exclusion Criteria

A novel article review form was developed by authors and the studies that searched before have noted on it. The relevant details such as field, year, authors etc. were also recorded in article review form. The criteria of inclusion or exclusion determined as in Figure 5.
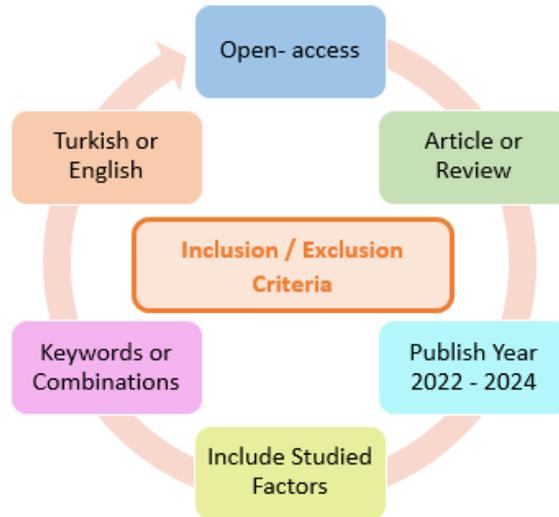
Figure 5: Inclusion / Exclusion Criteria

Studies that do not include any of the keywords, that are repetitive studies, that do not specify the field of research, that not in year range, which are not full-text were left out from the research. And review include 27 articles given in Table 1.

Table 1: AHA studies included in the research.

#	Main Focus Area	Study
1	Algorithm Development and Engineering Optimization	[1, 7, 10, 12, 15-18]
2	Energy Systems and Power Optimization	[11, 14, 19-23]
3	Medical Imaging and Disease Detection	[9, 24-26]
4	Control Systems	[13, 27]
5	Agriculture and Plant Disease Detection	[28]
6	Textile Industry	[29]
7	IoT and Social Systems	[8]
8	Comparative and Review Studies	[2, 4, 30]

RESULTS AND DISCUSSION

In this study, one of the questions stated earlier, "How does the AHA perform across different optimization tasks in various fields?" can be answered as follows. AHA has demonstrated as an effective optimization tool in a wide range of fields (Figure5) such as energy systems [11, 19, 22], healthcare [9, 24, 25], machine learning [28, 29], agriculture [28], control systems [27]. Across all these fields, AHA consistently performs well in solving complex, multi-objective optimization tasks, particularly excelling in environments characterized by uncertainty and variability. It particularly excels in domains where uncertainty and variability are high. Its flexibility and robustness enable it to outperform traditional algorithms like Particle Swarm Optimization (PSO) and Grey Wolf Optimizer (GWO). AHA shows exceptional capability in optimizing systems where traditional methods struggle, such as handling real-time data fluctuations and multi-parameter optimization. Additionally, its chaotic and hybrid versions have pushed AHA's potential even further, making it a suitable choice for highly dynamic and uncertain environments like renewable energy and healthcare diagnostics.
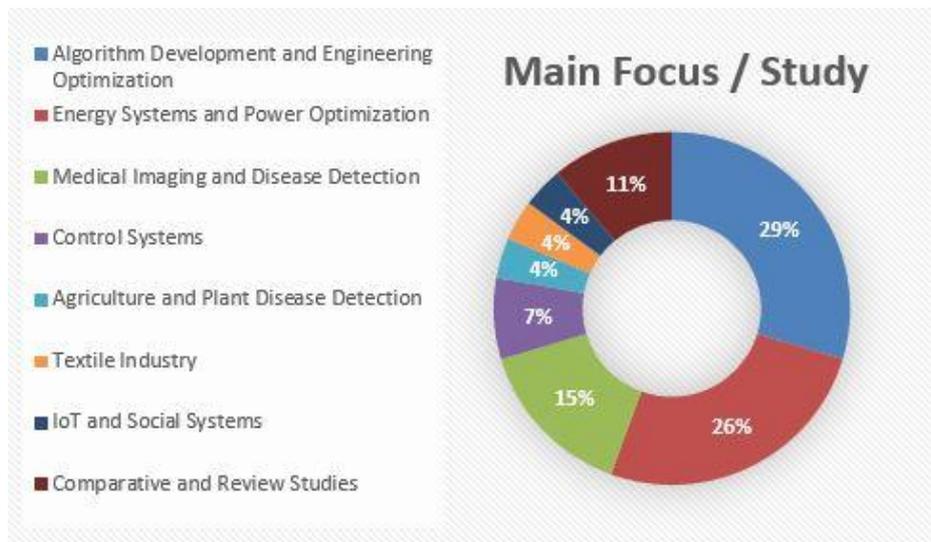
Figure 5: Main focus areas of studies

It has seen that AHA is flexible, easy to implement and well-suited for many types of optimization problems. Its memory-based learning and hybridization capabilities enhance its adaptability. But on the other hand it may face challenges with premature convergence, computational inefficiency in large-scale problems, and balancing exploration and exploitation. Advantages and disadvantages of AHA are explained in Table 2.

Table 2: Advantages and disadvantages of AHA.

Aspect	Advantages	Disadvantages
Exploration and Exploitation	Balances exploration (global search) and exploitation (local refinement) through distinct foraging behaviors	Can face challenges in maintaining a balance between exploration and exploitation when dealing with complex problems.

Flight Maneuvers	Utilizes axial, diagonal, and omnidirectional flight maneuvers, offering flexibility in navigating the search space	Flight maneuvers can sometimes lead to premature convergence if not well-optimized
Foraging Strategies	Incorporates guided, territorial, and migratory foraging, which helps prevent premature convergence	In real-world problems, fine-tuning the foraging strategies can be computationally expensive
Memory-Based Learning	The use of the Visit Table allows memory-based optimization, enhancing learning from previous solutions	Memory requirements increase with larger problem spaces, which can slow down the algorithm
Application Versatility	Demonstrates success in a wide range of applications (engineering, robotics, energy systems, feature selection)	Needs additional modifications (e.g., hybridization) to be competitive in highly complex or dynamic real-world problems
Algorithm Simplicity	Relatively simple structure and easy to implement	Lacks the same maturity and extensive tuning as older metaheuristic algorithms like PSO /GA
Hybridization Potential	Easily hybridized with other techniques (e.g., Lévy flight) to enhance performance	Hybrid versions often introduce additional parameters, making the algorithm more complex to tune
Computational Efficiency	Effective at solving medium-scale optimization problems, with relatively fast convergence in these scenarios	Computational performance may degrade with large-scale, high-dimensional optimization problems

The researches that hybridized AHA with other optimization algorithms like PSOAHA [15], BFAHA [25], CAHA [10], AAHA [19], MAHA [9] consistently show improvements in convergence speed, accuracy, and the ability to escape local minima, making them more suitable for complex, real-world problems. They also demonstrate enhanced performance in multi-objective optimization tasks, handling large-scale datasets, and improving the interpretability of models in domains like healthcare and energy management. In versatility range AHA's ability to adapt to different fields is the solid fact of its general-purpose nature. AHA has been applied not only in static optimization problems but also in dynamic, real-time optimization environments like energy systems and control systems. In comparison with other metaheuristic algorithms, AHA is fast in convergence, superior at handling local minima, and highly efficient in multi-objective optimization tasks. Medical diagnostics and renewable energy systems are two fields that benefit greatly from it because it provides high adaptability and accuracy under uncertain conditions.

CONCLUSION AND FUTURE WORK

AHA has also been shown to be able to handle a wide range of problems by its several hybrid versions via balancing the principles of exploration and exploitation. Besides multi-objective optimization capability and fast convergence, it is an invaluable tool that could be used by researchers and practitioners across a wide range of industries. Although AHA has performed well in comparison with other metaheuristic algorithms, some areas should also be further explored. In fact, as supported by existing literature, the potential of its applications has continued to expand, with positive results already obtained from various domains; however, further analysis is required to tap into it, especially in real-time and high-dimensional

optimization. In the near future, one can also try to apply and test AHA on real-world domains such as machine learning, Educational Data Mining, and industrial optimization problems. Integrating AHA into the pipeline of machine learning at points requiring feature selection, tuning parameters, or model optimization would be something new. Thus, this would enhance the performance of ML algorithms by equipping them with stronger optimization capabilities. Hybrid approaches based on combinations of AHA with other metaheuristics, such as Ant Colony Optimization or Differential Evolution, may provide an improved optimization performance for higher-dimensionality applications while leveraging the strengths of multiple algorithms.

REFERENCES

[1] W. Zhao, L. Wang, and S. Mirjalili, "Artificial hummingbird algorithm: A new bio-inspired optimizer with its engineering applications," *Computer Methods in Applied Mechanics and Engineering,* vol. 388, p. 114194, 2022/01/01/ 2022, doi: https://doi.org/10.1016/j.cma.2021.114194.

[2] M. Hosseinzadeh *et al.*, "A Survey of Artificial Hummingbird Algorithm and Its Variants: Statistical Analysis, Performance Evaluation, and Structural Reviewing," *Archives of Computational Methods in Engineering,* pp. 1-42, 2024.

[3] M. S. Uzer and O. Inan, "A novel feature selection using binary hybrid improved whale optimization algorithm," *The Journal of Supercomputing,* vol. 79, no. 9, pp. 10020-10045, 2023.

[4] F. BELLİ and H. BİNGÖL, "Performance Comparison of Biology based Metaheuristics Optimization Algorithms using Unimodal and Multimodal Benchmark Functions," *Turkish Journal of Science and Technology,* vol. 18, no. 1, pp. 157-167, 2023.

[5] J. Wang, Y. Li, G. Hu, and M. Yang, "An enhanced artificial hummingbird algorithm and its application in truss topology engineering optimization," *Advanced Engineering Informatics,* vol. 54, p. 101761, 2022/10/01/ 2022, doi: https://doi.org/10.1016/j.aei.2022.101761.

[6] M. Karakoyun, O. Inan, and İ. Akto, "Grey Wolf Optimizer (GWO) Algorithm to Solve the Partitional Clustering Problem," *International Journal of Intelligent Systems and Applications in Engineering,* vol. 7, no. 4, pp. 201-206, 12/12 2019, doi: 10.18201/ijisae.2019457231.

[7] G. Hu, J. Zhong, C. Zhao, G. Wei, and C.-T. Chang, "LCAHA: A hybrid artificial hummingbird algorithm with multi-strategy for engineering applications," *Computer Methods in Applied Mechanics and Engineering,* vol. 415, p. 116238, 2023/10/01/ 2023, doi: https://doi.org/10.1016/j.cma.2023.116238.

[8] M. A. Elaziz, A. Dahou, M. A. Al-Betar, S. El-Sappagh, D. Oliva, and A. O. Aseeri, "Quantum Artificial Hummingbird Algorithm for Feature Selection of Social IoT," *IEEE Access,* vol. 11, pp. 66257-66278, 2023, doi: 10.1109/ACCESS.2023.3290895.

[9] E. Kıymaç and Y. Kaya, "A novel automated CNN arrhythmia classifier with memory-enhanced artificial hummingbird algorithm," *Expert Systems with Applications,* vol. 213, p. 119162, 2023/03/01/ 2023, doi: https://doi.org/10.1016/j.eswa.2022.119162.

[10] V. Bhattacharjee, P. K. Roy, and C. Chattoraj, "Chaotic artificial hummingbird algorithm applied to elementary machine design problems," *Progress in Artificial Intelligence,* pp. 1-27, 2024.

[11] N. Alamir, S. Kamel, T. F. Megahed, M. Hori, and S. M. Abdelkader, "Developing an artificial hummingbird algorithm for probabilistic energy management of microgrids considering demand response," *Frontiers in Energy Research,* vol. 10, p. 905788, 2022.

[12] L. Wang, L. Zhang, W. Zhao, and X. Liu, "Parameter Identification of a Governing System in a Pumped Storage Unit Based on an Improved Artificial Hummingbird Algorithm," *Energies,* vol. 15, no. 19, p. 6966, 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/19/6966.

[13] R. Yang, Q. Zhou, Q. Mo, L. Gan, and R. Hu, "Using an Improved Artificial Hummingbird Algorithm for Vision-Guided Optimization And Grasping of Multi-Objective Robots," *Journal of Physics: Conference Series,* vol. 2365, no. 1, p. 012052, 2022/11/01 2022, doi: 10.1088/1742-6596/2365/1/012052.

[14] S. Haddad, B. Lekouaghet, M. Benghanem, A. Soukkou, and A. Rabhi, "Parameter Estimation of Solar Modules Operating Under Outdoor Operational Conditions Using Artificial Hummingbird Algorithm," *IEEE Access,* vol. 10, pp. 51299-51314, 2022, doi: 10.1109/ACCESS.2022.3174222.

[15] K. Chen, L. Chen, and G. Hu, "PSO-Incorporated Hybrid Artificial Hummingbird Algorithm with Elite Opposition-Based Learning and Cauchy Mutation: A Case Study of Shape Optimization

for CSGC–Ball Curves," *Biomimetics,* vol. 8, no. 4, p. 377, 2023. [Online]. Available: https://www.mdpi.com/2313-7673/8/4/377.

[16]   W. Zhao, Z. Zhang, S. Mirjalili, L. Wang, N. Khodadadi, and S. M. Mirjalili, "An effective multi-objective artificial hummingbird algorithm with dynamic elimination-based crowding distance for solving engineering design problems," *Computer Methods in Applied Mechanics and Engineering,* vol. 398, p. 115223, 2022/08/01/ 2022, doi: https://doi.org/10.1016/j.cma.2022.115223.

[17]   B. A. M. Yahya A. Alhamdany, "Use Two Kind Hybridization of the Chaotic Peafowl Algorithm with the Hummingbird Algorithm," *Journal of Al-Qadisiyah for Computer Science and Mathematics,* vol. 14, no. 4, pp. 151-161, 2022.

[18]   H. Bakır, "A novel artificial hummingbird algorithm improved by natural survivor method," *Neural Computing and Applications,* pp. 1-25, 2024.

[19]   V. Kansal and J. Dhillon, "Ameliorated artificial hummingbird algorithm for coordinated wind-solar-thermal generation scheduling problem in multiobjective framework," *Applied Energy,* vol. 326, p. 120031, 2022.

[20]   U. Waleed, A. Haseeb, M. M. Ashraf, F. Siddiq, M. Rafiq, and M. Shafique, "A Multiobjective Artificial-Hummingbird-Algorithm-Based Framework for Optimal Reactive Power Dispatch Considering Renewable Energy Sources," *Energies,* vol. 15, no. 23, p. 9250, 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/23/9250.

[21]   A. Ramadan, M. Ebeed, S. Kamel, E. M. Ahmed, and M. Tostado-Véliz, "Optimal allocation of renewable DGs using artificial hummingbird algorithm under uncertainty conditions," *Ain Shams Engineering Journal,* vol. 14, no. 2, p. 101872, 2023.

[22]   M. S. Abid, H. J. Apon, K. A. Morshed, and A. Ahmed, "Optimal planning of multiple renewable energy-integrated distribution system with uncertainties using artificial hummingbird algorithm," *IEEE Access,* vol. 10, pp. 40716-40730, 2022.

[23]   A. M. Sadoun, I. R. Najjar, G. S. Alsoruji, M. S. Abd-Elwahed, M. A. Elaziz, and A. Fathy, "Utilization of Improved Machine Learning Method Based on Artificial Hummingbird Algorithm to Predict the Tribological Behavior of Cu-Al2O3 Nanocomposites Synthesized by In Situ Method," *Mathematics,* vol. 10, no. 8, p. 1266, 2022. [Online]. Available: https://www.mdpi.com/2227-7390/10/8/1266.

[24]   A. A. Malibari *et al.*, "Artificial Hummingbird Algorithm with Transfer-Learning-Based Mitotic Nuclei Classification on Histopathologic Breast Cancer Images," *Bioengineering,* vol. 10, no. 1, p. 87, 2023. [Online]. Available: https://www.mdpi.com/2306-5354/10/1/87.

[25]   L. Feng, Y. Zhou, and Q. Luo, "Binary hybrid artificial hummingbird with flower pollination algorithm for feature selection in parkinson's disease diagnosis," *Journal of Bionic Engineering,* vol. 21, no. 2, pp. 1003-1021, 2024.

[26]   M. A. Elaziz, A. Dahou, S. El-Sappagh, A. Mabrouk, and M. M. Gaber, "AHA-AO: Artificial Hummingbird Algorithm with Aquila Optimization for Efficient Feature Selection in Medical Image Classification," *Applied Sciences,* vol. 12, no. 19, p. 9710, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/19/9710.

[27]   L. Abualigah, S. Ekinci, D. Izci, and R. A. Zitar, "Modified Elite Opposition-Based Artificial Hummingbird Algorithm for Designing FOPID Controlled Cruise Control System," *Intelligent Automation & Soft Computing,* vol. 38, no. 2, 2023.

[28]   S. P. Sankareshwaran, G. Jayaraman, P. Muthukumar, and A. Krishnan, "Optimizing rice plant disease detection with crossover boosted artificial hummingbird algorithm based AX-RetinaNet," *Environmental Monitoring and Assessment,* vol. 195, no. 9, p. 1070, 2023/08/23 2023, doi: 10.1007/s10661-023-11612-z.

[29]   Z. Zhou, Y. Hu, Z. Zhu, and Y. Wang, "Fabric Wrinkle Objective Evaluation Model with Random Vector Function Link Based on Optimized Artificial Hummingbird Algorithm," *Journal of Natural Fibers,* vol. 20, no. 1, p. 2163026, 2023/04/24 2023, doi: 10.1080/15440478.2022.2163026.

[30]   A. Thongsamai, S. Chansombat, and S. Sooncharoen, "The applications of Artificial Hummingbird Algorithm (AHA) in the optimization problems: A review of the state-of-the-art," *Engineering & Applied Science Research,* vol. 51, no. 2, 2024.

RESUME

Lecturer Esma ÖZEL

Esma ÖZEL graduated from Süleyman Demirel University, Computer and Control Teaching Department in 2011. She started working as an academician in the Selçuk University's IT department in 2014 and continues to do so. She continues to practice academic studies on data mining during her master's degree.

Assist. Prof. Dr. Onur İNAN

Onur İNAN is an Assistant Professor in the Technology Faculty at Selçuk University. He completed both his master's and Ph.D. at Selçuk University. He has published articles on various optimization algorithms and data mining. His primary research interests include machine learning, artificial intelligence, and neural networks.

- 129 -

TIME, FREQUENCY, AND TIME-FREQUENCY FEATURE SETS FOR DEEP LEARNING METHODS IN EPILEPTIC SEIZURE PREDICTION

Atakan DAŞDEMİR[1], Humar Kahramanlı ÖRNEK[2]

INTRODUCTION

TODAY, epilepsy is treated with methods such as medication and surgical procedures that remove the part of the brain causing seizures. However, surgery is only an option for a small number of patients. Approximately 30%-40% of epilepsy patients cannot be treated with medication because they have a drug-resistant form of epilepsy [1]. The global prevalence of epilepsy is estimated to be around 1%, and a significant proportion of these patients are classified as untreatable. The prediction of uncontrolled epileptic seizures has been a focus of research for many years due to the potential risks associated with such seizures. n epileptic seizure prediction research, most researchers have utilized publicly available EEG datasets, with the CHB-MIT dataset being particularly prominent. Processing raw EEG data, including data from the CHB-MIT dataset, without any form of preprocessing often results in negative impacts on predictive performance. It is well established that feature extraction has a direct and positive influence on classification outcomes. Feature extraction involves isolating key characteristics from recorded EEG data to generate a feature vector. This approach not only highlights the most critical features for the classifier but also reduces the dimensionality of the feature vector [2]. When reviewing studies on epileptic seizure prediction, it becomes evident that signal analysis methods continue to play a significant role in deep learning-based approaches, as they did in traditional methods in the past.

SIGNAL ANALYSIS METHODS

Given their positive impact on classification outcomes, signal analysis methods play a crucial role in epileptic seizure prediction studies. These methods can be categorized into three main groups: time domain, frequency domain, and time-frequency domain approaches.

Time Domain

Time domain analysis involves examining the EEG signal as a time series. It focuses on applying statistical processes adapted to the characteristics of the signals to derive a set of key parameters and subsequently analyzing these parameters [2]. Within the time domain, signal analysis is typically conducted using statistical calculations. By performing these operations, parameters such as maximum, minimum, mean, standard deviation, entropy, and average power values are determined. This approach is particularly suitable for sudden signal fluctuations, such as those seen in epilepsy, and was the earliest method employed for seizure prediction.

---

[1] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0000-0002-8794-3559, ataknet@gmail.com
[2] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0000-0003-2336-7924, hkahramanli@selcuk.edu.tr

The time domain consists of both linear and nonlinear methods. Linear methods are generally characterized by fundamental statistical measures like mean and variance, offering low computational costs. In contrast, nonlinear methods include approaches such as chaotic and entropy-based techniques. Until the 1990s, linear methods were predominant in studies. However, with increased recognition of signal irregularities, calculations involving entropy were deemed more appropriate. Ahmadi and et al. with the calculation of the power spectral density and adopting its frequency components as probability density functions which are used in the calculation of Shannon entropy. In the next step, 2 classifiers which are support vector machine (SVM) and K-nearest neighbor (KNN) classifier were used as predictors of epileptic seizures. Their proposed algorithm can predict occurrence of a seizure using the first 9 minutes of a 10-minute interval before the seizure. The proposed method using SVM achieved sensitivity of 83.8% and specificity of 71%. KNN classifier achieved sensitivity of 83.8% and specificity of 67.8% [3].

Frequency Domain

Frequency domain analysis involves the detailed examination of information by breaking down a signal into its frequency (sinusoidal) components through statistical methods and Fourier Transformation (FT). Unlike time domain analysis, it focuses on the spectral components of a signal without providing any information about time. Non-stationary EEG signals typically comprise events occurring at different frequencies, with frequency representing the rate at which these events occur. The frequency domain, also known as spectral analysis, is more sensitive to the presence of artifacts compared to statistical methods within the time domain. It is frequently used during the feature extraction phase in traditional signal analysis methods.

One of the most commonly utilized frequency domain features is the Power Spectral Density (PSD). PSD illustrates how the energy or power of a signal is distributed across various frequency components. For example, increased power in the lower frequency bands of an EEG signal can reflect different brain activities, whereas shifts in higher frequency components may signify distinct neurological events.

Studies conducted by [4], [5] and [6] utilized Spectral Power Analysis for feature extraction. In their approach, the EEG data was segmented into windows, and the spectral power characteristics were calculated for each window. These researchers highlighted that this method is computationally efficient, making it suitable for portable warning systems. Moreover, they found that Spectral Power Analysis can yield promising results for seizure prediction, particularly in high gamma frequency bands.

Time-Frequency Domain

Time-frequency domain analysis is a method that characterizes the energy distribution of EEG and other signals across both time and frequency, decomposing the signal into a two-dimensional representation, with one axis representing time and the other representing frequency. Analyzing a signal solely in the time domain can make it challenging to extract information about how energy is distributed across various frequencies and how these frequencies change over time. Similarly, frequency domain analysis, which disregards the time aspect by focusing solely on frequency and assuming that the EEG signal spectrum remains

constant, can obscure the timing of frequency component changes, leaving certain characteristic details undetected.

The Short-Time Fourier Transform (STFT) is a widely used and relatively simple method for time-frequency analysis. It converts a one-dimensional EEG time series signal into a two-dimensional matrix representing both time and frequency, commonly referred to as a spectrogram. This approach utilizes a sliding window technique, applying traditional spectral estimation methods (such as the Fourier Transform, Welch's method, or the Autoregressive (AR) method) to segments of a non-stationary signal, dividing the data into smaller, assumed stationary fragments. The segmentation is accomplished using a window function that determines the size of each segment. As the window function shifts along the signal's measurement period, the Fourier Transform is applied to the data within each window, capturing the signal's time-varying frequency characteristics and producing a two-dimensional time-frequency function. In essence, the signal is broken into multiple short-term segments that are treated as stationary, with each segment undergoing spectral estimation through windowing. The window function then shifts along the time axis to continue the transformation [7], [8], [9], [10]. Troung and colleagues [11] applied STFT to transform the Freiburg and CHB-MIT datasets, achieving a sensitivity of 89.8% and a false positive rate (FPR) of 0.17 using CNN classification.

METHODOLOGY

Upon reviewing the studies for the prediction of epilepsy attacks, it was observed that the process shown in Figure 1 was commonly applied.
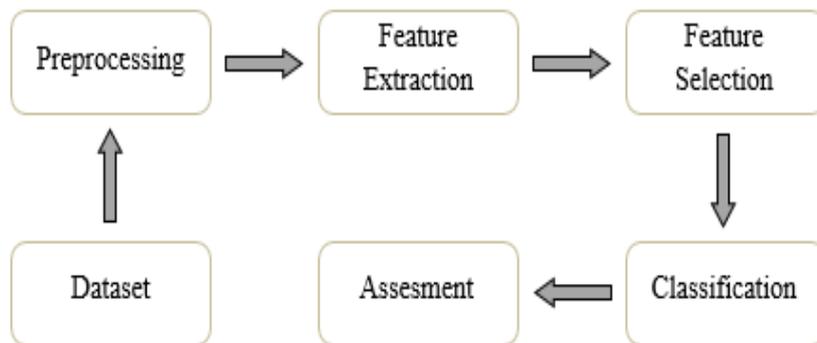
Figure 1: A common process in predicting epileptic seizures.

This study will not use the "Feature Selection" step. The process steps to be applied are as follows:

Dataset and Preprocessing

In this study, the CHB-MIT dataset, which is publicly available on PhysioNet.org, was utilized. This dataset comprises recordings from 23 patients and was collected using a 10-20 EEG recording system with 22 electrodes, sampled at a frequency of 256 Hz in a one-dimensional format. Certain patient data include issues such as missing or corrupted channels and recording

time inconsistency. The EEG data of five randomly selected patients were processed individually using the Pyedflib library.

In the preprocessing stage, raw EEG data were cleaned by removing artifacts and organizing the data. High-amplitude signals are typically classified as artifacts. The primary aim of the preprocessing stage is to reduce noise by detecting and eliminating these artifacts, thereby enhancing the accuracy of the classification process. To achieve this, frequencies within the ranges of 57-63 Hz and 117-123 Hz were filtered using a Butterworth band-stop filter to remove power line noise. Additionally, a high-pass filter was applied to eliminate frequencies below 1 Hz, reducing low-frequency noise and fluctuations. EEG signals were then segmented into 5-second windows for analysis. As depicted in Figure 2, EEG data were divided into three categories: interictal, preictal, and real-preictal periods. The interictal period represents normal brain activity between seizures and includes EEG recordings at least four hours away from seizure events; it was labeled as "Label 0" during the evaluation stage. The preictal period covers the 30-minute interval preceding seizure onset, excluding the 3-minute period immediately before the seizure, known as the Seizure Prediction Horizon (SPH), and was labeled as "Label 1." The real-preictal period, defined as the SPH, is utilized to assess the model's effectiveness in real-time applications.
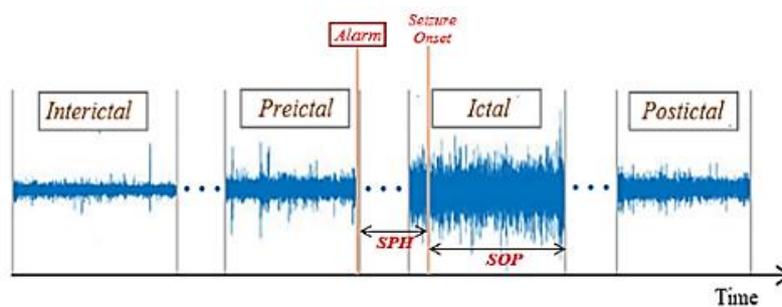
Figure 2: EEG Dataset periods.

The filtered and segmented 5-second EEG data were prepared for feature extraction.

Feature Extraction

The feature extraction phase involves applying various methods to the filtered EEG data. Traditional approaches typically rely on signal analysis techniques, although some studies utilize deep learning methods for automatic feature extraction. Feature extraction entails identifying key characteristics from recorded EEG data and creating a feature vector. In this study, feature extraction was performed using specific methods for three different domains. These methods include:

1. Within the scope of the time domain, feature extraction was performed using Shannon Entropy. The time domain enables the analysis of EEG signals across temporal windows, facilitating the extraction of more meaningful information by isolating specific frequency ranges. In this study, the EEG signal was processed within a designated window size to create filtered time series data for each window. This approach allows for the analysis of changes within each time interval while preserving the temporal resolution of the signal. Shannon Entropy was used to calculate the entropy of the filtered data separately for each channel, providing insights into the complexity and irregularity of the signal in each

window. This information can be critical for detecting neurological events.

2.   In the frequency domain, the Power Spectral Density (PSD) method was employed for detailed spectral analysis of the EEG signal. PSD for EEG signals was calculated using the Welch method, utilizing the *scipy.signal.welch* function. The PSD data was then transformed through logarithmic scaling and normalization to ensure comparability. Analyzing EEG data based on its frequency components is a critical method for revealing the spectral properties of the signal. This method involves creating a scalogram by applying various filtering operations to the EEG signal and calculating the PSD. It analyzes the power spectral density of the signal over defined time windows, enabling the tracking of changes at specific frequencies.

3.   In the time-frequency domain, the cleaned EEG signal was transformed into a spectrogram using the signal.spectrogram function. The spectrogram analyzes the frequency components of the signal over time based on the Short-Time Fourier Transform (STFT). As previously mentioned, spectrogram transformation allows for the simultaneous visualization and analysis of both temporal and frequency components of the signal. During this process, the signal is divided into time windows of specific lengths, and the Fourier Transform is applied to each window. This results in a matrix that illustrates how the signal evolves over time, with each matrix element representing the signal's power within a given time and frequency range. The generated spectrograms were further transformed through logarithmic scaling and normalized to the interval [0,1].

Classification

The data obtained from the three different feature spaces were classified using similar deep learning algorithms. For classification, simple and straightforward methods were utilized. Standard CNN models were created according to the index dimensions of the data obtained during feature extraction. Interictal and preictal spectrogram files, excluding those associated with the current seizure, were selected for training purposes. For training and validation, 75% of the total number of files were used for training, while the remaining 25% were allocated for validation. This approach allowed for monitoring the general performance of the model and helped prevent overfitting. The model was trained over a specified number of steps and epochs, and training was halted once the validation accuracy reached 97.5%. During the testing phase, the trained model was evaluated on the interictal and preictal data of the excluded seizure. The applied methods are shown in Figure 3.
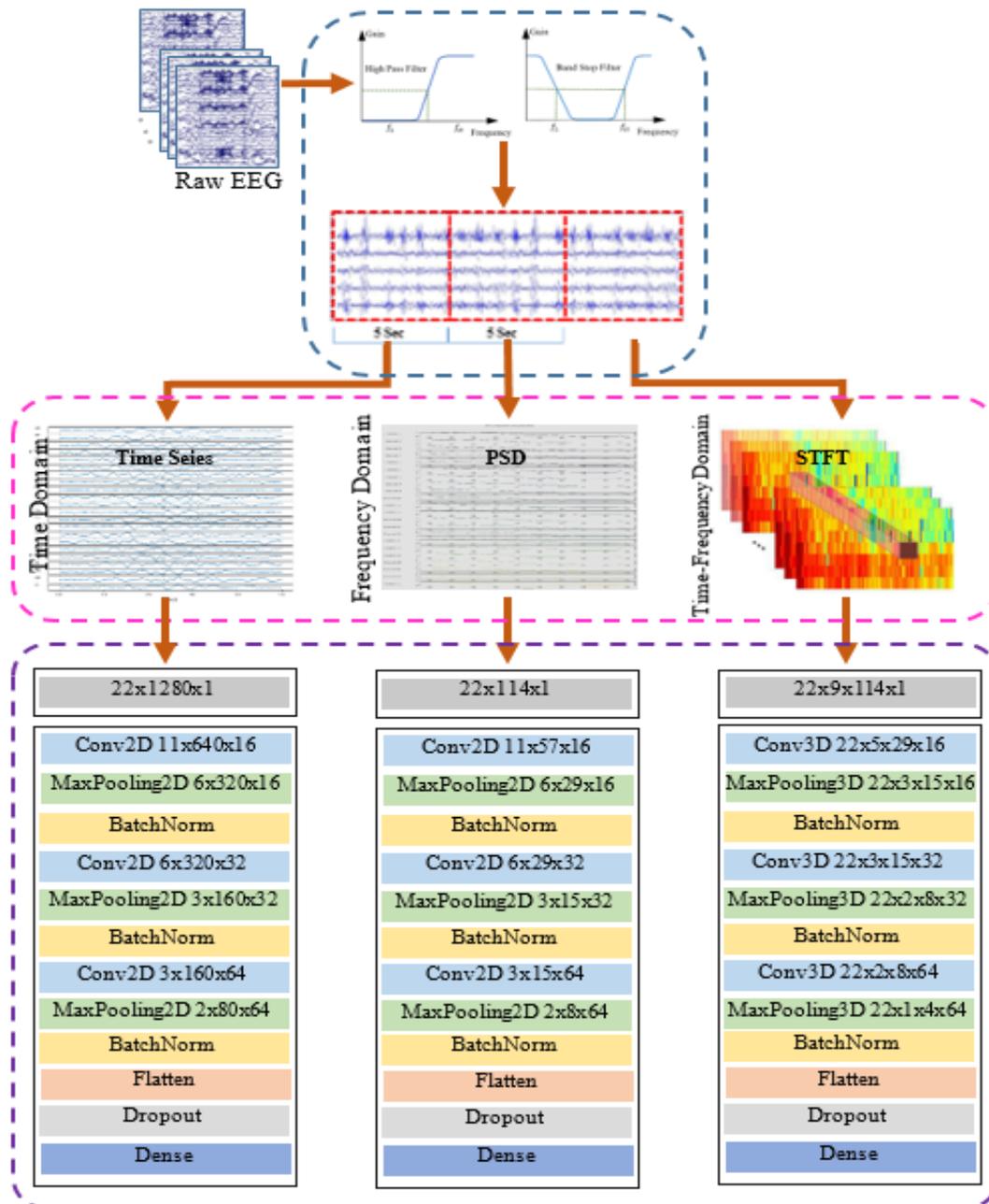
Figure 3: Application of the methods used in the study.

In the CNN models, dimensionality was reduced by applying pooling operations following each convolutional layer. Batch Normalization was applied after each pooling layer to enhance training stability. The Flatten layer was used to convert the 3D/2D data into a 1D structure, and a 50% dropout was applied to mitigate overfitting. In the Dense layer, a 'softmax' activation function was used to produce a 2-neuron output.

Assessment

The model was trained using the training data, and hyperparameter tuning was conducted with the validation data. The model's performance was evaluated on unseen test data. Predictions were made on the resulting set of interictal and real-preictal spectrograms. The evaluation results are presented in Figure 4.

Prediction Value		Actual Value				
		Pozitif (1)	**Negatif (0)**			
	Pozitif (1)	TP	FP	**Accuracy**	**Sensitivity**	**Specificity**
	Negatif (0)	FN	TN	$\dfrac{TP+TN}{TP+TN+FP+FN}$	$\dfrac{TP}{TP+FN}$	$\dfrac{TN}{TN+FP}$
		TP: The number of '1' in the test and prediction. FP: The number of '0' on the test and '1' on the prediction.				
		FN: The number of '1' on the test and '0' on the prediction. TN: The number of '0' on the test and prediction.				

Figure 4: Performance evaluation.

In predictions made on interictal data, false positive (FP) results corresponded to the label [1, 0]. For this evaluation, eight consecutive positive predictions (FP) were counted as a false alarm (FP). Similarly, in predictions made on preictal data, true positive (TP) results corresponded to the label [0, 1]. Eight consecutive positive predictions (TP) were counted as a true alarm (TP). If eight positive predictions did not occur within 10 consecutive predictions, it was considered a false negative (FN). The results were evaluated using accuracy, sensitivity scores.

Result and discussion

In this section, the impact of feature extraction across three different domains on the results is presented through tables and graphs. Table 1 displays the ACC, Sensitivity scores for five randomly selected patients, evaluated using the specified models. No optimization techniques were applied to enhance the models in this study. Consequently, this may have contributed to lower results.

As shown in Table 1, the Time Series + CNN model achieved higher average scores across all metrics. However, the computational cost and time consumption of this model are higher than others. A comparison of the models based on their average ACC scores is illustrated in Figure 5.

Table 1: Performance results of the three applied models.

STFT + CNN					
Patient	Gender	Age	No. Of seizure	ACC(%)	SEN(%)
Chb1	Female	11	7	84.54	84.65
Chb2	Male	11	3	49.53	75.86
Chb5	Female	7	5	48.2	49.02
Chb8	Female	3.5	5	48.74	48.7
Chb16	Female	7	8	52.86	53.96
			Average	56.77	62,04
PSD + CNN					
Patient	Gender	Age	No. Of seizure	ACC(%)	SEN(%)
Chb1	Female	11	7	86.79	86.87
Chb2	Male	11	3	31.74	43.57
Chb5	Female	7	5	47.32	48.36
Chb08	Female	3.5	5	50.17	50.14
Chb16	Female	7	8	55.10	56.21
			Average	54.62	57.43
Time Series + CNN					
Patient	Gender	Age	No. Of seizure	ACC(%)	SEN(%)
Chb1	Female	11	7	87.71	87.71
Chb2	Male	11	3	55.62	85.99
Chb5	Female	7	5	66.96	68.68
Chb08	Female	3.5	5	80.56	81.3
Chb16	Female	7	8	77.60	78.2
			Average	73.69	80.37
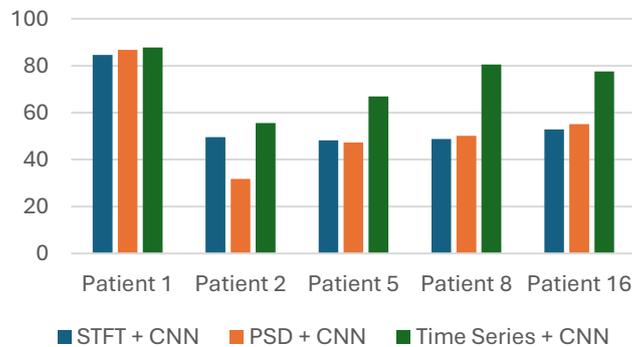
Figure 5: Comparison of accuracy results.

CONCLUSION

In this study, to observe the impact of feature extraction models on the results, different features were derived by applying specified signal analysis methods to the CHB-MIT EEG dataset. Presenting the results of these features in tables, particularly within the context of epileptic seizure prediction applications, is believed to be beneficial for future research in this field.

REFERENCES

[1] L. Kuhlmann, L., Lehnertz, K., Richardson, M. P., Schelter, B., & Zaveri, H. P. (2018). Seizure prediction - ready for a new era. Nature reviews. Neurology, 14(10), 618–630. https://doi.org/10.1038/s41582-018-0055-2

[2] Mohamed E, Yusoff M and Malik A, (2018) Comparison of EEG signal decomposition methods in classification of motor-imagery. BCI, Multimed Tools Appl. 77:21305-213027.G. R. Faulhaber, "Design of service systems with priority reservation," in Conf. Rec. 1995 IEEE Int. Conf. Communications, pp. 3–8.

[3] A. Ahmadi and H. Soltanian-Zadeh, "Epileptic Seizure Prediction Using Spectral Entropy-Based Features of EEG," 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA), Tehran, Iran, 2019, pp. 124-129, doi: 10.1109/PRIA.2019.8785984. Motorola Semiconductor Data Manual, Motorola Semiconductor Products Inc., Phoenix, AZ, 1989.

[4] Bandarabadi, M., Rasekhi, J., Teixeira, C.A., Karami, M.R. (2015).On the proper selection of preictal period for seizure prediction.Epilepsy and Behavior 46:158-66.

[5] Direito, B., Teixeira, C.A., Sales, F., Castelo-Branco, M., Dourado, A.A. (2017). Realistic seizure prediction study based on multiclass SVM. International Journal Of Neural Systems, 27, pp 1750006- 21.

[6] Zhang, Z. and Parhi, K.K. (2016). Low-Complexity Seizure Prediction From iEEG/sEEG Using Spectral Power and Ratios of Spectral Power. IEEE Transactions

[7] Sezer E (2008) Epilepsi teşhisi için EEG sinyal analizi. Selçuk Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi, Konya, Türkiye.

[8] Demren E (2015) Dalgacık dönüşümünün fourier dönüşümü ile karşılaştırılması ve uygulama. İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi, İstanbul, Türkiye.

[9] Orhan U (2011) EEG işaretlerinden epilepsi hastalığının teşhisi için yeni yaklaşımlar. Karaelmas Üniversitesi Fen Bilimleri Enstitüsü Doktora Tezi, Zonguldak, Türkiye.

[10] Hu L, Zhang Z (2019) EEG signal processing and feature extraction. SpringerBriefs in Applied Sciences and Technology Press, Singapore.

[11] Truong, N.D., Nguyen, A.D., Kuhlmann, L., Bonyadi, M. R., Yang, J., and Kavehei, O. (2017). A generalised seizure prediction with convolutional neural networks for intracranial and scalp electroencephalogram data analysis. arXiv:1707.01976.

RESUME

Prof. Dr. Humar KAHRAMANLI ÖRNEK

Humar KAHRAMANLI ÖRNEK graduated from Selçuk University, Faculty of Science and Letters, Department of Mathematics in 1995. Örnek completed both his master's degree and PhD. degree at Selçuk University, Institute of Sciences. Örnek, who has been working at Selçuk University since 2004, has received title of Professor in 2024. Örnek has many national and international scientific papers and publications related to the field of computer engineering.

Atakan DAŞDEMİR

DAŞDEMİR graduated from Selçuk University, Department of Computer Education in 2013 and Department of Computer Engeneering in 2019. Daşdemir completed his master's degree at Selçuk University, Institute of Sciences, Department of Information Technologies Engineering. Daşdemir continues his doctoral education at Selçuk University, Institute of Sciences, Department of Computer Engineering.

ALTERNATIVE DISTRIBUTIONS IN WIND SPEED FORECASTING AND A COMPARISON

Fatih YILMAZ[1], Ahmet PEKGÖR[2], İsmail ARSEL[3], Sema SERVİ[4], Onur İNAN[5], Aşır GENÇ[6], Galip OTURANÇ[7]

INTRODUCTION

Recently, since global warming and environmental problems have increased, sustainable energy sources gained considerable attention. Renewable energy sources have a significant role in reducing the environmental damage caused by fossil fuels. Accordingly, wind energy stands out as one of the clean and sustainable energy sources. Wind energy has a crucial role in generating electricity worldwide and its share is increasing each day [1], [2].

Due to its geographical location and climate, Türkiye has a high wind energy potential. Different wind speeds in various regions of the country, and the efficiency of wind turbines directly affect the success of wind energy projects. Thus, accurate estimation of wind speed data is extremely important to reduce uncertainties in energy production and to support the planning of wind energy projects [3].

In meeting wind energy demand, wind speed forecasts, selection of the site, design and operation of wind power plants are critical. Accurate forecasts contribute to optimizing energy production, reducing operating costs, and reducing environmental impacts. In this regard, analyzing wind speed data with statistical modeling methods provides valuable information to decision-makers in the energy sector [4].

In the literature, the Weibull distribution is often preferred for modeling wind speed data since it is a right-skewed distribution. This distribution is adopted to better understand the characteristics of wind speed and probability distributions. In modeling, each distribution is defined by certain parameters and estimating these parameters is critical to improve the accuracy of wind speed forecasts.

[1] Ministry of National Education, Konya/Türkiye, Orcid: https://orcid.org/0000-0001-7852-6756, yilmazoglu@gmail.com

[2] Necmettin Erbakan University, Konya/Turkiye, Orcid: https://orcid.org/0000-0001-9446-79, apekgor@erbakan.edu.tr

[3] Karamanoğlu Mehmetbey University, Karaman/Türkiye, Orcid: https://orcid.org/0000-0001-8570-8443, ismail.arsel@gmail.com

[4] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0000-0003-2069-9085, semaservi@selcuk.edu.tr

[5] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0009-0005-3053-9482, oinan@selcuk.edu.tr

[6] Necmettin Erbakan University, Konya/Turkiye, Orcid: https://orcid.org/0000-0002-0339-6050, agenc@erbakan.edu.tr

[7] Karamanoğlu Mehmetbey University, Karaman/Türkiye, Orcid: https://orcid.org/0000-0003-3809-9694, goturanc@kmu.edu.tr

In this paper, right-skewed Gamma and Rayleigh distributions, which are alternatives to the Weibull distribution, are employed to forecast wind speeds for 2003 from daily wind speed data obtained from five regions in Türkiye for the years 1998-2002. The performance of the forecasts found using Weibull, Rayleigh, and Gamma distributions will be evaluated with the help of the Mean Square Error (MSE).

MODELLING METHODS

In this study, Weibull, Gamma, and Rayleigh distributions will be used to model wind speed data. The common characteristic of these distributions is that they are right-skewed distributions.

Weibull Distribution

The Weibull distribution is commonly used to model the distribution of wind speed. The probability density function of the Weibull distribution, where α is the shape parameter and β is the scale parameter,

$$f(x; \alpha, \beta) = \begin{cases} \left(\frac{\alpha}{\beta}\right)\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}, & x \geq 0 \\ 0, & dy. \end{cases}$$

and the distribution function is,

$$F(x; \alpha, \beta) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\left(\frac{x}{\beta}\right)^{\alpha}}, & x \geq 0 \end{cases}$$

[5].

Gamma Distribution

Time is a continuous variable, and the Gamma distribution is widely used in the literature for modeling time between events and processes. The probability density function and the distribution function of the Gamma distribution, where α is the shape parameter and β is the scale parameter, respectively, are

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, & x \geq 0 \\ 0, & dy. \end{cases}$$

$$F(x; \alpha, \beta) = \begin{cases} 0, & x < 0 \\ \frac{1}{\Gamma(\alpha)} \gamma\left(\alpha, \frac{x}{\beta}\right), & x \geq 0 \end{cases}$$

[5].

Rayleigh Distribution

The Rayleigh distribution is useful for modeling the total magnitude of quantities that have two components perpendicular to each other (e.g., the x and y directions) and those components

have normal distributions. In nature, it is common for two components of an event to occur randomly in directions perpendicular to each other, and the quantities resulting from the combination of these components can be modeled by the Rayleigh distribution. The Rayleigh distribution has a scale parameter β, and probability density and distribution functions, respectively, are;

$$f(x;\beta) = \begin{cases} \dfrac{x}{\beta^2} e^{\left(-\frac{x^2}{2\beta^2}\right)} & , \quad x \geq 0 \\ 0 & , \quad dy. \end{cases}$$

$$F(x;\beta) = \begin{cases} 0 & , \quad x < 0 \\ 1 - e^{\left(-\frac{x^2}{2\beta^2}\right)} & , \quad x \geq 0 \end{cases}$$

[5].

Parameter Estimation

Parameter estimation is the process of estimating the unknown parameters of a model or distribution from the available data. The most well-known parameter estimation method is the Least Squares (LS) method. The purpose of this method is to minimize the sum of squares of the errors. The squared loss function with ordered observed values $x_{(i)}$, parameter vector $\boldsymbol{\theta}$ and inverse empirical distribution function $F^{-1}(p; \boldsymbol{\theta})$ are,

$$\sum_{i=1}^{n} \left(x_{(i)} - F^{-1}(p; \boldsymbol{\theta})\right)^2$$

Here p is the experimental probability value. For the p value in the literature, the following formula is widely used

$$\mathbf{p} = \frac{i - c}{n + 1 - 2c} \quad , \qquad 0 \leq c \leq 1 , \qquad i = 1,2,3,\dots,n-1$$

where $\mathbf{p} = \left\{\dfrac{i-0.5}{n}, \dfrac{i}{n+1}, \dfrac{i-0.3}{n+0.4}\right\}$

[6], [7].

Performance Evaluation

The performance of each model will be evaluated utilizing MSE values. MSE, which is the mean of the squared structure of the missing form, is calculated using the following formula,

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2$$

[8].

FINDINGS

For the application, the researchers used daily average wind speed data obtained from the General Directorate of Meteorology for the years 1998-2003 at Zonguldak, Erzincan, Van,

Kayseri, and Muğla measurement stations. The data for the years 1998-2002 were used to estimate the parameter values of Weibull, Gamma, and Rayleigh distributions. They used 2003 data for experimental probability and distribution performances.

Parameter estimation values for the distributions obtained from annual average wind speed data are available in Table 1. According to Table 1, it is clear that the parameter estimation values vary according to the regions, although, the variation in the experimental probability values is not significant.

MSE values according to regions and distributions, following parameter estimation, are available in Tables 2, 3, and 4. It is obvious from those tables that the MSE values of the Rayleigh distribution show the best performance compared to other distributions both annually and in all months except for some months in Kayseri and Muğla regions.

Table 1: Experimental probability value of annual average wind speed data and parameter estimation values of distributions by region

Experimental probability values	Regions	Weibull		Gamma		Rayleigh
		α	β	α	β	β
$(i - 0.5)/n$	ZONGULDAK	0.57292	1.49254	0.60855	1.45364	1.90576
	ERZINCAN	0.79937	1.24646	0.83153	1.17465	1.12697
	VAN	0.55485	1.45140	0.57297	1.39498	1.94238
	KAYSERI	0.70673	1.30495	0.77404	1.2665	1.30532
	MUGLA	0.69524	1.30853	0.75563	1.2592	1.32944
$i/(n + 1)$	ZONGULDAK	0.55930	1.45381	0.59823	1.43133	1.90664
	ERZINCAN	0.80203	1.25398	0.82092	1.16193	1.12747
	VAN	0.54744	1.43492	0.59737	1.4567	1.94336
	KAYSERI	0.69425	1.28969	0.77158	1.26477	1.30591
	MUGLA	0.69556	1.31584	0.74700	1.24712	1.33003
$(i - 0.3)/(n + 0.4)$	ZONGULDAK	0.56346	1.46252	0.60244	1.44006	1.90605
	ERZINCAN	0.78273	1.22943	0.82807	1.17075	1.12717
	VAN	0.55972	1.47614	0.5961	1.45235	1.94277
	KAYSERI	0.69974	1.29564	0.77334	1.26633	1.30557
	MUGLA	0.68744	1.29715	0.74648	1.2449	1.32969

Table 2: MSE values for regions and distributions for experimental probability value $p = (i - 0.5)/n$

		January	February	March	April	May	June	July	August	September	October	November	December	Annual
ZONGULDAK	Weibull	10.35429	15.85995	11.27172	6.07326	7.16602	5.57641	7.84505	11.13935	9.69735	7.38067	9.65550	10.77730	11.29471
	Gamma	0.95516	1.26711	1.21892	0.70898	1.50579	1.10603	1.53648	1.53248	1.23803	0.83567	0.65916	0.53257	0.76960
	Rayleigh	**0.07652**	**0.38065**	**0.19781**	**0.11671**	**0.46456**	**0.22379**	**0.36031**	**0.34698**	**0.18120**	**0.35297**	**0.13929**	**0.13942**	**0.05883**
ERZİNCAN	Weibull	0.38614	0.97236	2.87461	2.37479	2.00425	3.12664	3.05661	2.89917	1.00489	0.24827	0.57637	0.34003	1.36888
	Gamma	0.26532	0.24822	0.79729	0.61771	0.61010	0.93475	1.17372	1.03884	0.45161	0.22170	0.39305	0.20129	0.45987
	Rayleigh	**0.09035**	**0.11027**	**0.13705**	**0.12686**	**0.10515**	**0.18160**	**0.28613**	**0.24267**	**0.04223**	**0.02801**	**0.09058**	**0.03712**	**0.02031**
VAN	Weibull	9.06228	7.24530	11.97402	10.66408	12.76616	16.97946	15.09014	14.98983	13.04561	6.91975	5.86942	4.63771	13.91022

	Gamma	1.89847	0.76691	1.50901	0.77506	1.62582	3.12779	2.18677	2.76879	2.31277	0.79615	0.75784	0.67552	1.39080
	Rayleigh	**0.58516**	**0.08932**	**0.52418**	**0.39848**	**0.44065**	**1.13713**	**0.59995**	**0.93143**	**0.76088**	**0.24484**	**0.24833**	**0.19753**	**0.28784**
KAYSERİ	Weibull	0.77112	0.48070	5.52623	5.75365	2.81918	3.42086	3.93241	3.74409	1.63953	0.48315	1.87620	1.68454	2.24856
	Gamma	0.24203	**0.41988**	1.03948	0.54650	0.85747	1.20656	1.47389	1.28542	0.77058	0.31301	0.79889	0.72008	0.45776
	Rayleigh	**0.12395**	1.21264	**0.20594**	**0.25219**	**0.21300**	**0.29936**	**0.40032**	**0.34214**	**0.15940**	**0.30187**	**0.21460**	**0.54035**	**0.16295**
MUĞLA	Weibull	1.14692	2.28388	2.54620	1.54039	3.42943	6.28917	8.24439	5.84900	3.35113	1.71906	0.31390	0.69038	3.13586
	Gamma	0.29341	0.64223	0.57817	0.57148	1.10999	1.44624	1.91069	1.39807	0.87404	0.44025	**0.05286**	**0.12413**	0.57633
	Rayleigh	**0.09140**	**0.35743**	**0.10252**	**0.12321**	**0.22760**	**0.32111**	**0.49888**	**0.28953**	**0.13435**	**0.05065**	0.17173	0.17935	**0.03139**

According to Table 2, for all months and regions and annually as well, the Rayleigh distribution stands out with the smallest MSE values. However, it is noteworthy that Gamma distribution has the smallest MSE values in November and December for Muğla region.

Table 3: MSE values for regions and distributions for experimental probability value $p = i/(n+1)$

		January	February	March	April	May	June	July	August	September	October	November	December	Annual
ZONGULDAK	Weibull	6.84047	9.75807	7.20721	3.87929	4.74551	3.86246	5.46902	7.30667	6.55107	4.94855	5.83226	6.40694	10.21751
	Gamma	0.61634	0.93355	0.80836	0.53424	1.13684	0.83518	1.15545	1.13706	0.91911	0.63363	0.37393	0.29129	0.72220
	Rayleigh	**0.05419**	**0.39955**	**0.14672**	**0.11697**	**0.38883**	**0.17139**	**0.28471**	**0.26985**	**0.12060**	**0.37739**	**0.15179**	**0.16646**	**0.05850**
ERZİNCAN	Weibull	0.28490	0.71419	1.84358	1.59555	1.27184	2.07580	2.20864	2.11104	0.73710	0.18211	0.37997	0.22453	1.21353
	Gamma	0.18824	0.16546	0.56830	0.45024	0.43966	0.70502	0.91551	0.82598	0.29488	0.14634	0.25479	0.11423	0.43029
	Rayleigh	**0.09471**	**0.13714**	**0.10548**	**0.12212**	**0.08965**	**0.13803**	**0.23408**	**0.19016**	**0.02877**	**0.03217**	**0.08208**	**0.04439**	**0.02009**
VAN	Weibull	5.82447	4.96979	7.52912	6.26349	8.12092	11.15282	9.85335	10.21912	8.72008	4.52728	3.33491	2.93060	12.18790
	Gamma	1.43190	0.46474	1.12808	0.53856	1.19561	2.42077	1.64577	2.12469	1.76449	0.53318	0.52860	0.46734	1.20986
	Rayleigh	**0.48234**	**0.07378**	**0.46397**	**0.43222**	**0.37205**	**0.95291**	**0.48066**	**0.77918**	**0.61899**	**0.24581**	**0.23892**	**0.20052**	**0.28312**
KAYSERİ	Weibull	0.47351	**0.38237**	3.71429	3.20700	1.93726	2.41303	2.97007	2.67500	1.14176	0.48636	1.31395	1.24649	2.10556
	Gamma	0.20070	0.61905	0.76541	0.36926	0.68146	0.93125	1.15991	1.01549	0.57790	**0.31521**	0.54431	0.63150	0.43960
	Rayleigh	**0.14170**	1.35146	**0.16045**	**0.27911**	**0.19219**	**0.24800**	**0.33063**	**0.27849**	**0.13041**	0.33854	**0.18493**	**0.56126**	**0.16998**
MUĞLA	Weibull	0.75922	1.51411	1.65700	1.24378	2.43883	4.52689	5.76009	4.01968	2.26997	1.13538	0.13985	0.38059	2.75388
	Gamma	0.17527	0.54507	0.42461	0.40053	0.80997	1.02513	1.49070	1.01828	0.63245	0.28996	**0.03096**	**0.07955**	0.53620
	Rayleigh	**0.11273**	**0.38622**	**0.10480**	**0.12210**	**0.17305**	**0.24169**	**0.40822**	**0.22070**	**0.10676**	**0.04739**	0.20890	0.22153	**0.03063**

Table 3 indicates that for all months and regions and annually as well, the Rayleigh distribution stands out with the smallest MSE values as it is the case in Table 2. However, it is interesting that Gamma distribution has the smallest MSE values in November and December for Muğla

region, once again. Furthermore, Weibull distribution has the smallest MSE value for Kayseri region in February.

Table 4: MSE values for regions and distributions for experimental probability value $p = (i - 0.3)/(n + 0.4)$

		January	February	March	April	May	June	July	August	September	October	November	December	Annual
ZONGULDAK	Weibull	8.67458	12.50071	9.01640	5.36647	6.10821	4.76378	6.37149	9.15104	8.09438	6.21885	7.63026	8.61196	10.97632
	Gamma	0.77376	1.11683	1.01148	0.62685	1.32119	0.96338	1.34971	1.34135	1.06403	0.70802	0.52678	0.40132	0.75187
	Rayleigh	**0.06371**	**0.38585**	**0.17247**	**0.11603**	**0.42869**	**0.19902**	**0.32482**	**0.31054**	**0.15181**	**0.36220**	**0.14305**	**0.14895**	**0.05859**
ERZİNCAN	Weibull	0.32688	0.76562	2.26770	2.05680	1.65690	2.68607	2.38342	2.51348	0.87610	0.23037	0.44590	0.30985	1.39514
	Gamma	0.23201	0.19438	0.66663	0.52429	0.54006	0.83681	1.05106	0.91055	0.35422	0.17883	0.33128	0.14601	0.44490
	Rayleigh	**0.09205**	**0.12179**	**0.12186**	**0.12397**	**0.09779**	**0.16096**	**0.26185**	**0.21813**	**0.03550**	**0.02965**	**0.08628**	**0.03973**	**0.02020**
VAN	Weibull	7.77207	6.33422	9.66854	8.47805	10.24891	13.99454	12.36292	12.60229	10.72616	5.77279	4.48771	3.88783	12.52912
	Gamma	1.72167	0.61565	1.34054	0.65871	1.42974	2.76273	1.89055	2.49496	2.05671	0.67976	0.66204	0.49612	1.26163
	Rayleigh	**0.53704**	**0.07917**	**0.49524**	**0.41157**	**0.40890**	**1.05152**	**0.54448**	**0.86117**	**0.69489**	**0.24469**	**0.24285**	**0.19806**	**0.28577**
KAYSERİ	Weibull	0.60466	**0.39535**	4.62261	4.17296	2.27360	2.96230	3.36686	3.20112	1.37296	0.46074	1.61551	1.48668	2.19499
	Gamma	0.21638	0.50762	0.89679	0.45904	0.75234	1.07978	1.31915	1.17542	0.66444	**0.30276**	0.68971	0.67460	0.44895
	Rayleigh	**0.13204**	1.27570	**0.18445**	**0.26283**	**0.20376**	**0.27554**	**0.36808**	**0.31266**	**0.14589**	0.31899	**0.20049**	**0.55051**	**0.16615**
MUĞLA	Weibull	0.96398	1.90162	1.99579	1.33945	2.99245	5.23490	6.92515	5.09187	2.95571	1.34340	0.20723	0.50727	3.06271
	Gamma	0.23537	0.59293	0.49079	0.45911	0.96740	1.21773	1.64739	1.21625	0.77760	0.33999	**0.04141**	**0.09277**	0.56566
	Rayleigh	**0.10018**	**0.36918**	**0.10258**	**0.12180**	**0.20177**	**0.28362**	**0.45650**	**0.25712**	**0.12087**	**0.04815**	0.18807	0.19745	**0.03098**

Considering Table 4, it is obvious that the output is similar to the one in Table 3. Once more, for all months and regions and annually as well, the Rayleigh distribution stands out with the smallest MSE values as. However, the Gamma distribution has the smallest MSE values in November and December for Muğla region, and the Weibull distribution has the smallest MSE value for Kayseri region in February.

CONCLUSION

In this paper, the researchers used daily wind speed data obtained from five regions in Türkiye and made wind speed forecasts for the year 2003 by utilizing the data between the years 1998-2002. The accuracy of the forecasts obtained using Weibull, Rayleigh, and Gamma distributions was evaluated utilizing the MSE criterion. The results reveal that the Rayleigh distribution has better-fit performance than the other two distributions. The results also indicate that the Rayleigh distribution better reflects the characteristics of the wind speed data, consequently, giving more reliable results in wind speed forecasting models. It is, particularly, obvious that the Rayleigh distribution has higher accuracy compared to Weibull and Gamma distributions at low and medium values of wind speed. Moreover, the change in the experimental probability values used in the models did not have a significant effect on the results. The findings suggest that the Rayleigh distribution can be preferred for wind energy projects. The findings also emphasize the importance of such statistical modeling to better assess the country's wind energy

potential. Doing similar analyses for wind energy plays a crucial role in increasing efficiency in energy production and ensuring environmental sustainability as well.

REFERENCES

[1] Ember, "Data into action," https://ember-energy.org/.

[2] S. M. Valdivia-Bautista, J. A. Domínguez-Navarro, M. Pérez-Cisneros, C. J. Vega-Gómez, and B. Castillo-Téllez, "Artificial Intelligence in Wind Speed Forecasting: A Review," *Energies (Basel)*, vol. 16, no. 5, p. 2457, Mar. 2023, doi: 10.3390/en16052457.

[3] M. A. Benmahdjoub, A. Mezouar, M. Ibrahim, L. Boumediene, Y. Saidi, and M. Atallah, "Accurate estimation of effective wind speed for wind turbine control using linear and nonlinear Kalman Filters," *Arab J Sci Eng*, vol. 48, no. 5, pp. 6765–6781, May 2023, doi: 10.1007/s13369-022-07498-7.

[4] H. Chen, "A comprehensive statistical analysis for residuals of wind speed and direction from numerical weather prediction for wind energy," *Energy Reports*, vol. 8, pp. 618–626, Nov. 2022, doi: 10.1016/j.egyr.2022.07.080.

[5] F. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*, 4th ed. New York, USA: McGraw-Hill, 2002.

[6] M. , A. Stephens, "Tests Based on EDF Statistics," in *Goodness-of-fit techniques*, vol. 68, R. B. D'Agostino and M. A. Stephens, Eds., New York: Marcel Dekker Inc., 1986, ch. 4, pp. 97–194.

[7] A. Genc, M. Erisoglu, A. Pekgor, G. Oturanc, A. Hepbasli, and K. Ulgen, "Estimation of wind power potential using Weibull distribution," *Energy Sources*, vol. 27, no. 9, pp. 809–822, Jul. 2005, doi: 10.1080/00908310490450647.

[8] F. M. Dekking, C. Kraaikamp, H. P. Lopuhaä, and L. E. Meester, "Efficiency and mean squared error," in *A modern introduction to probability and statistics*, London: Springer, 2005, pp. 299–311. doi: 10.1007/1-84628-168-7_20.

RESUME

Dr. Fatih YILMAZ

Dr. Fatih YILMAZ graduated from the Middle East Technical University, Faculty of Education, department of foreign language education in 2007. He completed his master's degree at Selcuk University, Faculty of Education, department of Educational Administration Supervision Planning and Finance, in 2009, and his PhD. at Necmettin Erbakan University in 2019. He has a special interest in applied statistics and completed his master's degree at Necmettin Erbakan University, Department of Statistics in 2024. Yılmaz, who has been working at the Ministry of National Education since 2007, is married and has a son.

Assoc. Prof. Dr. Ahmet PEKGÖR

Associate Professor Dr. Ahmet PEKGÖR graduated from Ankara University, Faculty of Science, Department of Statistics in 1998. He completed his master's degree at Selçuk University, Department of Statistics in 2004, and completed his PhD. at Selcuk University, Department of Mathematics in 2010. Pekgör worked as a research assistant at Selcuk University, Faculty of Science, Department of Statistics between 2001-2014. He started working as an Assistant Professor at Necmettin Erbakan University, Faculty of Science, Department of Statistics in 2014. He received the title of Associate Professor in 2023 and has still been working as a vice dean at Necmettin Erbakan University, Faculty of Science. Pekgör, having several national and international papers and publications, is married and has two sons.

Assist. Prof. Dr. İsmail ARSEL

Assistant Professor Dr. İsmail ARSEL graduated from Ankara University, Faculty of Science, Department of Physics in 1980. He completed his master's degree at Ankara University, Institute of Science, and his PhD. at Ege University, Solar Energy Institute Department of Energy Technologies in 1989. He worked as an assistant Professor at Batman University, Faculty of Science between 2012-2017. He has been working as an assistant Professor at Karamanoğlu Mehmetbey University, Kamil Özdağ Faculty of Science since 2018.

Assistant Prof. Dr. Sema SERVİ

Sema SERVI is a Assistant Professor Dr. in Computer Engineering at Selcuk University, Faculty of Technology. She completed her master's and doctorate degrees at Selcuk University, Faculty of Science, Department of Mathematics. Servi has published articles on various national and international journals on topics such as applied mathematics and artificial intelligence. Her main research interests include numerical approximation solution methods, optimization algorithms and data mining.

Assist. Prof. Dr. Onur İNAN

Onur İNAN is an Assistant Professor in the Technology Faculty at Selçuk University. He completed both his master's and Ph.D. at Selçuk University. He has published articles on various optimization algorithms and data mining. His primary research interests include machine learning, artificial intelligence, and neural networks.

Prof. Dr. Aşır GENÇ

Prof. Dr. Aşır GENÇ graduated from Ankara University, Faculty of Science, Department of Mathematics in 1986. He completed both his master's degree and PhD. at Ankara University, Institute of Science, Department of Statistics in 1990 and 1997, respectively. Genç started working as a research assistant at Ankara University, Faculty of Science, Department of Mathematics between 1987-1990 and worked as a lecturer at Ankara University, Faculty of Science, Department of Statistics between 1990-1997. He started working as an Assistant Professor at Selcuk University in 1998 and received the title of Associate Professor in 2006 and the title of professor in 2011 and worked as a professor at Selcuk University till 2018. Genç, who has several national and international papers and publications in statistics, has been working as a professor at Necmettin Erbakan University, Faculty of Science, Department of Statistics since 2018. Genç is married and has three sons.

Prof. Dr. Galip OTURANÇ

Prof. Dr. Galip OTURANÇ graduated from Ankara University, Faculty of Science, Department of Mathematics in 1987. He completed both his master's degree and his PhD. at Ege University in 1989 and 1991, respectively. He worked as a research assistant at Ege University, in the Department of Mechanical Engineering for a while. He worked as an Assistant Professor in the Department of Mathematics at Selçuk University between 1995-2018. He received the title of associate professor in 2006, and the title of professor in 2011. He has been working at Karamanoğlu Mehmetbey University since 2018. Oturanç, has been working in Applied Mathematics, is married, and has two children.

TIME-FREQUENCY ANALYSIS OF BIOMEDICAL SIGNALS IN SLEEP APNEA

Göksu DEMİRCİ[1], Adem GÖLCÜK[2]

INTRODUCTION

Sleep apnea is a prevalent sleep disorder characterized by repeated episodes of breathing cessation or reduction during sleep [1, 2]. This condition disrupts sleep architecture, impedes the transition to deeper sleep stages, and is often associated with symptoms like excessive daytime sleepiness, loud snoring, and fragmented sleep patterns. Sleep apnea generally manifests in three forms: obstructive sleep apnea (OSA), central sleep apnea (CSA), and mixed sleep apnea, a combination of the two [3]. Among these, OSA is the most common and arises when the muscles supporting the soft tissues in the throat relax, causing partial or complete obstruction of the upper airway.

Biomedical signal processing has become essential in understanding, diagnosing, and monitoring sleep apnea [4-6]. Various physiological signals such as electrocardiograms (ECG), electroencephalograms (EEG), and oxygen saturation levels (SpO2) are frequently recorded in polysomnography (PSG) tests, the standard method for sleep apnea diagnosis [7]. This comprehensive monitoring approach allows researchers and clinicians to capture and interpret time- and frequency-domain characteristics of these signals, which provides insight into the patient's overall physiological state during sleep.

Recent studies have utilized time-frequency analysis to examine the behavior of SpO2 and ECG signals, which are essential indicators of respiratory and cardiovascular function, respectively [6, 8]. The focus on time-frequency analysis is driven by its ability to capture transient signal features that static time- or frequency-only analyses might overlook. Techniques such as Short-Time Fourier Transform (STFT) and Fast Fourier Transform (FFT) enable researchers to assess how signal frequency components change over time [9], offering a dynamic perspective on the body's physiological responses. These analyses can identify specific markers or abnormalities, such as the presence of low-frequency components in the ECG that may indicate bradycardia or variability in SpO2 indicating respiratory distress [9].

In this study, we aim to explore the time-frequency characteristics of SpO2 and ECG signals associated with sleep apnea episodes, using the open-source PSG-Audio dataset available on the scidb.cn database. By applying both time-domain and frequency-domain analyses, this research seeks to identify patterns and dominant frequencies related to apnea episodes [10]. The results of this study may aid in the development of more effective diagnostic and monitoring

---

[1] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0009-0001-1375-4202, demircigoksu@gmail.com
[2] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0000-0002-6734-5906, adem.golcuk@selcuk.edu.tr

tools, helping clinicians make informed decisions about the management and treatment of sleep apnea [10, 11].

MATERIALS AND METHODS

Dataset

The PSG-Audio dataset used in this study comprises polysomnographic data of 212 patients and is publicly available on the scidb.cn platform [12]. This dataset was created through collaboration between the Sleep Research Unit at Sismanoglio Amalia Fleming General Hospital in Athens and the Department of Electrical and Electronic Engineering at the University of West Attica. Each patient's data includes several physiological signals collected during sleep, such as EEG, EOG, EMG, ECG, airflow, body position, and SpO2 levels [13]. Additionally, each record is accompanied by a corresponding audio file that captures ambient sounds, including any sounds indicative of sleep apnea events like snoring or gasping [14]. The associated publication provides further insights into the dataset's structure and methodology [12].

Table 1: Information of 20 channels in the data set

Channel label	Description	Sampling frequency (Hz)	Physical minimum (a.u.)	Physical maximum (a.u.)	Digital minimum	Digital maximum
EEG A1-A2	Electroencephalogram (reference channe)	200	-313	313	-32768	32767
EEG C3-A2	Electroencephalogram (channel 1)	200	-313	313	-32768	32767
EEG C4-A1	Electroencephalogram (channel 2)	200	-313	313	-32768	32767
EOG LOC-A2	Electroculogram (left eye)	200	-313	313	-32768	32767
EOG ROC-A2	Electroculogram (right eye)	200	-313	313	-32768	32767
EMG Chin	Electromyogram (placed on chin)	200	-78	78	-32768	32767
Leg 1	Leg movement signal (right leg)	200	-78	78	-32768	32767
Leg 2	Leg movement signal (left leg)	200	-78	78	-32768	32767
ECG I	Electrocardiogram	200	-8333	8333	-32768	32767
RR	RR interval in the ECG	10	0	200	0	200
Snore	Contact microphone (placed on trachea)	500	-100	100	-32768	32767
Flow Patient	Thermistor	100	-100	100	-32768	32767
Flow Patient	Pressure cannula	100	-100	100	-32768	32767
Effort THO	Respiratory belt (thoracic volume changes)	100	-100	100	-32768	32767
Effort ABD	Respiratory belt (abdomen volume changes)	100	-100	100	-32768	32767
SpO2	Oxymeter	1	0	102.3	0	1023
Body	Position of the body (S=supine, R=right, L=left)	1	0	255	0	255
PulseRate	The pulse rate extracted by the ECG	1	0	255	0	255
Tracheal	High-quality contact microphone (placed on trachea)	48000	-100	100	-32768	32767
Microphone	Ambient microphone (placed ~1 m above patient head)	48000	-100	100	-32768	32767

For this study, we specifically extracted ECG and SpO2 signals from the dataset due to their high relevance in diagnosing sleep apnea. ECG signals reflect heart activity, offering insights into heart rate and rhythm variability, while SpO2 signals represent blood oxygen saturation levels, essential for understanding respiratory patterns.

Data Preprocessing

The ECG and SpO2 signals were processed in the MATLAB environment to ensure accuracy and consistency in subsequent analyses. Key steps in preprocessing included:

1)Signal Extraction: Using the MATLAB ReadEDF function, we extracted the relevant data columns—column 9 for ECG and column 16 for SpO2. Each patient's signal data was stored in EDF format, and time vectors were created to standardize each signal's time series, allowing for synchronization across all analyses.

2)Noise Filtering: Biomedical signals often contain noise due to environmental factors, patient movement, or electrical interference.

3)Segmentation: The signals were divided into segments based on time intervals, enabling the application of Short-Time Fourier Transform (STFT) for localized frequency analysis [15, 16]. This segmentation allowed for the observation of frequency changes within specific time frames, essential for detecting transient apnea episodes.

Time Domain Analysis

In the time domain, we calculated statistical parameters such as mean, standard deviation, minimum, and maximum for both ECG and SpO2 signals [17, 18]. For ECG signals, these statistics help in understanding the variability of the heart rate over time, while for SpO2, they provide insights into the stability of oxygen saturation levels. These measurements provide a basis for assessing whether the signals exhibit any significant deviations during apnea events, such as sudden drops in SpO2 or irregular heartbeats.
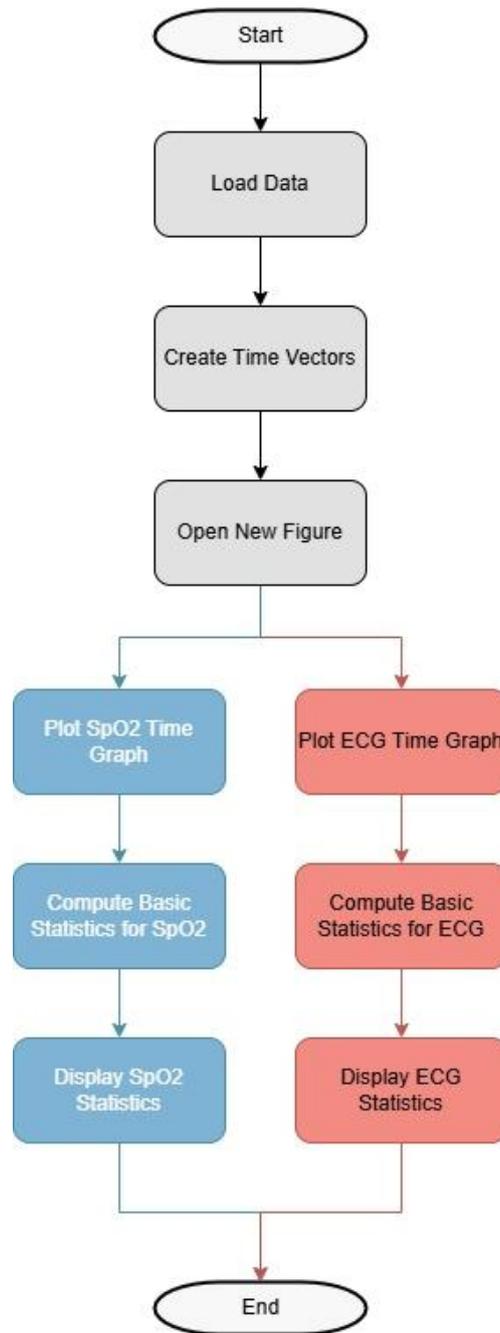
Figure 2: Flowchart of SpO$_2$ and ECG Data Processing Steps

Additionally, time-series plots were generated to visualize the ECG and SpO2 fluctuations over time. The time domain representation of SpO2, for example, can reveal patterns of oxygen desaturation, which is commonly associated with apnea episodes. ECG time plots, on the other hand, can illustrate instances of bradycardia or tachycardia, which may correlate with apnea events.
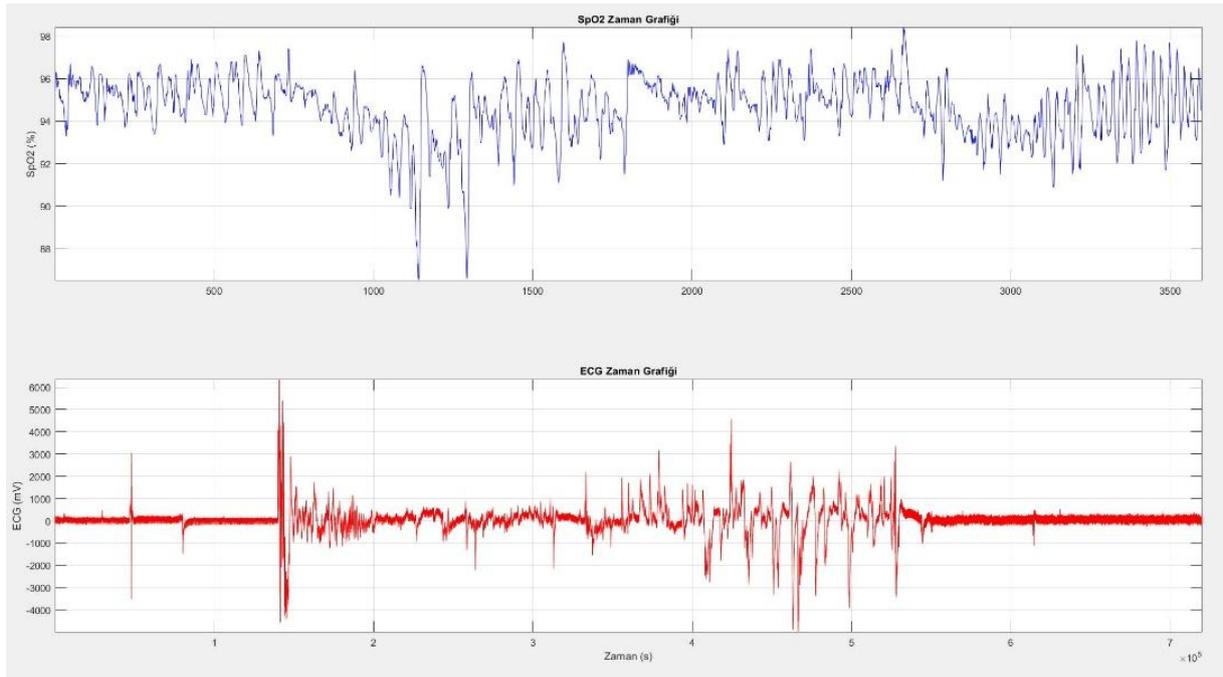
Figure 3: SpO2 – ECG Time Domain Graphs

Frequency Domain Analysis

Frequency domain analysis was performed using the Fast Fourier Transform (FFT), which decomposes the signals into their frequency components [19, 20]. FFT was applied to determine the dominant frequencies in both ECG and SpO2 signals, offering insights into the rhythmic patterns of these physiological signals. This analysis is particularly important for ECG signals, where a dominant low frequency could indicate a slow heart rate typical in apnea patients.
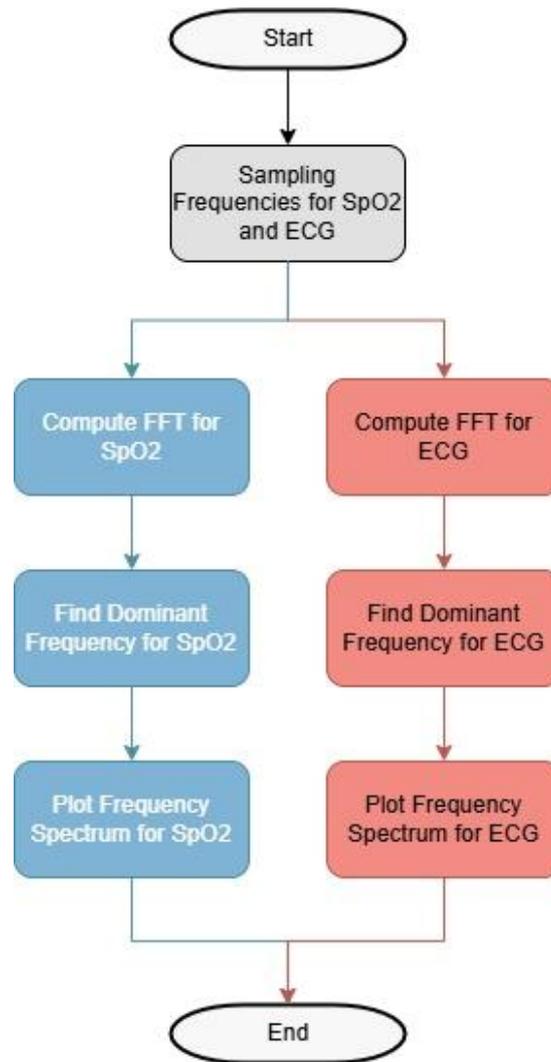
Figure 4: Flowchart of Frequency Analysis for SpO₂ and ECG Data

The sampling frequency for ECG was set to 200 Hz, capturing a high-resolution view of heart rate variability, while SpO2 was sampled at 1 Hz, as oxygen saturation changes tend to occur more slowly. Through the FFT process, the prominent frequency peaks were identified, allowing us to determine which frequency bands are most active during apnea episodes. For instance, the analysis revealed that the dominant frequency for SpO2 signals was 0 Hz, suggesting a relatively stable oxygen level in the absence of apnea [21].
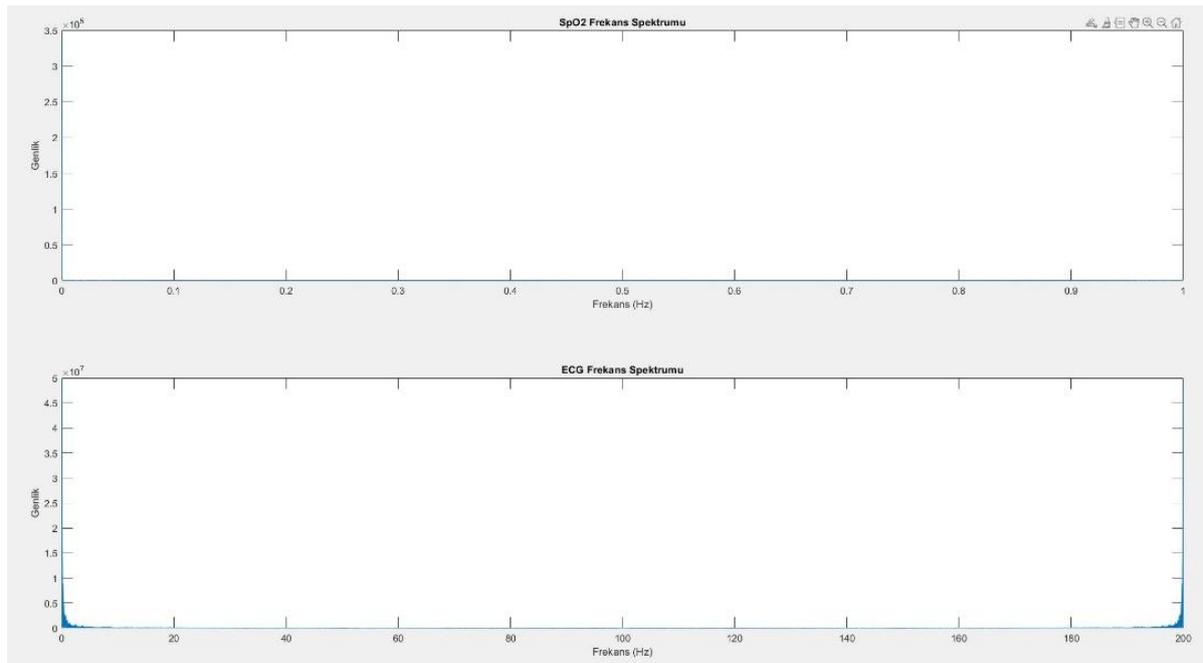
Figure 5: SpO2 – ECG Frequency Domain Graphs

Short-Time Fourier Transform (STFT) Analysis

The Short-Time Fourier Transform (STFT) was applied to both ECG and SpO2 signals to capture changes in frequency content over time. STFT divides each signal into overlapping windows and computes the Fourier transform for each segment, producing a spectrogram where frequency components are displayed over time.
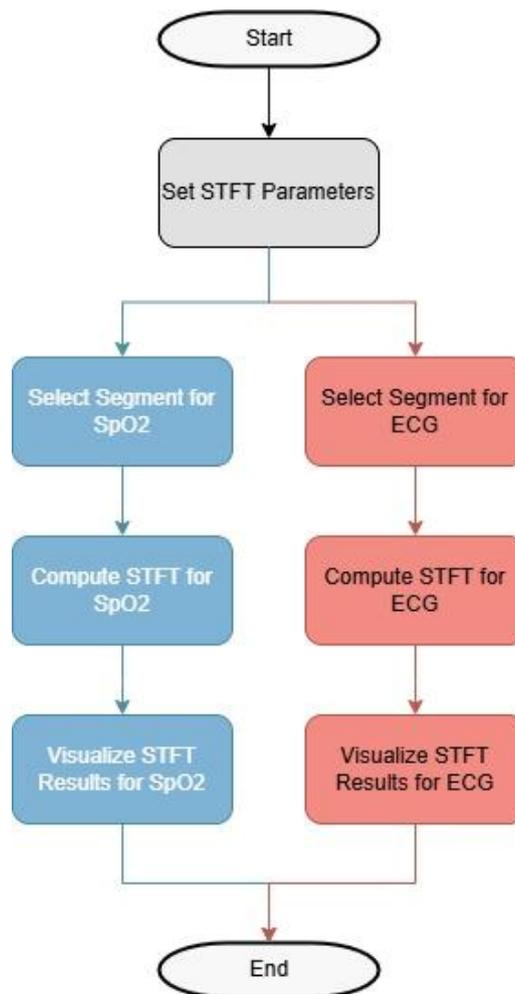
Figure 6: Flowchart of STFT Analysis for SpO₂ and ECG Data

For STFT, we used a Hamming window with a size of 256 data points and an overlap of 75% to balance frequency and time resolution. This choice of parameters provided a detailed view of how frequency components evolved, capturing transient phenomena that would be difficult to detect with traditional FFT.

The STFT analysis for SpO2 signals displayed a predominantly blue spectrum, indicating low-power frequency components, which align with the expected stable oxygen saturation during non-apnea periods. In contrast, the ECG spectrogram revealed distinct bands of higher intensity, corresponding to the periodic nature of heartbeats. These periodic fluctuations were indicative of a regular heart rhythm, but some segments displayed abnormal spikes, which may be associated with apnea episodes.
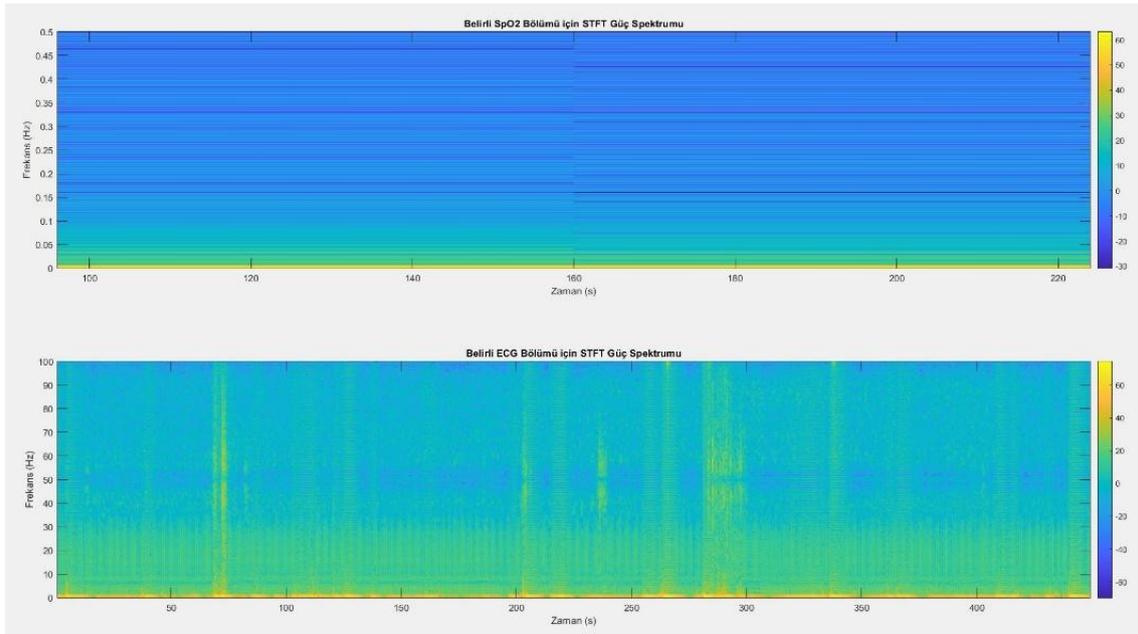
Figure 7: STFT Power Spectrum

The Short-Time Fourier Transform (STFT) analysis was conducted with a Hamming window of 256 points, a 75% overlap, and an FFT point size of 512. These parameters were chosen to balance time and frequency resolution. ECG signals were sampled at 200 Hz, allowing for a high temporal resolution, while SpO2 signals were sampled at 1 Hz, reflecting the slower changes in oxygen saturation. These settings ensured that transient frequency events associated with apnea episodes were captured effectively.

RESULTS

The analysis revealed distinct patterns in both time and frequency domains that characterize the physiological changes associated with sleep apnea. In the time domain, SpO2 signals showed occasional desaturation events, with oxygen levels dropping below the standard range during apnea episodes. These events were short-lived, with SpO2 levels typically returning to baseline once the apnea episode ended. ECG time domain analysis indicated irregularities in heart rate, including bradycardia during apnea events, which aligns with previous studies associating apnea with slowed heart rate.

The FFT analysis provided insights into the dominant frequency components for both signals. For SpO2, the dominant frequency remained at 0 Hz, reflecting a generally stable oxygen saturation level over time. However, during apnea episodes, the amplitude of low-frequency components increased, indicating periodic oxygen desaturation. For ECG, the dominant frequency was found to be 0.0125 Hz, which aligns with a regular, slow heart rhythm. This frequency pattern was punctuated by irregular spikes during apnea events, suggesting heart rate variability related to respiratory distress.

The STFT analysis produced spectrograms that highlighted changes in frequency components over time. The ECG spectrogram, for instance, showed bursts of high power at specific frequencies, potentially indicating arrhythmias or irregular heartbeats associated with apnea.

The SpO2 spectrogram confirmed that oxygen saturation was stable under normal conditions but exhibited low-frequency fluctuations during apnea events, indicating transient desaturation.

The application of STFT further allowed for temporal and spectral analysis that distinguished the differences in physiological states during apnea and non-apnea periods. Through a detailed examination of the spectrograms generated from STFT, we identified that the SpO2 signal, generally stable during regular breathing, exhibited distinct low-frequency oscillations during apnea events. These oscillations correspond to periods of decreased oxygenation, consistent with respiratory cessation or significant reduction in airflow. Similarly, the ECG spectrogram displayed frequency components that varied with each apnea episode, highlighting a rise in low-frequency heart rate variability (HRV) components, which are typically associated with autonomic responses to hypoxic conditions.

During non-apnea periods, both SpO2 and ECG signals displayed regular patterns without sudden fluctuations. In contrast, apnea episodes were marked by abrupt changes in both the amplitude and frequency of the signals. The ECG signal, for instance, showed signs of sinus arrhythmia and occasional bradycardia, likely reflecting the physiological response to intermittent hypoxia. These findings are consistent with the literature that associates sleep apnea with increased parasympathetic activity, particularly during episodes of oxygen desaturation.

Table 1: Statistical Analysis

Parameter	SpO2	ECG
Mean	%94.64	40.94 mV
Standard Deviation	1.44	599.12 mV
Maximum	%98.4	6361.61 mV
Minimum	%86.5	-4996.74 mV
Length	3600	720000

The statistical analysis of the SpO2 signal revealed an average value of 94.64% with a standard deviation of 1.44%, indicating overall stability with minor fluctuations. ECG signals exhibited a mean value of 40.94 mV and a standard deviation of 599.12 mV, reflecting significant variability consistent with cardiac activity. Maximum and minimum values for both signals were within expected physiological ranges, providing further validation of signal reliability.

The length parameter, representing the total number of data points, was 3600 for SpO2 and 720,000 for ECG. For SpO2, this corresponds to one data point per second over a one-hour period, while for ECG, sampled at 200 Hz, it represents a total of one hour of recording at high temporal resolution.

DISCUSSION

The results of this study demonstrate the potential of time-frequency analysis in identifying and characterizing the unique physiological changes associated with sleep apnea. By focusing on SpO2 and ECG signals, we have shown that it is possible to capture relevant biomarkers for apnea episodes, which could be invaluable in clinical diagnostics and automated monitoring systems.

Physiological Implications

The study supports the hypothesis that sleep apnea episodes cause temporary changes in respiratory and cardiovascular systems. Periodic desaturation in SpO2 levels and heart rate irregularities indicate the body's response to hypoxic stress, with low-frequency oscillations in both signals showing efforts to restore oxygen levels. These findings align with previous studies linking apnea severity to autonomic dysregulation.

The presence of low-frequency components in ECG during apnea episodes, particularly those linked to bradycardia, further illustrates apnea's impact on cardiovascular health. Frequency analysis may help predict long-term cardiovascular risks, as prolonged exposure to low-frequency HRV components may signal underlying heart conditions, such as arrhythmias, exacerbated by apnea.

The dominant frequency for ECG signals, identified as 0.0125 Hz, suggests a low heart rate consistent with bradycardia, particularly during apnea episodes. Similarly, the SpO2 signal's dominant frequency of 0 Hz indicates a stable oxygen saturation, reflecting the absence of significant respiratory distress during non-apnea periods. These findings align with the hypothesis that apnea episodes disrupt normal physiological rhythms, with observable impacts on both cardiovascular and respiratory systems.

Methodological Contributions

The combined use of FFT and STFT effectively distinguishes between apnea and non-apnea periods in PSG data. FFT offers a static view of dominant frequencies, while STFT captures transient changes, providing a complete picture of apnea events. This approach can be valuable for real-time monitoring, where continuous frequency tracking could trigger alerts.

Preprocessing steps, including filtering and segmentation, were essential to reliable analysis. By removing high-frequency noise, we focused on significant apnea-related frequencies. Data segmentation allowed us to target apnea episodes directly, enhancing the accuracy of physiological pattern identification.

CONCLUSION

This study demonstrates that time-frequency analysis of ECG and SpO2 signals offers insights into the physiological changes associated with sleep apnea. By applying FFT and STFT, we identified biomarkers in these signals correlating with apnea episodes, providing a foundation for future automated diagnostics. Key takeaways include:

1) Identification of Key Biomarkers: Low-frequency oscillations in SpO2 and ECG signals during apnea episodes serve as potential biomarkers, with implications for both diagnosis and long-term monitoring of sleep apnea patients.

2) Differentiation of Apnea and Non-Apnea Periods: The use of time-frequency analysis enabled a clear distinction between apnea and regular breathing periods, suggesting that this method could be implemented in real-time monitoring systems to detect apnea events as they occur.

3) Clinical Relevance: The observed changes in ECG and SpO2 signals provide insights into the cardiovascular and respiratory responses to apnea. These findings support existing research that links sleep apnea with cardiovascular conditions, particularly those related to autonomic dysregulation.

4) Implications for Future Research: This study emphasizes the importance of multi-signal analysis in understanding the full scope of sleep apnea's impact on physiological systems. Future studies may benefit from integrating additional bio signals, such as EEG or respiratory effort signals, to capture a broader range of physiological responses.

Future Directions

Further research is needed to validate these findings in larger patient populations and to explore the potential of integrating these time-frequency analysis methods into clinical practice. Specifically, future studies could focus on the development of machine learning algorithms that use these biomarkers to automatically classify apnea severity and predict cardiovascular risk. The integration of multiple signal types could also provide a more comprehensive assessment of sleep apnea, leading to better diagnosis, treatment, and management of this condition.

Additionally, the application of advanced time-frequency analysis techniques, such as wavelet transforms, could enhance the resolution of transient features in the data, potentially revealing additional biomarkers that could improve the accuracy of automated apnea detection systems. Finally, real-time implementation of these techniques in wearable devices or remote monitoring systems could provide continuous apnea monitoring, benefiting both patients and healthcare providers in managing this chronic condition.

REFERENCES

[1] L. M. Donovan and V. K. Kapur, "Prevalence and Characteristics of Central Compared to Obstructive Sleep Apnea: Analyses from the Sleep Heart Health Study Cohort," Sleep, vol. 39, no. 7, pp. 1353-1359, 2016, doi: 10.5665/sleep.5962.

[2] K. A. Franklin and E. Lindberg, "Obstructive sleep apnea is a common disorder in the population—a review on the epidemiology of sleep apnea," Journal of Thoracic Disease, vol. 7, no. 8, pp. 1311-1322, 2015. [Online]. Available: https://jtd.amegroups.org/article/view/4797.

[3] K. Banno and M. H. Kryger, "Sleep apnea: Clinical investigations in humans," Sleep Medicine, vol. 8, no. 4, pp. 400-426, 2007/06/01/ 2007, doi: https://doi.org/10.1016/j.sleep.2007.03.003.

[4] T. M. Rutkowski, "Data–Driven Multimodal Sleep Apnea Events Detection," Journal of Medical Systems, vol. 40, no. 7, p. 162, 2016/05/18 2016, doi: 10.1007/s10916-016-0520-7.

[5] T. H. Tran, Z. M. Yuldashev, and E. V. Sadykova, "A multilevel system for diagnosis of sleep apnea," in 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 29 Jan.-1 Feb. 2018 2018, pp. 1235-1238, doi: 10.1109/EIConRus.2018.8317316.

[6] M. Bahrami and M. Forouzanfar, "Sleep Apnea Detection From Single-Lead ECG: A Comprehensive Analysis of Machine Learning and Deep Learning Algorithms," IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-11, 2022, doi: 10.1109/TIM.2022.3151947.

[7] D. S. Morillo, J. L. R. Ojeda, L. F. C. Foix, D. B. Rendon, and A. Leon, "Monitoring and Analysis of Cardio Respiratory and Snoring Signals by using an Accelerometer," in 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 22-26 Aug. 2007 2007, pp. 3942-3945, doi: 10.1109/IEMBS.2007.4353196.

[8] P. Moridian et al., "Automatic diagnosis of sleep apnea from biomedical signals using artificial intelligence techniques: Methods, challenges, and future works," WIREs Data Mining and Knowledge Discovery, vol. 12, no. 6, p. e1478, 2022, doi: https://doi.org/10.1002/widm.1478.

[9] R. Atri and M. Mohebbi, "Obstructive sleep apnea detection using spectrum and bispectrum analysis of single-lead ECG signal," Physiological Measurement, vol. 36, no. 9, p. 1963, 2015/08/10 2015, doi: 10.1088/0967-3334/36/9/1963.

[10] I. D. Castro, C. Varon, T. Torfs, S. Van Huffel, R. Puers, and C. Van Hoof, "Evaluation of a Multichannel Non-Contact ECG System and Signal Quality Algorithms for Sleep Apnea Detection and Monitoring," Sensors, vol. 18, no. 2, p. 577, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/2/577.

[11] A. R. Hassan and M. A. Haque, "An expert system for automated identification of obstructive sleep apnea from single-lead ECG using random under sampling boosting," Neurocomputing, vol. 235, pp. 122-130, 2017/04/26/ 2017, doi: https://doi.org/10.1016/j.neucom.2016.12.062.

[12] G. Korompili et al., "PSG-Audio, a scored polysomnography dataset with simultaneous audio recordings for sleep apnea studies," Scientific Data, vol. 8, no. 1, p. 197, 2021/08/03 2021, doi: 10.1038/s41597-021-00977-w.

[13]   A. Q. Javaid, C. M. Noble, R. Rosenberg, and M. A. Weitnauer, "Towards detection of sleep apnea events by combining different non-contact measurement modalities," in 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 16-20 Aug. 2016 2016, pp. 5307-5310, doi: 10.1109/EMBC.2016.7591925.

[14]   S. D. Bank. PSG-Audio Dataset, doi: https://doi.org/10.11922/sciencedb.00345.

[15]   H. E. Romero, N. Ma, G. J. Brown, and E. A. Hill, "Acoustic Screening for Obstructive Sleep Apnea in Home Environments Based on Deep Neural Networks," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 7, pp. 2941-2950, 2022, doi: 10.1109/JBHI.2022.3154719.

[16]   Y. Huang, L. Chen, and Q. Huang, "Fine-Grained Detection of Apnea-Hypopnea Events Based on Transformer Network in Audio Recordings," in 2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP), 21-23 April 2023 2023, pp. 580-585, doi: 10.1109/ICSP58490.2023.10248848.

[17]   G. SharanYadav, S. Yadav, and P. Prachi, "Time and Frequency Exploration of ECG Signal," International Journal of Computer Applications, vol. 67, pp. 5-8, 2013.

[18]   S. Kansal, P. P. Bansod, and A. Kumar, "Statistical Approach for Determination of ECG Markers," in 2015 International Conference on Computational Intelligence and Communication Networks (CICN), 12-14 Dec. 2015 2015, pp. 446-451, doi: 10.1109/CICN.2015.93.

[19]   L. Song and F. Yu, "The Time-Frequency Analysis of Abnormal ECG Signals," in Life System Modeling and Intelligent Computing, Berlin, Heidelberg, K. Li, L. Jia, X. Sun, M. Fei, and G. W. Irwin, Eds., 2010// 2010: Springer Berlin Heidelberg, pp. 60-66.

[20]   A. F. Hussein, S. J. Hashim, A. F. A. Aziz, F. Z. Rokhani, and W. A. W. Adnan, "Performance Evaluation of Time-Frequency Distributions for ECG Signal Analysis," Journal of Medical Systems, vol. 42, no. 1, p. 15, 2017/11/29 2017, doi: 10.1007/s10916-017-0871-8.

[21]   G. C. Lui, D. Wu, K. W. Cheung, H. F. Ma, and K. Y. Szeto, "Time warping of apneic ECG signals using genetic algorithm," in 2016 IEEE Congress on Evolutionary Computation (CEC), 24-29 July 2016 2016, pp. 178-184, doi: 10.1109/CEC.2016.7743793.

RESUME

Göksu DEMİRCİ

Göksu Demirci is a graduate student at the Department of Computer Engineering, Faculty of Technology, Selcuk University. He completed his bachelor's degree in mechanical engineering at Akdeniz University in 2016 and is currently pursuing his master's degree in Computer Engineering at Selcuk University. Göksu specializes in biomedical signal processing, focusing on the analysis and interpretation of physiological signals commonly used in medical diagnostics. His education includes feature extraction, filtering techniques, and the development of AI-based models for classification and disease diagnosis, particularly in applications such as sleep apnea detection. Additionally, he has gained expertise in machine learning, artificial intelligence, embedded systems, and computational intelligence through his coursework and projects.

Assoc. Prof. Dr. Adem GÖLCÜK

Adem Golcuk is an Associate Professor at the Department of Computer Engineering, Faculty of Technology, Selcuk University. He completed his master's degree in 2010 and his doctorate in 2017 at Selcuk University. Golcuk has published articles in various national and international journals on topics such as biomedical devices, embedded systems, machine learning and image processing. His main research interests include embedded systems, RF communication, machine learning and sensor data analysis.

- 165 -

A SIMPLIFIED RULE-BASED EXPERT SYSTEM DESIGN FOR EVALUATION OF NEPHRITIS RELATED PATIENT SYMPTOMS

Fatih BAŞÇİFTÇİ[1], Kamil Aykutalp GÜNDÜZ[2], Hüseyin Salih ÖZGÖKÇE[3]

INTRODUCTION

Nephritis is the inflammation of glomeruli, or blood thinning filters, in the filtration unit called nephron in our kidneys. The general structure of Nephron is shown in Fig 1. The task of fine filters, called glomeruli, is to filter the blood and remove excess fluids, electrolytes, and wastes from the body. Nephritis emerges in two forms, acute and chronic. Acute nephritis starts suddenly and is short term but requires serious treatment. Chronic nephritis gradually evolves and gradually worsens because of a time period that can last for months. However, it is a long-term disease [1-4].
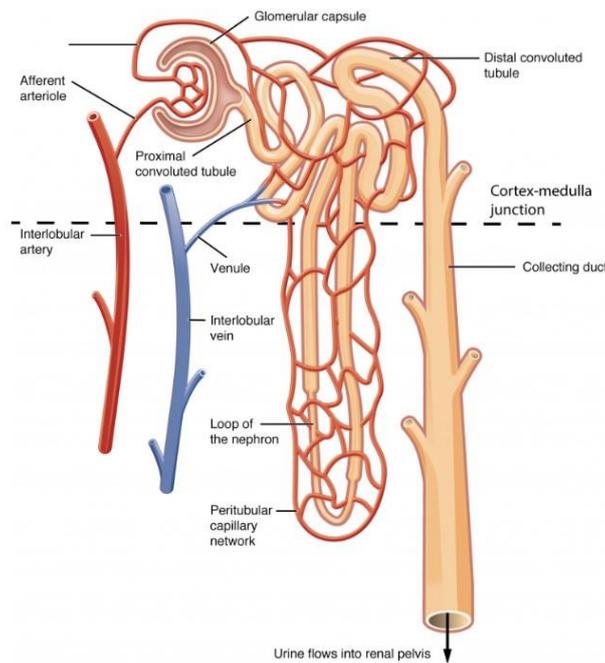
Figure 1: Nefron's General Construction [2].

Nephritis changes glomerular build up in the kidneys by inflammatory cell proliferation. This leads to glomerular damage in the kidney. Damaged glomeruli cannot function and cause a decrease in blood flow. As a result, urine output also decreases. In other words, this condition,

[1] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0000-0003-1679-7416, basciftci@selcuk.edu.tr
[2] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0000-0002-2290-5447, aykutalp@selcuk.edu.tr
[3] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0009-0001-1762-0084, ozgokcesalih@gmail.com

called oliguria, causes the urine to be removed from the body less than normal. Thus, waste products called urea are kept in the body and stored in the blood, which leads to the disease we call uremia. Therefore, the red blood cells infiltrate of the injured glomeruli, resulting in the appearance of blood in the urine that we call hematuria. Reduction of blood flow activates the hormonal renin-angiotensin-aldosterone system (RAAS), which controls blood pressure and fluid balance. In this case, the body's water retention and mild hypertension occurs as a natural result. In cases where RAAS is too active, the blood pressure will also be high. When the glomeruli are damaged, the proteins cannot function to prevent the passage of urine and to keep them in the blood vessels. In this case, much more protein will lead to the urine than normal, which is known as proteinuria. Loss of essential proteins in the body because of nephritis may cause many vital symptoms in the future. The most serious complication caused by nephritis is the loss of proteins that protect blood against excessive clotting. Loss of these proteins can trigger strokes that result from sudden clotting of the blood [3, 5].

Another renal disease, Tubulointerstitial Nephritis (TIN), is known for tubulo-interstitial edema, fibrosis, atrophy and infiltration of inflammatory cells. It occurs because of tubular damage. TIN can be observed both acutely and chronically [6].

Chronic interstitial nephritis, which occurs in about 15-30% in our country, has indications such as tuberous atrophy and dilatation, tubulo-intersis fibrosis, infiltration of mononuclear cells. It is among the common causes of chronic kidney disease. The course of the disease continues with progressive loss of kidney function. [7].

Henoch-Schönlein purpura (HSP) is a common disease in children 5-10 years of age. HSP is a systemic vascular inflammation that usually manifests as skin, joint, gastrointestinal tract, and kidney involvement. Nephritis is an important stage that is likely to be encountered in the long-term course of HSP disease. Since death rates are basically related to glomerulonephritis, kidney involvement should be monitored for a long time [8].

The purpose of this study is to decide nephritis (kidney inflammation) disease, by evaluating seven different risk factors specific to the disease. For this purpose, a Medical Expert System (MES) has been developed. Using scientific articles, the seven identified symptoms were converted to binary values according to certain conditions (eg, 1 if the fever > 38, 0 else). Using this MES created to help doctors make decisions, (27) 128 different rules were obtained. 128 different rules created By Boolean Functions Simplification method, reduced to 6 rules covering all the rules. The MES rule base was established with reduced rules. This rule base was evaluated on the data of 120 patients.

MATERIAL AND METHOD

Nephritis (Inflammation of the Kidney) Disease

Nephritis occurs when the body temperature is 38 degrees and above. It is usually checked by examining risk factors such as gastroenteritis, back pain, urinary pressure, mass pain, urethral burn, itching, swelling, urinary incontinence [9].

Dataset

In this work, the database named "Acute Inflammations Data Set" consisting of data from 120 potential patients was used. This database was created for the paper "Application of rough sets in the presumptive diagnosis of urinary system diseases" presented by Czerniak (2002) on Artificial Intelligence and Security Conference (ACS' 2002) in 9th International Information Systems [9,10].

Expert Systems

The first examples of Expert Systems (ES) initiated at Stanford University in the 1970's. According to Prof. Dr. Edward Feigenbaum from Stanford University, ES is a clever software that realizes the solutions of the hard problems which can be solved by an expert who has the knowledge and required ability to cope with, by modeling the expert's information and logical derivation mechanisms [11-13]. Figure 2 shows the common structure of the ES. It basically consists of a database, rule base and parser.
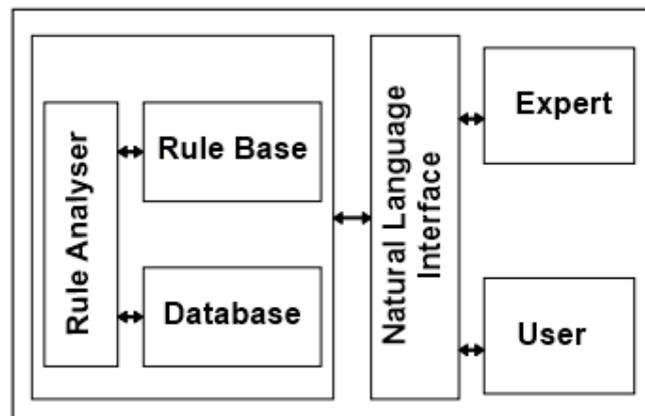
Figure 2: Common Structure of ES [14].

ES has application areas in many fields such as medical field diagnosis and treatment planning, financial area analysis, industrial planning, automation, debugging in operating systems, sales consultancy [15-17].

When the expert system was developed in the early period, LISP and PROLOG, which are artificial intelligence programming languages (AI), were generally used. Today, Expert System can be coded in traditional languages (TL) such as C, C ++, C #, Visual Basic. The main difference between AI and TL is that AI can provide indefinite outputs by working with indefinite data. AI works with symbolic and descriptive rules, relational knowledge bases and inference mechanisms [16].

The first widely developed example of expert systems is the MYCIN named system at Stanford University which is written in LISP. This system is used to diagnose diseases caused by meningitis and bacteria in the blood. An experiment was conducted with ten specialist doctors to compare the success of MYCIN in diagnosis and treatment. As a result of the experiment, it

was observed that the recommendations of MYCIN were 65% accurate while the success of the diagnosis of specialist doctors varied between 42.5% and 62.5% [11].

INTERNIST was designed in 1970 for the diagnosis of internal diseases. The patient's story, laboratory results and clinical appearance are entered into the system. The system identifies possible diagnoses according to the data. As a result, it chooses the most appropriate diagnosis for the disease and presents it to the doctor. The Current Disease Program (CDP) is designed for the diagnosis of kidney failure disease. CDP forms diagnostic recommendations by using 36 findings of kidney failure disease, and thus models doctors' reasoning. Another system called PHARMA-2 provides decision support to doctors in the treatment of medicines with a database based on evaluation techniques of pharmacologists [18].

MOLGEN, the most well-known expert system in the field of genetics, offers decision support to molecular genetic specialists in DNA material replacement and redesign. The Heart Disease Program (HDP) is the largest expert system in the diagnosis of heart attack. When laboratory tests and age information are entered into the system by the doctor, it can determine the type of heart attack [19].

Isabel is a pediatric decision support tool used by doctors in hospitals and clinics in England. The system can only be used for pediatric diseases. When the patient's symptoms are entered into the system, the system offers recommendations to the doctor by determining the most appropriate diagnosis from the possible diagnoses related to the system [20].

The ES called DXplain, which was started to be used in Massachusetts General Hospital in 1987, was able to diagnose 2000 diseases with 4700 diagnoses. Users are able to query the likelihood of a particular disease appearing in a disease in the system and the frequency of any disease in the presence of a disease [21].

As of 1989, it is estimated that 2200 systems were developed in the United States, 1200 systems were being developed in Europe, and 85 systems were used in total. The main reason for the rapid expansion of expert systems is the increase in the quality of the decisions made and, at least, the cost reduction [15].

Simplification of Boolean Functions

Providing equivalent logic function by reducing the number of elements of logic functions used in logic circuits is an important engineering problem. For this reason, expressions are simplified by applying Boolean rules to find the function that has the same value as the logical function that is realized by at least logical function element [22].

For example, when the Boolean algebraic rules are applied to the function $F(x, y, z) = x.y.z + x.\bar{y}.z$ expressed as the sum of products, $F(x, y, z) = (y + \bar{y})(xz) = 1.(xz) = xz$ is obtained [22].

As will be seen, simplification of the boolean functions will lead to the development of improved software and circuits from cost, time and other important aspects [23].

APPLICATION

In this study, an MES was designed according to the symptoms of disease for the diagnosis of nephritis. While the MES rule base was established, the rules in Nephrit diagnostics were reduced to fewer rules by reevaluating boolean functions with simplification techniques to minimize adverse effects on system and evaluation performance. It was aimed to determine whether the patients were suffering from nephritis with the help of the MES created using reduced rules. The accuracy of the system was determined by evaluating the MES by the Acute Inflammations Data Set, which contains 120 patient data generated by Jacek Czerniak from the Polish Academy of Sciences.

Seven different risk factors were used in the MES rule base for the diagnosis of nephritis. Using these risk factors, 7 input symbols and 1 exit symbol were created. These generated symbols are shown in Table 1.

Table 1. Input and Output Values

Input No	Input Symbol	Input Name	Output No	Output Symbol	Output Name
1	G1	Fever	1	Ç1	Nephritis
2	G2	Nausea			
3	G3	Lumbar Pain			
4	G4	Urinate Frequent			
5	G5	Painful Urination			
6	G6	Burning of the urethra, itching, swelling			
7	G7	Inflammation of the Urinary Bladder			

Patient data for 7 entries and 1 output symbol generated within the data obtained from the studies for nephritis were reanalyzed according to the presence of fever at 38º C and above. The analyzed data and symbols are shown in Table 2.

Table 2. Redefining the Input and Output Values by Fever

No	Input and Output Symbol	Definition
1	G1	If the fever is 38 and above
2	G2	If you have nausea or not
3	G3	If you have Lumbar Pain or not
4	G4	If you have Urinate Frequent or not
5	G5	If you have Painful Urination or not
6	G6	If you have Burning of the urethra, itching, swelling or not
7	G7	If you have Inflammation of the Urinary Bladder or not
8	Ç1	If you have Nephritis or not

7 entries and 1 exit symbols used for diagnosis of nephritis were converted to values in the binary (0,1) number system according to the patient's fever state. The symbols of the disease risk factors, the actual values of entry and exit and the corresponding values according to the binary number system are shown in Table 3. According to their actual values, if the conditions that determine the risk factors of the disease are present, they are indicated as "yes" in the Table

3, otherwise "no" statement is written on the related cell of the Table. According to the boolean functions, if the conditions are specified, it is indicated as 1 in the binary number system and 0 otherwise.

Table 3. Data by Real and Boolean Functions

G1	G2	G3	G4	G5	G6	G7
(a) Real values						
35,5	no	yes	no	no	no	no
37,9	no	no	yes	no	no	yes
38	no	yes	yes	no	yes	no
(b) Boolean Functions						
0	0	1	0	0	0	0
0	0	0	1	0	0	1
1	0	1	1	0	1	0

128 different rules were produced from 7 risk factors included in the study. According to Boolean functions, a truth table of 1's and 0's is created with rules consisting of 7 inputs and 1 output. In this truth table, an output state is determined for each state. The input and output symbols forming the precision table with binary numerical values are presented in Table 4.

Table 4. Truth table

No	G1	G2	G3	G4	G5	G6	G7	Ç1
1	0	0	0	0	0	0	0	0
2	0	0	0	1	0	0	1	0
3	0	0	0	0	1	1	0	0
…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…
127	1	1	1	1	1	0	0	0
128	1	1	1	1	1	1	1	1

128 different Boolean functions handled in the truth table were reduced to 6 cases by simplification. In this way, instead of controlling 128 different states, 6 states obtained by simplification will be checked. These situations, which are obtained through reduction, cover all of the 128 cases. The number of Boolean functions obtained as a result of the simplification process improves the efficiency of MES as well as making identification faster and more efficient. In this way diagnosis will be made by evaluating all possible situations. The states obtained as a result of the simplification process are shown in Table 5.

Table 5. Simplification table

Output Symbol	Output States						
	G1	G2	G3	G4	G5	G6	G7
Y1	1	1	-	-	-	-	-
Y1	1	-	1	-	-	-	-
Y1	1	-	-	1	-	-	-
Y1	1	-	-	-	1	-	-
Y1	1	-	-	-	-	1	-
Y1	1	-	-	-	-	-	1

As seen in Table 5, the rule base for the MES prepared for the disease diagnosis were established with the reduced states obtained by the Boolean Function Simplification method. With the help of this established rule base, all possibilities that may arise in relation to nephritis have been evaluated. As a result, the rule base created by Boolean Function Simplification technique has been reduced to 6 rules including all rules. The symbol "1" used in Table 5 indicates the symptom found in the patient. The symbol "-" indicates that the indication is not important in the patient. A rule base is established according to two of the reduced states of MES in Table 5 (1 - - 1 - - -, 1 - - - - 1 -) and is shown in Table 6. When the MES rule base is created, the "IF condition THEN action" structure is used.

Tablo 6. Creation of MES rule base

Symbolic Display	Display with Normal Values	Display with Binary Values
IF G1 **AND** G4 **THEN** Ç1	**IF** Fever $\geq$ 38 **AND** Urinate Frequent=Available **AND** **THEN** Nephritis = Available	**IF** Fever = 1 **AND** Urinate Frequent = 1 **AND** **THEN** Nephritis = 1
IF G1 **AND** G6 **THEN** Ç1	**THEN** Fever $\geq$ 38 **AND** Burning of the urethra, itching, swelling = Var **THEN** Nephritis = Available	**THEN** Fever = 1 **AND** Burning of the urethra, itching, swelling = 1 **THEN** Nephritis = 1

In Table 6, MES rule base generation with normal values and binary values is shown for the two cases. The rule base created by the binary values exemplified in Table 6 is used in the designed MES. Here the patient data is converted into binary values and made available for the rule base. In Table 7, there are 5 samples showing patient data showing binary values. In Table 6, MES rule base generation with normal values and binary values is shown for the two cases. The rule base created by the binary values exemplified in Table 6 is used in the designed MES. Here the patient data is converted into binary values and made available for the rule base. In Table 7, there are 5 samples showing patient data showing binary values.

Table 7. Display of real patient data with binary values

Real Patient Data						
G1	G2	G3	G4	G5	G6	G7
36,8	no	no	yes	yes	yes	yes
37,3	no	yes	no	no	no	no
38	no	yes	yes	no	no	yes
40	no	no	no	no	no	no
41,1	yes	yes	yes	yes	yes	yes

Binary Values						
G1	G2	G3	G4	G5	G6	G7
0	0	0	1	1	1	1
0	1	0	0	0	0	0
1	0	1	1	0	0	1
1	0	0	0	0	0	0
1	1	1	1	1	1	1

As the examples in Table 7 show, the data in the Acute Inflammations Data Set from the Polish Academy of Sciences, created by Jacek Czerniak, have been converted into binary values. The actual patient data converted was evaluated with the designed MES. The accuracy of MES was determined, and the results are shown in Table 8.

Table 8. Accuracy Value of Developed MES

Application Values	Number of Data	Accuracy Rate
Inflammation of the Urinary Bladder	50	100%
Non-Inflammation of the Urinary Bladder	70	100%
Total number of patients	120	
Overall Accuracy Rate (%)		100%

The accuracy rate of the generated MES was evaluated based on the actual patient data in the database consisting of 120 persons and presented in Table 8. As a result of the evaluation, the number of Nephritis patients was determined as 50 persons with MES rule base. This number is equal to the patient data in the database. The accuracy of the developed MES was determined as 100% success.

In Keles' (2008) study, an ES designed to assist doctors in the thyroid disease diagnosis called Expert System for Thyroid Disease (ESTDD) was designed. The system is built using Microsoft .Net software architecture and SQL Server 2005 database. The neuro-fuzzy method was used in the diagnosis of the diseases. ESTDD can diagnose thyroid disease with an accuracy of 95.33%. ESTDD benefits from 5 findings of thyroid disease when diagnosed [24].

A strong rule base has been devised in the study of Keles (2010) for the breast cancer diagnosis. The NEFCLASS named vehicle was used for the rule base based on the neuro-fuzzy method. There are 9 rules and 3 findings for diagnosis. The positive predictive value of the rule base was 75% and the negative predictive value was 93%. As a result of the study, an ES called Ex-DBC was introduced [25].

Hiesmayr (1993) designed a KBWean named ES for changes in the ventilation parameters of patients who were mechanically oxygenated in the intensive care unit after surgery for heart and chest infarction. In the experiment with 10 randomly selected patients, ES reacted 131 minutes ago according to the physician in intensive care unit and offered recommendations for frequent and very small changes in the mechanical ventilation system settings [26].

In Prabhudesai (2008), an ES was developed to assist doctors in the diagnosis of acute appendicitis. The data from a training hospital were tested with data from 60 patients using artificial neural networks technique. Symptoms, hematological evaluations, age and gender data were used for diagnosis. When the system is applied to the data, the positive predictive value reaches 96% and the negative predictive value reaches 100% [27].

In order to analyze the hematology laboratory tests, a specialist system with web interface Fuzzy input was designed by Başçiftçi (2011). 816 patients from the Selcuk University Faculty of Medicine laboratory were entered into the ES. As a result of the study, it was seen that the ES evaluated 92% success rate in females, 91% in males and 92% in general [28].

In the study of Kingsland (1982), a ES called AI / RHEUM was developed to help doctors diagnose rheumatic diseases. AI / RHEUM can diagnose 26 different diseases. The system has 94% successful diagnoses on 384 real patient data [29].

In the study of Ivandic (1996); A computerized physician decision support system for the determination of the amount of cretinin, total protein, albumin, microglobulin, macroglobulin and glucosaminidase in urea has been developed to diagnose Proteinuria, hematuria and leukocytosis diseases with the expert system called Urea Protein Discrimination Expert System (UPES). In the study, a database consisting of 129 urea analyzes of 94 patients was evaluated with the established rulebase, and 98% of the patients were found to be suitable for the diagnosis of 4 specialists [30].

Expert Systems for Thyroid Disease Diagnosis:

2021 - "A Hybrid Intelligent System for Thyroid Disease Diagnosis". A novel expert system proposal based on machine learning and fuzzy logic [31].

2022 - "THYROIDEX: An Intelligent System for Automated Thyroid Disease Diagnosis" [32]. A web-based expert system using deep learning techniques

Expert Systems for Breast Cancer Diagnosis:

2020 - "An Expert System for Early Breast Cancer Diagnosis" [33]. Proposal for a sensitive rule-based expert system

2021 - "BRCAEX: An AI-Powered Breast Cancer Risk Assessment Tool" [34]. An expert system for risk assessment using genetic data

Expert Systems for Monitoring Respiratory Parameters:

2021 - "A Knowledge-Based System for Mechanical Ventilation Management" [35]. Intelligent decision support system for intensive care patients

2022 - "VENTEX: An Expert System for Optimizing Ventilator Settings" [36]. An expert system using machine learning and rule-based approach

Expert Systems for Acute Appendicitis Diagnosis: 2020 - "APPENDICITIS: An AI-Assisted Diagnostic Tool for Acute Appendicitis" [37]. A deep learning based system using symptoms and imaging data

2021 - "ExpertAppend: An Expert System for Rapid Appendicitis Diagnosis" [38]. A new system that combines rule-based and machine learning approaches Current Expert Systems for Other Medical Fields:

2022 - "DiabetEx: An Intelligent System for Diabetes Management" [39]. Intelligent decision support system for diabetes patients

2023 - "CardioNet: An Expert System for Cardiovascular Disease Diagnosis" [40].

A new expert system used in the diagnosis of heart diseases These examples reflect current studies and post-2020 publications similar to the given medical expert systems. Research in the field of intelligent decision support systems in the medical field is progressing rapidly.

CONCLUSION

The main objective of this study is to develop and evaluate a simplified rule-based expert system for use in the medical field. In contrast to the complex expert systems frequently used in the existing literature, this study adopts a simpler and more user-friendly approach.

In the first phase of the study, 7 risk factors that play the most important role in the diagnosis of the target disease were identified through interviews with expert physicians. Then, an expert system of 128 rules was designed based on these 7 risk factors. However, it was anticipated that this complex rule set would be difficult to be used effectively by doctors in practice. Therefore, in the second stage, a simpler and more understandable decision support model was developed by reducing the number of rules in the system.

In order to evaluate the performance of the designed simplified expert system, 120 patient data obtained from the Acute Infections Dataset were used. The system is tested for its success in making the correct diagnosis for each patient. The results show that the proposed simplified expert system has an accuracy rate of 100%.

This high performance demonstrates that the designed system can be effectively used in clinical applications. Indeed, this system, which enables earlier, faster and more accurate diagnosis, may also contribute to reducing the negative health outcomes of patients. As a result, this study is an important contribution to the development of medical decision support systems.

REFERENCES

[1] http://www.ilacpedia.com/hastaliklar/nefrit-glomerulit, Available: 24 January 2016.

[2] Arıcan, A., 2011, Nefritler, http://www.adnanarican.com/2011/11/nefritler, Available: 5 December 2016.

[3] https://tr.wikipedia.org/wiki/Nefrit, Available: 24 January 2016.

[4] Akpolat, T., 2016, Nefrit, http://tekinakpolat.com/wp-content/uploads/2016/11/nefrit.pdf, Available: 6 December 2016].

[5] Klinik Biyokimya Uzmanları Derneği, Proteinüri, http://labtestsonline.org.tr/understanding/conditions/proteinuria, Available: 4 January 2017].

[6] Parmaksız, G., Cengiz, N., Canpolat, T., Baskın, E., Noyan, A., 2012, Kronik Tübülointerstisyel Nefrit: Vaka Sunumu, Çocuk Dergisi, 12 (4), 186-188.

[7] Kaynar, K., 2014, Kronik İnterstisyel Nefrit: Tanım, Patogenez, Epidemiyoloji, Patoloji, Klinik Tablo ve Tedavi, Türkiye Klinikleri J Neprol-Special Topics, 7(3), 15-18.

[8] Ersen, A., Aydınöz, S., Karademir, F., Süleymanoğlu, S., Meral, C., Özkaya, H., Göçmen, İ., 2009, Çocukluk Dönemi Henoch-Schönlein Purpurası: 42 Olgunun Retrospektif Analizi, Dirim Tıp Gazetesi, 84(2), 35-41.

[9] Czerniak, J., Zarzycki, H., 2003, Application of rough sets in the presumptive diagnosis of urinary system diseases, Artifical Inteligence and Security in Computing Systems, ACS'2002 9th International Conference Proceedings, Kluwer Academic Publishers, 2003, 41-51.

[10] Czerniak, J., 2009, Acute Inflammations Data Set [online], Machine Learning Repository Center for Machine Learning and Intelligent Systems, Available: http://archive.ics.uci.edu/ml/datasets/Acute+Inflammations, 11 January 2017.

[11] Doğaç, A., 1990, Uzman Sistemler, TMMOB EMO Elektrik Mühendisliği, 373, 87-91.

[12] Başçiftçi, F., İncekara, H., 2011, Web based medical decision support system application of Coronary Heart Disease diagnosis with Boolean functions minimization method, Expert Systems with Applications, 38 (11), 14037-14043.

[13] Demirhan, A., Kılıç, Y., Güler, İ., 2010, Tıpta Yapay Zeka Uygulamaları, Yoğun Bakım Dergisi, 9 (1), 31-41.

[14] Serhatlıoğlu, S., Hardalaç, F., 2009, Yapay Zeka Teknikleri ve Radyolojiye Uygulanması, Fırat Tıp Dergisi, 14 (1), 01-06.

[15] Bilginoğlu, F., 1993, İşletmenin Karar Sürecini Destekleyen Uzman Sistemler, Yönetim Dergisi, 15, 5-11.

[16] Bilge, U., 2007, Tıpta Yapay Zeka ve Uzman Sistemler, 4. Tıp Bilişimi Kongresi, Antalya, 1-3 Kasım, 108-112.

[17]   İçen, U., Günay, S., 2014, Uzman Sistemler ve İstatistik, İstatistik ve Aktüerya, 7, 37-45.

[18]   Canvar, T., 2002, İlaç Endüstrisinde Uzman Sistemler ve Uygulamaları, SAU Fen Bilimleri Enstitüsü Dergisi, 6 (1), 147-154.

[19]   Karim, A., Çelebi, F., Mohammed, A., 2016, Software Development for Blood Disease Expert System, Lecture Notes on Software Engineering, 4 (3), 179-183.

[20]   Özata, M., Aslan, Ş., 2004, Klinik Karar Destek Sistemleri ve Örnek Uygulamalar, Kocatepe Tıp Dergisi, 5 (2), 11-17.

[21]   Waxman, H., Worley, W., Computer-Assisted Adult Medical Diagnosis: Subject Review and Evaluation of a New Microcomputer-Based System, Medicine, 69 (3), 125-136.

[22]   Çölkesen, R., 2012, Bilgisayar Mühendisliğine Giriş, Papatya Yayıncılık, 44-56.

[23]   Başçiftçi, F., 2007, Simplification of Single-Output Boolean Functions by Exact Direct Cover Algorithm Based on Cube Algebra, EUROCON 2007 The International Conference on "Computer as a Tool", Warshaw September 9-12, 427-431.

[24]   Keleş, A., Keleş, A., 2008, ESTDD: Expert system for thyroid diseases diagnosis, Expert Systems with Applications, 34 (1), 242-246.

[25]   Keleş, A., Keleş, A., 2010, Extracting fuzzy rules for the diagnosis of breast cancer, Turkish Journal of Electrical Engineering and Computer Sciences, 2013 (21), 1495-1503.

[26]   Hiesmayr, M., Gamper, J., Neugebauer, T., Mares, P., Adlassnig, K., Haider, W., 1993, Clinical application of patient data management systems (PDMS): computer-assisted weaning from artificial ventilation (KBWEAN), Patient Data Management in Intensive Care, 4 (6), 129-138.

[27]   Prabhudesai, S., Gould, S., Rekhraj, S., Tekkis, P., Glazer, G., Ziprin, P., 2008, Artificial neural networks: useful aid in diagnosing acute appendicitis, World Journal of Surgery, 32 (2), 305-309.

[28]   Başçiftçi, F., İncekara, H., 2011, Web Ara Yüzü ile Hematoloji Laboratuarı Tahlillerinin Değerlendirilmesi için Bulanık Girişli Uzman Sistem Tasarımı, SDÜ Fen Bilimleri Enstitüsü Dergisi, 15 (1), 51-55.

[29]   Kingsland, L., Sharp, G., Kay, D., Weiss, S., Roeseler, G., Lindberg, D., 1982, An Expert Consultant System in Rheumatology: AI/RHEUM, Proceedings of the Annual Symposium on Computer Application in Medical Care, 748–752.

[30]   Ivandic, M., Hofmann, W., Guder, W., 1996, Development and Evolution of A Urine Protein Expert System, Clinical Chemistry, 42 (8), 1214-1222.

[31]   Li, Z., Qin, J., Zhang, X., Wan, Y. (2019). A Hybrid Intelligent Framework for Thyroid Diagnosis. In: Ning, H. (eds) Cyberspace Data and Intelligence, and Cyber-Living, Syndrome, and Health. CyberDI CyberLife 2019 2019. Communications in Computer and Information Science, vol 1138. Springer, Singapore. https://doi.org/10.1007/978-981-15-1925-3_32

[32]   Vasile CM, Udriștoiu AL, Ghenea AE, Popescu M, Gheonea C, Niculescu CE, Ungureanu AM, Udriștoiu Ș, Drocaş AI, Gruionu LG, Gruionu G, Iacob AV, Alexandru DO. Intelligent Diagnosis of Thyroid Ultrasound Imaging Using an Ensemble of Deep Learning Methods. Medicina (Kaunas). 2021 Apr 19;57(4):395. doi: 10.3390/medicina57040395. PMID: 33921597; PMCID: PMC8073676.

[33]   Aswathi, A & Antony, Anil. (2018). An Intelligent System for Thyroid Disease Classification and Diagnosis. 1261-1264. 10.1109/ICICCT.2018.8473349.

[34]   Liew, X.Y.; Hameed, N.; Clos, J. A Review of Computer-Aided Expert Systems for Breast Cancer Diagnosis. Cancers 2021, 13, 2764. https://doi.org/10.3390/cancers13112764

[35]   Dojat M, Brochard L, Lemaire F, Harf A. A knowledge-based system for assisted ventilation of patients in intensive care units. Int J Clin Monit Comput. 1992 Dec;9(4):239-50. doi: 10.1007/BF01133619. PMID: 1484275.

[36]   Bianchi, V., Giambusso, M., De Iacob, A. et al. Artificial intelligence in the diagnosis and treatment of acute appendicitis: a narrative review. Updates Surg 76, 783–792 (2024). https://doi.org/10.1007/s13304-024-01801-x.

[37]   Issaiy M, Zarei D, Saghazadeh A. Artificial Intelligence and Acute Appendicitis: A Systematic Review of Diagnostic and Prognostic Models. World J Emerg Surg. 2023 Dec 19;18(1):59. doi: 10.1186/s13017-023-00527-2. PMID: 38114983; PMCID: PMC10729387.

[38]   Guan Z, Li H, Liu R, Cai C, Liu Y, Li J, Wang X, Huang S, Wu L, Liu D, Yu S, Wang Z, Shu J, Hou X, Yang X, Jia W, Sheng B. Artificial intelligence in diabetes management: Advancements, opportunities, and challenges. Cell Rep Med. 2023 Oct 17;4(10):101213. doi: 10.1016/j.xcrm.2023.101213. Epub 2023 Oct 2. PMID: 37788667; PMCID: PMC10591058.

[39]   Ahn, Imjin & Na, Wonjun & Kwon, Osung & Yang, Dong & Park, Gyung-Min & Gwon, Hansle & Kang, Hee & Jeong, Yeon & Yoo, Jungsun & Kim, Yunha & Jun, Tae & Kim, Y.H.. (2021). CardioNet: a manually curated database for artificial intelligence-based research on cardiovascular diseases. BMC Medical Informatics and Decision Making. 21. 10.1186/s12911-021-01392-2.

[40]   Z. Tafa, N. Pervetica and B. Karahoda, "An intelligent system for diabetes prediction," 2015 4th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 2015, pp. 378-382, doi: 10.1109/MECO.2015.7181948.

RESUME

Prof. Dr. Fatih BAŞÇİFTÇİ

Fatih BASÇİFTCİ was born in Konya, Turkey, in 1974. Dr. Basciftci has graduated from the Electronic and Computer Education Department at Marmara University with a B.Sc. degree in 1997 and from the Department of Computer Systems Education at Selcuk University with an M.Sc. degree in 2000. He received a Ph.D. degree in Electric and Electronic Engineering from Selcuk University, Konya in 2006. From 1998 to 2002 he was a research assistant in the Electronic and Computer Education Department of the same university. From 2002 to 2007 he was a lecturer of the same department of Selcuk University, Konya, Turkey. From 2007 to 2012 he was Assist. Professor in Electronic and Computer Education Department of the same university. From 2012 to 2017 he was Assoc. Professor of the Department of Computer Engineering. Since 2017 he has been a Professor of the Department of Computer Engineering. He served as the Head of the Computer Engineering Department between 2015 and 2023. His research interest includes Switching Theory and Computer Architecture on which he has published over 200 papers.

Dr. Kamil Aykutalp GÜNDÜZ

Kamil Aykutalp GÜNDÜZ graduated from Kyrgyzstan-Turkey Manas University, Department of Computer Engineering in 2009. Gündüz completed his master's degree at Kyrgyzstan-Turkey Manas University, Institute of Science and Technology and his PhD degree at Selçuk University, Institute of Science and Technology. Gündüz, who has been working as a lecturer at Selçuk University Kadınhanı Faik İçil Vocational School since 2013, received the title of Dr. in 2023. Gündüz has many national and international scientific articles and publications in the field of Informatics.

Hüseyin Salih ÖZGÖKÇE

Hüseyin Salih ÖZGÖKÇE is an SME Specialist in the Information Technology Department at KOSGEB. He completed his master's degree at Gazi University Informatics Institute. He continues his Ph.D. at Selçuk University Institute of Science. His primary research interests include Information Security and Cryptography, Software and Systems Engineering.

PIONEERING PERSPECTIVES IN MODERN ELECTRICAL AND ELECTRONICS ENGINEERING

AN EVALUATION OF SCHEDULING ALGORITHM IN MPTCP SYSTEMS: A COMPREHENSIVE SURVEY

Atefeh Ahmadniai KHAJEKINI[1], Ali Hakan ULUSOY[2], Enver EVER[3], Mostafa Ayoubi MOBARHAN[4]

INTRODUCTION

In recent years, the number of mobile equipment surpasses the global population. These tools are equipped with multiple wireless interfaces, such as 3-5 G and WiFi systems. However, the standard Transmission Control Protocol (TCP) which is known as a stream network protocol for the Internet is able to work with only one access network per TCP connection. Thus, Multi-path TCP (MPTCP), as a modern extension of standard TCP, can set proper communication between two mobile devices (i.e., end-hosts) over multiple paths [1] [2]. The MPTCP presented by the Internet Engineering Task Force (IETF) to improve the level of reliability in wireless networks. The most common open-source implementation of MPTCP technology is presented in the Linux kernel. Packet scheduling algorithms greatly impact the performance of MPTCP systems, particularly in homogenous environments. If a MPTCP scheduler makes some wrong decisions, it can affect the performance of the network in both the heterogenous and homogenous cases [3]. An inappropriate packet scheduler leads to a decrease in throughput, poor path utilization, and higher download time, while a proper packet scheduler should be able to use the available interfaces to decrease the number of Out of Order (OfO) packets and increase network throughput [4] [5]. In this research paper, we provide a comprehensive overview of MPTCP systems. Then, we theoretically study the most common existing MPTCP schedulers. We also provide a comparison of different MPTCP schedulers. The main features of state-of-the-art MPTCP scheduling algorithm are presented and evaluated. Additionally, we analyze which schedulers achieve better performance in wireless lossy networks [6].

AN OVERVIEW OF THE MPTCP SYSTEMS

MPTCP technology allows multi-homed mobile devices to switch between networks and use several network connections concurrently to boost network throughput. Communication scenarios might be homogeneous or heterogeneous networks [7] [8]. For instance, a homogeneous scenario can be established by using two WiFi interfaces, while a heterogeneous

[1] Eastern Mediterranean University, North Cyprus, via Mersin 10, Turkey, Orcid: https://orcid.org/0009-0001-5002-1999 , atefeh.ahmadniai@emu.edu.tr
[2] Eastern Mediterranean University, North Cyprus, via Mersin 10, Turkey, Orcid: https://orcid.org/0000-0002-2992-9265, alihakan.ulusoy@emu.edu.tr
[3] Middle East Technical University, North Cyprus, via Mersin 10, Turkey, Orcid: https://orcid.org/0000-0001-6516-2770 , eever@metu.edu.tr
[4] Final International University, North Cyprus, via Mersin 10, Turkey, Orcid: https://orcid.org/0000-0002-0298-3384 , mostafa.mobarhan@final.edu.tr

scenario can be generated with one Ethernet, and one WiFi interface or with one LTE, and WiFi connection. Fig. 1 shows the architecture of MPTCP systems [9].
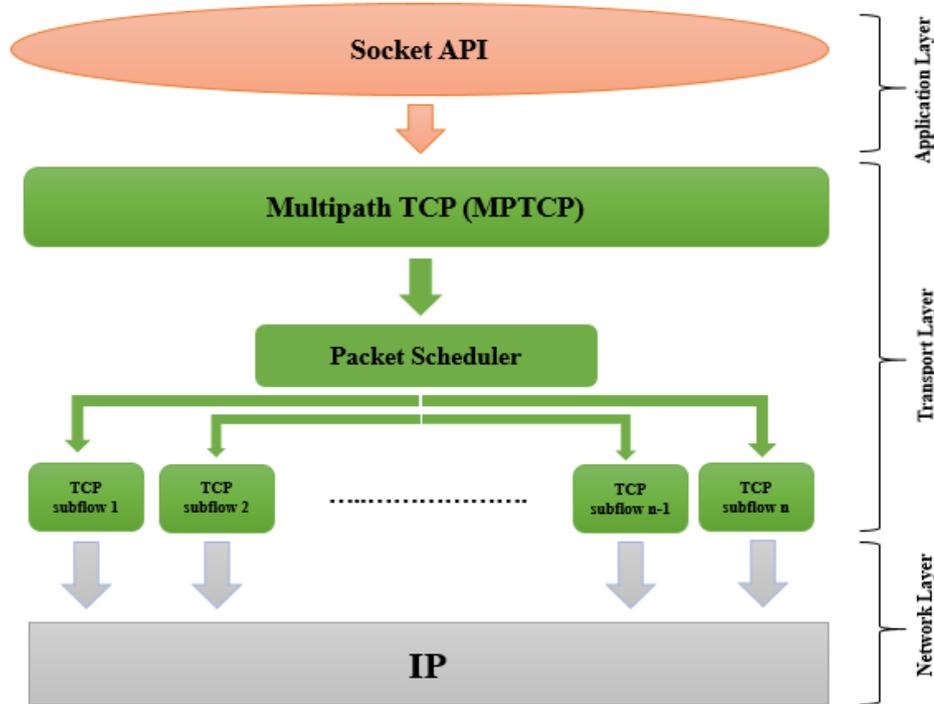

Figure 1: Architecture of MPTCP systems

As shown in Fig1, multiple network interfaces can handle the connections through different subflows. These MPTCP connections work like regular TCP connections, and the MPTCP scheduler is responsible for dividing user data among the subflows [10]. The decision regarding how to distribute user data across the subflows must be based on the network status for each subflow. The MPTCP systems provide several significant benefits as follows [10] [11]: The MPTCP can enable specific bandwidth aggregation. The system is very resilient to link failure difficulties. It has the potential to significantly improve mobility, optimize shared resources, and increase network capacity.

Background of MPTCP Schedulers

The MPTCP scheduling algorithms decide which packet must be sent on which path in the network. The most common schedulers make the scheduling decisions based on some key metrics provided by the TCP network layer like Congestion Windows (CWND), and Round-Trip Time (RTT). The Fig. 2 shows the scheduling process step by step [7]. The scheduler is invoked when the packets come from the upper layer (i.e., application layer) or when acknowledgements are obtained. The MPTCP scheduler starts to extracts some path features, including RTT values, and signal power. As a result, the most appropriate subflow can be chosen to establish the link [12].
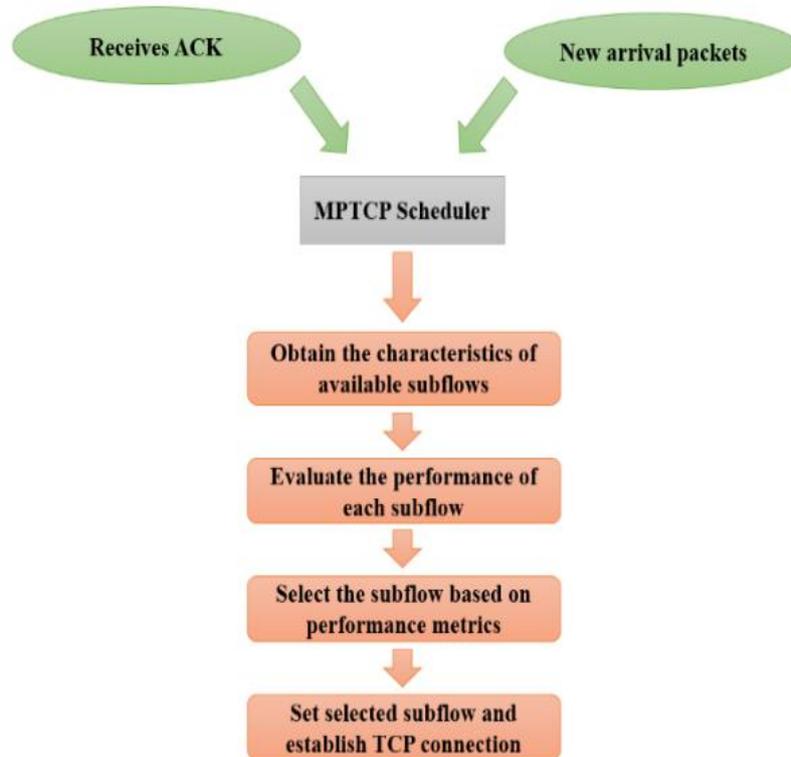
Figure 2: Scheduling process in the MPTCP

Notably, an appropriate MPTCP scheduler must be able to Present a reliable scheduling policy, especially when the systems have to deal with different network heterogeneity; Provide a condition to be friendly to standard TCP path; Balance a range of Quality of Service (QoS) metrics like reducing RTT, maximizing throughput; Transfer user segments from congested path to less congested paths properly [13-17].

Main challenges of the MPTCP and schedulers

The MPTCP protocols inherit some disadvantages of standard TCP. These are principally brought on by conventional loss-based congestion control methods. The main problem of the MPTCP technology is OfO packets in the receiving nodes [18]. When the packets are transmitted over different paths, the systems can face an increased end-to-end delay. The OfO problem occurs when a transmitted packets (i.e., data) which have greater sequence number arrives at the receiving node before others sent packet with a less sequence number. In addition, we have some high transmission error rates, particularly in different wireless systems due to the increased amount of OfO packets [18-24]. Another important issue with MPTCP is host buffer problem, which has yet to be completely addressed. As the receiver buffers are always limited, many OfO packets can ultimately fill the whole node buffer. This issue could be caused by a performance issue, such as slowing the user's transmission, which would result in Head of Line (HoL) blocking challenges. In this case (i.e., HoL blocking), the transmitted data (i.e., user packets) can be scheduled on quicker connections [25] [26]. These packets have to wait for other transmitted data (i.e., packets) which are assigned to the slower interface. This condition can produce severe congestion at the receiver buffer, increasing transmission delay unreasonably, especially in wireless networks with a high packet loss ratio [27-29].

RELATED WORKS

Many studies are presented to improve the performance of the MPTCP systems [30-35]. Furthermore, packet schedulers of MPTCP received Much attention in recent years. These proposed scheduling methods have different goals, applications, and approaches. In this section, we first review some important related works in MPTCP and scheduling methods briefly. Then, we discuss the main existing and proposed MPCTP schedulers in detail.

Round-Robin (RR) scheduler: When allocating user packets, the RR scheduler chooses one interface after another without using a logical preferring mechanism based on various characteristics such a lesser RTT. Each path is constrained by its CWND. Since there is no other alternative, the scheduler will allocate the next segment to the flow that receives the Acknowledgement (ACK) whenever it arrives [8] [14] [25].

Redundant scheduler: The MPTCP system uses a redundant scheduler to deliver the same packet over many subflows. By transferring the same data across several communication interfaces, this method lowers the amount of HoL blocking at the recipient. However, transmitting the same packet wastes a lot of bandwidth [9] [7].

MinRTT (Default) scheduler: The MinRTT scheduler is defined in the Linux implementation named Default. The MinRTT approach schedules packets to the subflow in its CWND space that has the lower RTT value. In the case of the HoL blocking in the MPTCP level, it can reduce CWND of the flow that caused the HoL block and retransmits packets via a faster flow [7].

BLEST (Blocking Estimation-Based scheduler): To reduce number of times faster subflow is unable to transmit a packet due to a lack of available space, the BLEST monitors the MPTCP send window. BLEST, like minRTT, schedules packets to the fastest (lowest RTT) subflow that is available. If not, sending the segment via a slower subflow is intended to prevent the faster flow from transmitting once it available again [11]. In heterogeneous networks, BLEST aims to minimize receiver buffer blocking.

Short Transfer Time First (STTF) scheduler: The concept behind STTF is pretty straightforward: compute the transmission time for each segment to schedule, taking into account data that is currently in flight. The STTF assigns packets to the flow that it believes will deliver them to the receiver in the shortest amount of time. To predict the increase of the CWND during the current RTT, the estimation considers the current status of congestion control [29].

Delay-Aware Packet (DAP) scheduler: DAP scheduler aims to send packets to the receiving node in proper manner. In order to transmit these packets to the receiver node, the DAP scheduler only needs to consider "not to block the receiver buffer". To ensure that segments arrive in order, the DAP scheduler should find which packets can be transmitted by which subflows by taking into account each subflow's CWND and forward delay [30][31].

OTIA method (Out-of-Order Transmission for In-Order Arrival scheduler): The OTIA Scheduler optimizes segment scheduling by assigning each segment to the path with the minimum transfer time or quickest route to its destination. The predicted transfer time over all possible subflows is calculated by the OTIA scheduler each time a given packet has to be re-scheduled [32].

Earliest Completion First (ECF) scheduler: The main aim of ECF is to find best use of the quickest of several different types of pathways. It considers the sender buffers' size, subflow's RTT estimates, and how much of their congestion window they occupy [12]. Additionally, ECF addresses path asymmetry-related performance degradation issues. The waiting time of fastest subflows are reduced. This method takes into account CWND parameter. It arranges the packets to subflow by taking into account the factors.

Forward Prediction Packet (FP) scheduler: The FP scheduler for multiple interface terminals with varying latencies is validated in SCTP method. if a subflow in the time of the scheduling gets free up CWND to retrieve fresh packets from the sender nodes, that calculates how long a new transmission will take [7]. Next, it calculates how many packets (N) can be sent concurrently across different paths during the specified period. Because the assessment is according to some cases that there would be no losses, the CWND can expand with each RTT.

As research on MPTCP scheduling algorithms continues to expand, Coninck et al. [15] present the main differences between standard TCP and existing MPTCP. The authors implement the MPTCP system on a range of Android smartphones. This evaluation checks the performance of MPTCP in mobile communication systems when these mobile devices work with critical applications like real-time video streaming and online gaming. In [16], authors state that the MPTCP provides increased bandwidth and improved resilience against failures of network paths, they also discuss the impact out-of-order packets on communication. The authors give specific instructions for reordering the packets. Beig et al. [17] present that by using a learning algorithm, the scheduler can determine the optimal case for the quality rate of the signal given the kind of interface and amount of the transmitted file. A method to check the effectiveness of long-lasting subflows in different heterogeneous wireless networks has been developed by Pokhrel and Mandjes [18]. They took into account a number of network parameters, including buffer sizes and the retransmission limit. Polese et al. [19] highlight research trends on congestion control methods, novel transport protocols, and multi-path capabilities. Although survey articles provide a broad understanding of the multi-path domain, lessons particular to the MPTCP highlight its unique characteristics. According to Pokhrel and Garg [20], extremely dynamic networks will require more than just optimization-based MPTCP techniques based on complexity of method and accuracy of scheduling method.

A novel throughput-based method was proposed by Pokhrel et al. [19] to schedule user's packets in different heterogeneous wireless networks. This method has a combined technique to provide a proper load balance as long as an appropriate Forward Error Correction (FEC). This can perform coupled congestion management MPTCP, focusing on time-sensitive systems. In addition, ReLeS method was proposed by Zhang et al. [22]. It uses DRL techniques to construct a Neural Network (NN) in order to determine the optimal distribution cases. Due to its determination of several QoS criteria, the reward function in use is complex. Peekaboo is a unique learning-based Multi-path Quick UDP Internet Connections (MPQUIC) scheduler that Wu et al. [23] presented. It continuously monitors the impact made by each path's current dynamicity level and chooses the best scheduling approach in accordance with that impact. The EDPF scheduling is a noteworthy network layer solution [24]. It sends the data via the subflow that sends these packets after dynamically estimating the following packet's sending time on each link. It has been demonstrated by Frommgen et al. [26] that for short streaming usages

such as HTTP application, subflow techniques according to RTT values can suffer from incorrect RTT values. However, in order to improve web services, some authors have already developed an MPTCP packet scheduler that can suppress QoS fluctuation [27]. Through actual experiments, they demonstrated that the proposed scheduler can do so more effectively than the default scheduler. Furthermore, [28] shows that the suggested scheduler reduces throughput more than the default scheduler.

The current literature on MPTCP scheduling presents multiple methods, each aimed at enhancing metrics such as latency, bandwidth efficiency, and resilience. Foundational schedulers such as RR and MinRTT may encounter challenges when operating within dynamic environments. Advanced methods like BLEST and ECF attempt to address certain problems like buffer blocking and path asymmetry, but their flexibility is limited because many of them rely on static parameters like RTT. While recent approaches use adaptive learning but bring high computational costs, highlighting the necessity for efficient, real-time schedulers that can manage diverse and quickly changing network conditions.

Evaluation of MPTCP scheduler

This section presents a thorough summary of relevant MPTCP scheduler works in Table 1. Although the MPTCP proposes a set of scheduling methods, we included the most popular schedulers along with their main concepts, benefits, and drawbacks.

Table 1: A comparison of scheduling algorithm in MPTCP

Scheduler's Name	Main idea of scheduler	Key advantages of schedulers	Main disadvantages of the schedulers
RR [25]	All paths are used in succession, one subflow after the another one.	It offers optimal load balancing and capacity aggregation. In symmetric subflows, it performs admirably.	It is unable to function effectively in asymmetric subflows, or heterogeneous routes.
MinRTT [7]	The MinRTT employs the faster path which has the shortest RTT value and suitable CWND size.	MinRTT can be very easy to set in many MPTCP systems.	MinRTT is not able to set appropriately with asymmetric subflows. This method cannot find number of user packets for each path well.
Redundant [9]	Redundant can utilize whole paths to transmit the same user's packets.	This method is fast and it needs less latency.	This scheduler provides more overhead in the receiving nodes.
DAP [31]	DAP employs all paths to handle HoL blocking issues.	DAP method can be good for different wireless subflows.	This method cannot handle many network failures effectively.
ECF [12]	The ECF is presented to increase network throughput by using all fast subflows.	This method uses optimal completion time. This policy can be useful for different heterogeneous scenarios.	ECF upsurge OfO packets which is inappropriate in the receiver side.
OTIAS [32]	OTIAS method gives a specified packet to a given subflow which has small departure time. This can	This scheduler employs less parameters. It also cuts transmission time	OTIAS suffers from high number of OfO packets and also this method is not able to handle reinjections.

	minimize transmission time.	that is important in real-time applications.	
BLEST [11]	BLEST avoids transferring user's packets on each slower subflow when a faster path is available.	The BLEST schedules packets for different heterogeneous scenarios by decreasing HoL blocking.	BLEST makes high errors and it suffers from performance degradation if losses arise regularly.
STTF [29]	STTF allocates unsent segments to subflows according to transfer time.	STTF can work fine for heterogeneous scenarios.	This method suffers from very high OfO packets in wireless systems.

Furthermore, Table 2 provides an overview of the schedulers, emphasizing their applications, overarching objectives, and measures that are used to determine scheduling. Table 2 displays how several schedulers are arranged in the solution space with respect to applications, metrics, and overall objectives. The multi-path performance can be enhanced in bulk transfer applications with higher throughput needs by using the schedulers LL [25] [7], ESPC [33], LTS [35], LWS [35], and HSR [35].

Table 2: An evaluation of involved metrics in MPTCP scheduler.

Schedulers	Main goals of each scheduler	Key application Of each scheduler	Important metrics						
			Size of CWND	RTT values	MSS values	Loss packets	ID of Subflow	Number of Inflight packets	segment size
ESPC [33]	Increase network throughput	Effective bulk transfer	✓	✓		✓			
LL [25]	Increase network throughput	Effective bulk transfer		✓					
LTS [35]	Increase network throughput	Effective bulk transfer	✓	✓				✓	
LWS [35]	Increase network throughput	Effective bulk transfer	✓	✓				✓	
HSR [35]	Increase network throughput	Effective bulk transfer	✓	✓	✓				
WRR [36]	Enhance load balancing level	Effective bulk transfer	✓						✓
BLEST [11]	Reduce number of retransmissions packets, and handle HoL-blocking issue	Effective bulk transfer	✓	✓	✓			✓	
ECF [12]	Not only improve utilization of faster paths, but also decrease completion time for transmission	video streaming systems	✓	✓					

OTIA [32]	Decrease completion time, handle high jitter, and abate broadcast time	Real-time systems	✓	✓					✓	
LL/RP [34]	Ease HoL-blocking issue	Wide-ranging app	✓	✓						
DAP-2 [31]	Moderate HoL-blocking problems, and exploit the probability of OfO segments	Wide-ranging app	✓	✓					✓	
RR [25]	Defined for academic goals	Wide-ranging app						✓		
LL/BM [34]	Moderate bufferbloat problem	Wide-ranging app	✓	✓						
STTF [29]	Cut completion time importantly	Web-based app	✓	✓						✓

Other schedulers with specialized functions, such as packet retransmission decrease, HoL-block moderation in the BLEST method, and load balancing optimization in WRR [36], are available for these applications. ECF [12] can optimize the usage of the fastest paths in the connection for video/audio streaming systems, enhancing QoS metrics in those streams. Without any particular applications, the schedulers LL/RP [34] and DAP-2 [31] try to reduce HoL-blocking.

The LL/BM [34] is able to work in common scenarios to mitigate buffer bloat issue. However, because it induces HoL-blocking, which results in poor performance, RR [25] is only suitable for test academy use. The OTIA [32] scheduler seeks to reduce transmission latency and jitter for real-time applications. The schedulers can shorten the completion time of applications that need short transmission periods. STTF [29] can be used to reduce the transmission time of interactive applications and the Web. In addition, we compare a range of MPTCP schedulers based on level of path utilization, and key constraints in Table 3. In conclusion, the MPTCP schedulers discussed above make an effort to enhance network performance using various strategies. To extract the key characteristics of various schedulers, we compare and analyze them in this paper.

Table 3: A comparison of MPTCP schedulers based on path utilization.

Scheduler	Path Utilization	Constraints
RR [25]	Simultaneous utilization of all paths	Indeterminate fairness
MinRTT [7]	A specific path at a time	It is unable to function on asymmetric paths.
DAP [31]	Use of simultaneous paths	Unable to handle network outages. It creates schedules without the ability to adapt to changes in the network. Schedule reinjections, or retransmissions, are not used.
ECF [12]	Prefers fast paths	It is vulnerable to head-of-line blocking
OTIA [32]	Use of simultaneous paths	Symmetric forward delays are assumed, and schedule reinjections, or retransmissions, are not used.
DEMS [37]	Simultaneous utilization of all path	For effective scheduling, rely on precise understanding of the data chunk boundaries.
CP [38]	Prefers fast paths	Needs network support

STMS [39]	Simultaneous utilization of all paths	Sensitivity of RTT errors
Qaware [40]	Simultaneous utilization of all path	Needs network support
BLEST [11]	Primarily fast paths	Sensitivity of RTT errors

CONCLUSION AND FUTURE WORKS

Multiple network interfaces are common in modern devices, and MPTCP is a promising technique to make use of this functionality. Nowadays, a wide spectrum of wireless communication channels has significant loss rates. MPTCP proposals a resolution for further reliable data communications than standard TCP in such lossy situations. The MPTCP design should utilize every selected TCP path in different multi-homed tools simultaneously and professionally. MPTCP is a potential transport layer protocol that offers greater robustness and throughput. A variety of factors influence MPTCP's performance. Among them, the number of subflows can be considered an MPTCP application layer parameter; In transport layer of network architecture, scheduling and congestion control methods are applied; and in network layer, systems employ MPTCP paths. In this study, we examine several schedulers and concentrate on scheduling methods. Scheduling methods can improve MPTCP throughput by ensuring that user's packets must be arrived in correct manner. However, the design concepts of the scheduling algorithms given here cannot account for every aspect of network performance, especially in wireless networks. We discovered that majority of the currently available schedulers are ineffective for bulk or video traffic in lossy networks. The parameter that has the biggest impact on MPTCP performance is path heterogeneity. Schedulers mostly use fast paths to handle OfO packets; nevertheless, that is not able to accomplish main goals of MPTCP such as high bandwidth aggregation. The overall performance of MPTCP is degraded by HoL-blocking at the receiver caused by such path heterogeneity. In order to address heterogeneity issue of wireless paths, the proposed MPTCP employs a responsive strategy, penalizing the paths which result in HoL-blocking via a retransmission mechanism. The failure is caused by the scheduling algorithm that chooses which subflow to send each packet. The goal of multi-path scheduling should be to receive the data in the correct order. Programs can continuously read data from the receive queue, which lowers HoL blocking and CWND limitation. In thispaper, we discussed the advantages and disadvantages of several MPTCP schedulers for different use situations. We discovered that neither multi-homing situation nor traffic use-cases could function effectively. We examined and outlined the properties of the MPTCP schedulers currently available in this study. Next, we looked at a few commonly used schedulers in different network conditions. To show the issues, this study examines many cutting-edge MPTCP schedulers (e.g., BLEST, MinRTT, OTIA, and ECF).

In the future, we hope to design an optimal scheduler to improve MPTCP performance, particularly in very lossy environments including different wireless scenarios. We will compare our suggested scheduler to the most advanced schedulers in a variety of network scenarios, notably those with considerable path heterogeneity. In addition, we plan to investigate how schedulers and congestion management methods interact and how this affects MPTCP performance in general.

REFERENCES

[1] K. Vidya, and S. Sudhir, "Exploring Multi-path TCP Schedulers in Heterogeneous Networks", International Journal of Information Communication Technologies and Human Development, pp. 1-11, 2022.

[2] M. Prakash, and A. Abdrabou, "On the fidelity of NS-3 Simulations of Wireless Multi-path TCP Connections", Sensors Journal, 2020.

[3] S. Afzal, V. Testoni, C. E. Rothenberg, P. Kolan, and I. Bouazizi, "A Holistic Survey of Wireless Multi-path Video Streaming", Journal of Network and Computer Applications, Vol. 212, 2023.

[4] Y. Xing, K. Xue, Y. Zhang, J. Han, J. Li, and D. S. L. Wei, "An Online Learning Assisted Packet Scheduler for MPTCP in Mobile Networks", In IEEE/ACM Transactions on Networking, pp. 1-16, 2023.

[5] N. Thakur, and K. Ashwini, "Analysing Schedulers of Multi-path TCP in Diverse Environment", International Conference on Advances in Computing, Communication Control and Networking, pp. 1337-1340, 2021.

[6] P. Kumar, F. Nida, and P. Saxena, "Performance Analysis of Multi-path Transport Layer Schedulers under 5G/B5G Hybrid Networks", 14th International Conference on Communication Systems, pp. 658-666, 2022.

[7] A. Khajekini, H. Amca, A. Ulusoy, and E. Enver, "Enhancing MPTCP Performance in Wireless Networks with PRS-MPTCP", Ad Hoc Networks, Vol. 162, 2024.

[8] W. Yang, D. Pingping, C. Lin, and T. Wensheng, "Loss-Aware Throughput Estimation Scheduler for Multi-path TCP in Heterogeneous Wireless Networks", IEEE Transactions on Wireless Communications, pp. 3336-3349, 2021.

[9] X. Yitao, K. Xue, Y. Zhang, J. Han, L. Jian, L. Jianqing, and L. Ruidong, "A Low-Latency MPTCP Scheduler for Live Video Streaming in Mobile Networks", IEEE Transactions on Wireless Communications, pp. 7230-7242, 2021.

[10] L. Li, K. Xu, L. Tong, K. Zheng, C. Peng, D. Wang, X. Wang, M. Shen, and R. Mijumbi, "A Measurement Study on Multi-path TCP with Multiple Cellular Carriers on High Speed Rails", Conference of the ACM Special Interest Group on Data Communication, pp. 161-175, 2018.

[11] S. Ferlin, A. Ozgu, M. Olivier, and B. Roksana, "BLEST: Blocking Estimation-based MPTCP Scheduler for Heterogeneous Networks", FIP Networking Conference and Workshops, pp. 431-439, 2016.

[12] L. Yeon, N. Erich, T. Don, and G. Richard, "ECF: An MPTCP Path Scheduler to Manage Heterogeneous Paths", 13th International Conference on emerging Networking EXperiments and Technologies, pp. 147-159, 2017.

[13] J. Han, K. Xue, X. Yitao, L. Jian, W. Wenjia, D. Wei, and X. Guoliang, "Leveraging Coupled BBR and Adaptive Packet Scheduling to Boost MPTCP", IEEE Transactions on Wireless Communications, pp. 1-14, 2021.

[14] V. Yedugundla, F. Simone, D. Thomas, A. Ozgu, K. Nicolas, H. Per, and B. Anna, "Is Multi-path Transport Suitable for Latency Sensitive Traffic?", Computer Networks, Vol. 105, 2016.

[15] C. Quentin, B. Matthieu, H. Benjamin, and B. Olivier, "Observing Real Smartphone Applications over Multipath TCP", IEEE Communications Magazine, Vol. 54, pp. 88-93, 2016.

[16] A. Alheid, A. Doufexi, and D. Kaleshi, "A Study on MPTCP for Tolerating Packet Reordering and Path Heterogeneity in Wireless Networks", Wireless Days journal, pp. 1-7, 2016.

[17] M. Beig, P. Daneshjoo, S. Rezaei, A. Movassagh, R. Karimi, and Y. Qin, "MPTCP Throughput Enhancement by Q-learning for Mobile Devices", In IEEE 20th International Conference on High Performance Computing and Communications, pp. 1171-1176, 2018.

[18] S. R. Pokhrel, and M. Mandjes, "Improving Multi-path TCP Performance over WiFi and Cellular Networks: An Analytical Approach", IEEE Transactions on Mobile Computing, pp. 2562-2576, Vol. 18, 2018.

[19] M. Polese, C. Federico, B. Elia, R. Filippo, Z. Andrea, Z. Michele, "A Survey on Recent Advances in Transport Layer Protocols", IEEE Communications Surveys & Tutorials, Vol. 21, No. 4, pp. 3584-3608, 2019.

[20]   S. R. Pokhrel, and S. Garg, "Multi-path Communication With Deep Q-Network for Industry 4.0 Automation and Orchestration", In IEEE Transactions on Industrial Informatics, Vol. 17, No. 4, pp. 2852-2859, 2021.

[21]   S. R. Pokhrel, and J. Choi, "Low-delay Scheduling for Internet of Vehicles: Load-balanced Multi-path Communication with FEC", IEEE Transactions on Communications, Vol. 67, pp. 8489-8501, 2019.

[22]   H. Zhang, W. Li, S. Gao, X. Wang, and B. Ye, "ReLeS: A Neural Adaptive Multi-path Scheduler based on Deep Reinforcement Learning", IEEE Conference on Computer Communications, pp. 1648-1656, 2019.

[23]   W. Hongjia, A. Ozgu, B. Anna, F. Simone, and C. Giuseppe, "Peekaboo: Learning-Based Multi-path Scheduling for Dynamic Heterogeneous Environments", IEEE Journal on Selected Areas in Communications, Vol. 38, No. 10, pp. 2295-2310, 2020.

[24]   K. Chebrolu, and R. R. Rao, "Bandwidth Aggregation for Real-time Applications in Heterogeneous Wireless Networks", In IEEE Transactions on Mobile Computing, Vol. 5, No. 4, pp. 388-403, 2006.

[25]   B. Kimura, L. Demetrius, and L. Antonio, "Packet Scheduling in Multi-path TCP: Fundamentals, Lessons, and Opportunities", IEEE Systems Journal, pp. 1445-1457, 2020.

[26]   A. Frommgen, J. Heuschkel, and B. Koldehofe, "Multi-path TCP Scheduling for Thin Streams: Active Probing and One-way Delay-awareness", IEEE International Conference on Communications (ICC), pp.1-7, 2018.

[27]   K. Noda, and Y. Ito, "Study of Multi-Path TCP Scheduler to Suppress QoS Fluctuation for Improving WebQoE", International Conference on Electronics, Information, and Communication (ICEIC), pp. 1-4, 2019.

[28]   A. Mondal, A. R. Kabbinale, S. Shailendra, H. K. Rath, and A. Pal, "PPoS: A Novel Sub-flow Scheduler and Socket APIs for Multi-path TCP (MPTCP)", Twenty Fourth National Conference on Communications, pp. 1-6, 2018.

[29]   P. Hurtig, G. Karl, B. Anna, F. Simone, A. Ozgu, and K. Nicolas, "Low-Latency Scheduling in MPTCP", IEEE/ACM Transactions on Networking, Vol. 27, pp. 302-315, 2019.

[30]   G. Sarwar, R. Boreli, E. Lochin, A. Mifdaoui, and G. Smith, "Mitigating Receiver's Buffer Blocking by Delay Aware Packet Scheduling in Multi-path Data Transfer", 27th International Conference on Advanced Information Networking and Applications Workshops, pp. 1119-1124, 2013.

[31]   N. Kuhn, E. Lochin, A. Mifdaoui, G. Sarwar, O. Mehani, and R. Boreli, "DAPS: Intelligent Delay-aware Packet Scheduling for Multi-path Transport", IEEE International Conference on Communications (ICC), pp. 1222-1227, 2014.

[32]   F. Yang, Q. Wang, and P. D. Amer, "Out-of-Order Transmission for In-Order Arrival Scheduling for Multi-path TCP", 28th International Conference on Advanced Information Networking and Applications Workshops, pp. 749-752, 2014.

[33]   F. Yang, P. Amer, and N. Ekiz, "A Scheduler for Multi-path TCP", 22nd International Conference on Computer Communication and Networks, pp. 1-7, 2013.

[34]   C. Paasch, S. Ferlin, O. Alay, and O. Bonaventure, "Experimental Evaluation of Multi-path TCP Schedulers", ACM SIGCOMM Workshop on Capacity Sharing, pp. 27-32, 2014.

[35]   B. L. Kimura, D. F. Lima, and A. Loureiro, "Alternative Scheduling Decisions for Multi-path TCP", In IEEE Communications Letters, Vol. 21, No. 11, pp. 2412-2415, 2017.

[36]   K. W. Choi, Y. S. Cho, J. W. Lee, S. M. Cho, and J. Choi, "Optimal Load Balancing Scheduler for MPTCP-based Bandwidth Aggregation in Heterogeneous Wireless Environments", Computer Communications, Vol. 112, pp. 116-130, 2017.

[37]   Yu Liu, C. Zhang, B. Li, and J. Niu, "DeMS: A Hybrid Scheme of Task Scheduling and Load Balancing in Computing Clusters", Journal of Network and Computer Applications, Vol. 83, pp. 213-220, 2017.

[38]   B. Hwan, J. Lee, "Constraint-Based Proactive Scheduling for MPTCP in Wireless Networks", Computer Networks, Vol. 91, pp. 548-563, 2015.

[39]   H. Shi, Y. Cui, X. Wang, Y. Hu, M. Dai, F. Wang, and K. Zheng, "STMS: Improving MPTCP Throughput under Heterogeneous Networks", USENIX Annual Technical Conference (USENIX ATC 18), 2018.

[40]   T. Shreedhar, N. Mohan, S. K. Kaul and J. Kangasharju, "QAware: A Cross-Layer Approach to MPTCP Scheduling", IFIP Networking Conference (IFIP Networking) and Workshops, pp. 1-9, 2018.

RESUME

Atefeh Ahmadniai KHAJEKINI

Atefeh Ahmadniai Khajekini received the BSc degree in Electrical and Electronic Engineering from the Islamic Azad University of Lahijan (Iran) in 2008. In October 2013, she joined the Department of Computer Engineering at the University of Guilan, Iran, as a full time MSc student. She received the MSc degree in 2015. She is a Ph.D. student under advisor Ali Hakan Ulusoy in the Electrical and Electronic Engineering Department of Eastern Mediterranean University (EMU). She is a senior instructor with the Electrical and Electronic Engineering Department at the Final International University in Northern Cyprus. Her research interests include computer networks, wireless networks, network security, cryptography algorithms, and mobile communications.

Prof. Dr. Ali Hakan ULUSOY

Prof. Dr. Ali Hakan Ulusoy was born on June 3, 1974, in Eskisehir, Turkey. He graduated from the double major program of the Department of Electrical and Electronic Engineering (EEE) and the Department of Physics at Eastern Mediterranean University (EMU) in 1996. Subsequently, he earned his M.S. and Ph.D. degrees in EEE from EMU in 1998 and 2004, respectively. In 2004, he joined the Department of Information Technology at EMU as an instructor. Over the years, he progressed through various academic ranks: he became a visiting assistant professor in 2005, an assistant professor in 2007, an associate professor in 2013, and finally, a professor in 2019. During his tenure at EMU, he held administrative roles at the Institute of Graduate Studies and Research (IGSR). He served as the assistant director from January 2012 to August 2017 and as the acting director from August 2017 to December 2019. Since January 2020, he has been serving as the director of IGSR. His research interests span across various areas, including wireless communications, receiver design, channel estimation, fuzzy systems, wireless networks, cloud computing, millimeter-wave communications, healthcare system development, and support vector machines for intrusion detection in vehicular ad-hoc networks.

Prof. Dr. Enver EVER

Prof. Dr. ENVER EVER received an M.Sc. degree in computer networks and a PhD in performance evaluation of computer networks and communication systems from Middlesex University, London, in 2004 and 2008, respectively. He was a PostDoctoral Research Associate with Bradford University for one year. He was a Lecturer/Senior Lecturer with the Computer and Communications Engineering Department at Middlesex University from 2008 to 2013. He is a Professor at the Middle East Technical University Northern Cyprus Campus. His current research interests include the Internet of Things, computer networks, wireless sensor networks, wireless multimedia sensor networks, artificial intelligence (machine learning/deep learning) applications, wireless communication systems, cloud computing, and performance/reliability modelling. He serves on various boards and program committees.

Mostafa Ayoubi MOBARHAN

Mostafa Ayoubi Mobarhan is an Assistant Professor in the Engineering Faculty at Final International University, where he has been a faculty member since 2022. He received his BSc degree in Electrical Engineering (Electronics) from Islamic Azad University of Lahijan, Iran, in 2008. In 2009, he joined the Department of Computer Engineering at the University of Guilan, Iran, as a full-time MSc student and earned his MSc degree in 2011. He later pursued his Ph.D. in the Computer Engineering Department of Eastern Mediterranean University (EMU), focusing on advanced topics in wireless networks and network security. Dr. Ayoubi Mobarhan's research interests include wireless networks, network security, cryptographic algorithms, and mobile communications. His work contributes to enhancing secure and efficient communication in modern networked environments. Since joining Final International University, he has been actively involved in both teaching and research, striving to inspire the next generation of engineers and researchers.

DESIGN OF A DIRECTIONAL COUPLER BASED CSRR MICROWAVE SENSOR TO ENHANCE SENSITIVITY

İbrahim GENÇ[1], Mahmut Ahmet GÖZEL[2], Mesud KAHRİMAN[3]

INTRODUCTION

Microwave sensors are developing technology for measuring the dielectric properties of materials with high precision. Characterising the dielectric properties of materials has a critical role in the production and material quality determination applications in many fields such as biomedical, defence, and agriculture [1-6]. Although the penetration of low-power electromagnetic (EM) waves into materials is limited, their interaction with the material is sufficient for non-destructive analysis. Non-destructive dielectric constant measurement methods can be divided into two main categories: (1) resonant methods, which detect shifts at a specific resonant frequency, and (2) non-resonant methods, which track changes in the properties of EM waves over a wide frequency range [1]. Non-resonant methods can measure the relative dielectric constant of various materials without considering the frequency sensitivity of the material. However, in these methods, the sensing sensitivity is lower, and performance is limited by the accuracy of the model used to convert the obtained data into dielectric constants [7,8]. Resonant methods provide high sensitivity and selectivity when determining the dielectric constant of a target sample (MUT) around the resonator [9].

Resonator-based planar microwave sensors, which are widely used in the literature, are based on split ring resonator (SRR) and complementary split ring resonator (CSRR) structures [9]. This is because these resonators are low-profile and non-destructive, and can be easily applied on a planar dielectric surface. A high electric field region is required in the resonator to measure the dielectric properties of the MUT with high sensitivity. With the SRR and CSRR structures, the high-intensity electric and magnetic fields necessary to determine the dielectric properties of a material can be generated [10]. A structure with high field distribution further increases the interaction between the loaded MUT and the resonator [11].

The dielectric properties of the MUT, the resonance frequency (f_r) shift in the sensor's unloaded and loaded states, and the amplitude change in the S-parameters or the change in the quality factor (Q) can be calculated with appropriate mathematical models. The resonance frequency of the sensor can be written as follows when unloaded.

[1] Isparta University of Applied Sciences, Isparta/Turkey, Orcid: https://orcid.org/0000-0002-1693-6685, ibrahimgenc@isparta.edu.tr
[2] Süleyman Demirel University, Isparta/Turkey, Orcid: https://orcid.org/0000-0002-0360-7188 , mahmutgozel@sdu.edu.tr
[3] Süleyman Demirel University, Isparta /Turkey, Orcid: https://orcid.org/0000-0003-0731-0936 , mesudkahriman@sdu.edu.tr

$$f_r = \frac{1}{2\pi\sqrt{L_o C_o}} \tag{1}$$

Here, L_0 and C_0 represent the sensor's effective inductance and capacitance, respectively. These values depend on the topology and dimensions of the sensor structure. Since the material to be characterized is non-conductive, L_0 can be considered constant. However, when the MUT is loaded onto the sensor, the capacitance of the test material, C_{MUT}, combines with the sensor's capacitance C_0. As a result, the overall capacitance changes to ($C_{eff} = C_o + C_{MUT}$), and due to this change in capacitance, the sensor's resonant frequency f_r shifts to f_{rMUT} [12].

$$f_{r_{MUT}} = \frac{1}{2\pi\sqrt{L_o C_{eff}}} \tag{2}$$

The quality factor is a fundamental characteristic parameter for many microwave circuits. It defines how much energy is dissipated in a system relative to the stored energy. Q determines the energy loss at the resonant frequency of a resonator. Higher Q means lower energy loss and increases the resonator's quality [13]. The unloaded quality factor (Q_0) refers to the Q for the microwave resonator without a MUT load and Q_0 can be calculated as follows [9].

$$Q_0 = \frac{f_r}{\Delta f} \tag{3}$$

Here Q_0 is the unloaded quality factor, fr is the unloaded resonant frequency of the microwave resonator and Δf is the bandwidth of the frequency ± 3 dB concerning the minimum transitions. The quality factor (Q_{MUT}) of the microwave resonator loaded with MUT can be expressed as follows [12].

$$Q_{MUT} = \frac{f_{r_{MUT}}}{\Delta f_{MUT}} \tag{4}$$

A sharp decrease in the resonant frequency leads to a higher Q value. Q can be written as a function of the loss tangent $tan(\delta)$ as follows. [14] .

$$Q = \frac{1}{tan(\delta)} \tag{5}$$

The sensitivity of microwave sensors is important to the performance of a sensor because it determines how accurately the sensor operates at a particular frequency. To measure a sensor's sensitivity, the sensor's resonant frequencies in the unloaded and MUT loaded states are considered. Since the relative permittivity and loss tangent of air are well known, the unloaded sensor is usually used as a reference model. When the sensor is loaded with MUT in the area where the electric field is strongest, the resonant frequency changes depending on the permittivity of the MUT. This relationship is linear and any change in the relative permittivity ($\Delta\varepsilon_r$) causes a change in the resonant frequency (Δf_r). To compare the sensitivities correctly, it has to be normalized concerning the empty sensor's resonant frequency (f_{rair}). If the empty

sensor is considered to be loaded with air, $\Delta\varepsilon_r$ is the difference between the electrical permittivity of the MUT ($\Delta\varepsilon_{rMUT}$) and the electrical permittivity of air (ε_{rair} =1). The normalized sensitivity $S(\%)$ is written as follows [5].

$$S(\%) = \frac{f_{r\,air} - f_{r\,MUT}}{f_{r\,air}(\varepsilon_{r\,MUT} - 1)} x100 \tag{6}$$

The existing resonator-based microwave sensor studies in the literature have mostly developed SRR or CSRR structures placed in and around the transmission line [4,8,10,12]. In addition, in a study where only a directional coupler was used as a microwave sensor, the sensor sensitivity was compared with resonator-based sensors by monitoring the changes in the S_{31} and S_{41} parameters of the coupler [15].

The directional coupler is a passive high-frequency measuring element with unidirectional transmission and four ports (Figure 1) that allows samples to be taken from reflected and transmitted signals in microwave systems [16]. Port 1 of the directional coupler is the input port, Port 2 is the transmission port, Port 3 is the coupling port and Port 4 is the isolation port. RF signal power is applied to the directional coupler from the input port and a small part of the applied power is transferred to the coupling port while a large part is transmitted to the transmission port. Theoretically, no power is transferred to the isolation port [17].

It has been determined that CSRR sensors provide higher sensitivity when compared to SRR sensors of the same size [18]. It has been stated that CSRRs are more sensitive to permeability and show superior performance in terms of resonance sharpness due to their electrical coupling to the transmission lines, narrow bandwidth, and low insertion loss [19].
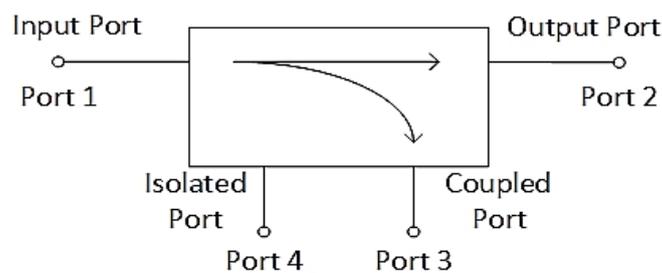

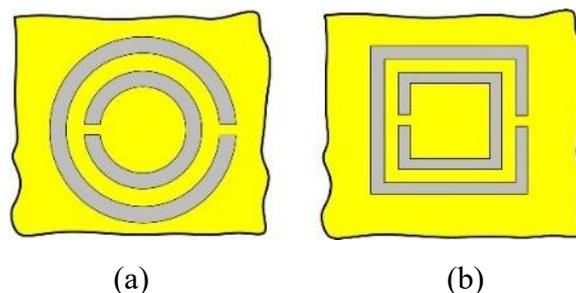Figure 1: Basic structure of directional coupler.

(a)   (b)

Figure 2: Basic CSRR structures a) circular CSRR b) rectangular CSRR.

This study proposes a new microwave sensor design using a directional coupler and CSRR structures. The proposed sensor design consists of a directional coupler with a -10dB coupling factor at 1.1 GHz placed on an easily available FR-4 substrate material with low cost and four identical CSRR structures placed to increase the sensitivity to the ground plane. The purpose of using the directional coupler in the study is to examine the changes in S_{31} and S_{41} parameters except for the S_{21} parameter, unlike the two-port transmission line sensor structures, and to compare the results. The simulation results show that the proposed method provides an accurate and cost-effective solution for determining the dielectric properties of the materials.

Microwave Sensor Design

In this study, a microwave sensor using CSSR structures at a directional coupler is designed in a simulation tool to increase the sensitivity. FR4 substrate material with dielectric constant ε_r=4.3, thickness 1.5 mm, and $tan(\delta)$ =0.025 was used for the sensor design. Figure 2 shows the structure of the sensor. The dimensions of the sensor are shown in Table 1.
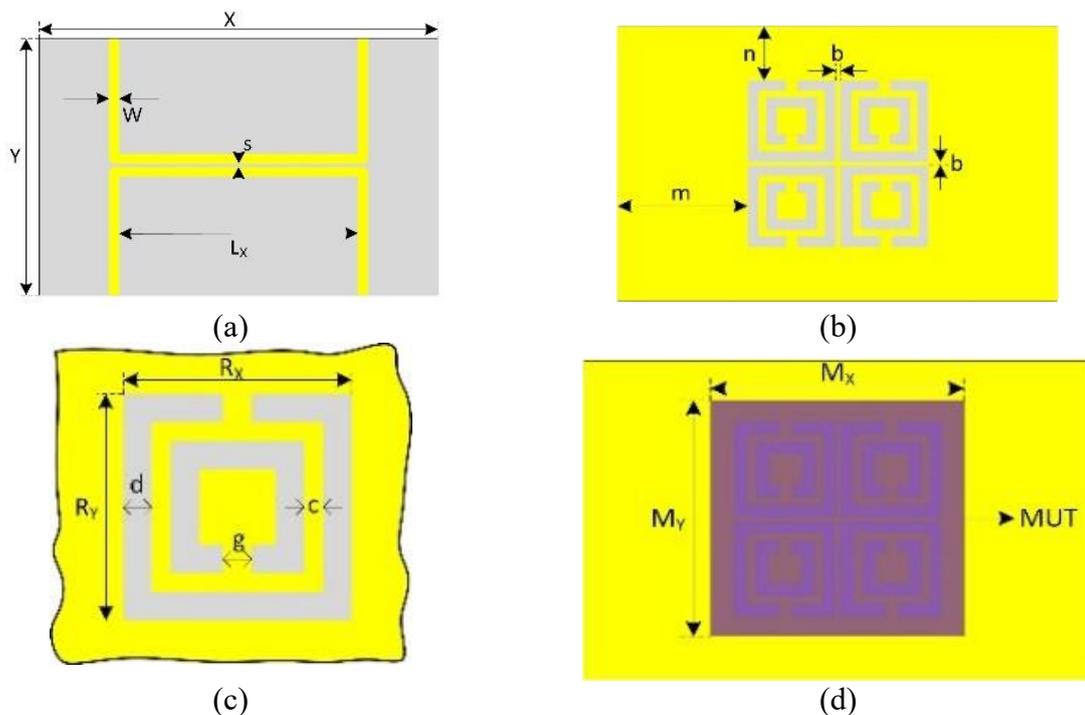

Figure 2: Structure of the microwave sensor a) front view b) rear view c) structure of a CSSR element d) placement of the MUT on the sensor.

Table 1: Dimensions of the microwave sensor

X	60 mm	m	17.75 mm	g	1.5 mm
Y	40 mm	n	7.75 mm	R_X	12 mm
W	1.5 mm	b	0.5 mm	R_Y	12 mm
s	0.5 mm	c	1 mm	M_X	30 mm
L_X	40 mm	d	1.5 mm	M_Y	30 mm

Figure 3 shows the S-parameter graphs obtained from the simulation of the directional coupler without the CSSR structures used in the sensor design. The directional coupler operates with a -10 dB coupling factor at 1.1 GHz. The S-parameter graph of the microwave sensor designed using the directional coupler and CSSR structures in Figure 2, obtained from the simulation when the MUT is unloaded, is presented in Figure 4.
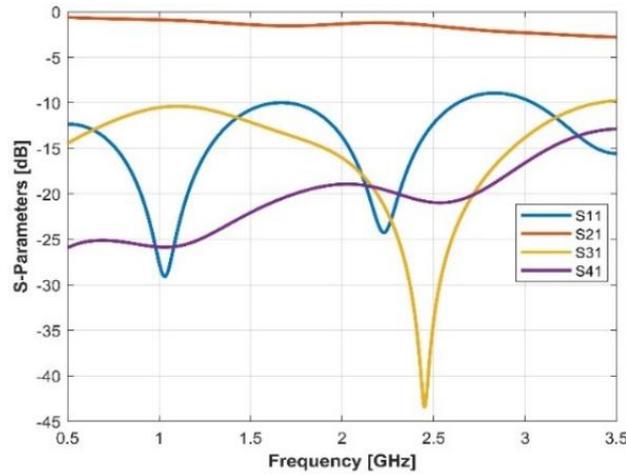

Figure 3: S-parameters obtained from the simulation results of the directional coupler used in the sensor design.
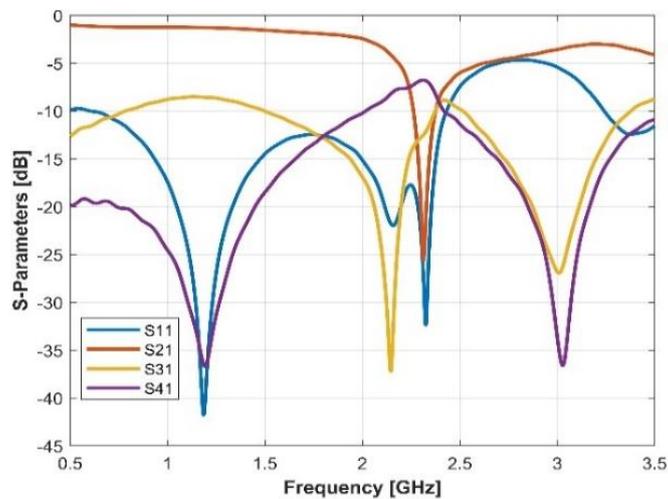
Figure 4: S-parameters of the designed microwave sensor obtained with MUT unloaded, simulation result.

When the graphs in Figure 2 and Figure 4 are compared, it is seen that distinctive new resonances are formed in parameters S_{21} and S_{41} due to the CSRR structures added to the directional coupler. It is also noteworthy that in parameter S_{31}, the resonance of the directional coupler at 2.45 GHz shifts to 2.144 GHz, and a new resonance appears at 3.008 GHz. The f_r and Q values obtained from the simulation of the unloaded microwave sensor are given in Table 2.

Table 2: Simulation results of the designed microwave sensor in no-load condition.

S-Parameters	f_{r1} (Ghz)	f_{r2} (Ghz)	Q_{fr1}	Q_{fr2}
S_{21}	2.315	-	140.3	-
S_{31}	2.144	3.008	111.7	26.8
S_{41}	1.193	3.032	14.8	75.8

Evaluation Of The Simulation Results

The microwave sensor is loaded with different MUTs as shown in Figure 2(d). The thickness of the MUTs is 1.25 mm, the $tan(\delta)$ values are kept constant at 0 and the ε_r values are changed between 1-10. The S_{21}, S_{31} and S_{41} parameters obtained from the simulation with different MUTs are shown in Figure 5. It is understood from Figure 5 that as the ε_r value of the MUT increases, the f_{rMUT} values get closer to each other. The reason for this is the decrease in sensitivity at high ε_r values. The f_r and $S(\%)$ values for each MUT are given in Table 3.
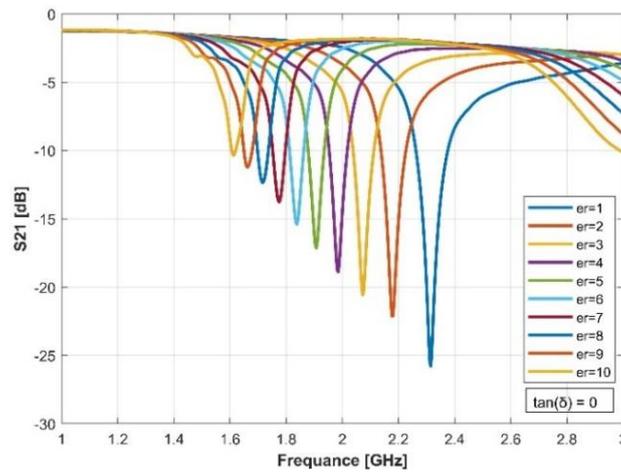
(a)
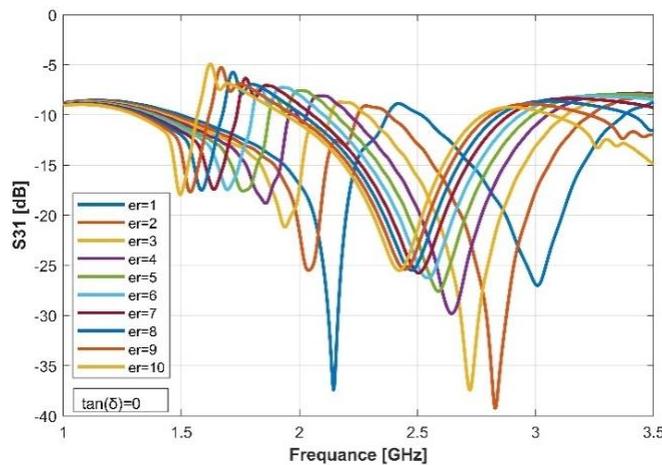
(b)
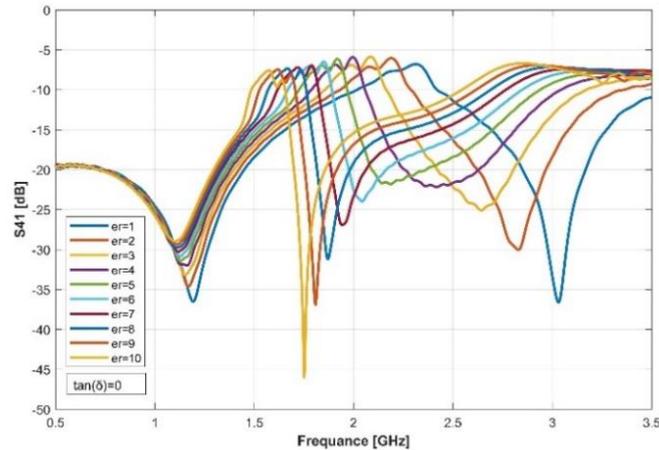
(c)

Figure 5: Variation of a) S21 parameter b) S31 parameter c) S41 parameter after loading the sensor with different MUTs with ε_r values ranging from 1 to 10

	S_{21}		S_{31}				S_{41}			
ε_r	f_r (GHz)	$S(\%)$	f_{r1} (GHz)	$S_{fr1}(\%)$	f_{r2} (GHz)	$S_{fr2}(\%)$	f_{r1} (GHz)	$S_{fr1}(\%)$	f_{r2} (GHz)	$S_{fr2}(\%)$
1	2.315	-	2.144	-	3.008	-	1.193	-	3.032	-
2	2.177	5.96	2.039	4.90	2.828	5.98	1.169	2.01	2.828	6.73
3	2.072	5.25	1.940	4.76	2.723	4.73	1.151	1.76	2.642	6.43
4	1.985	4.75	1.856	4.48	2.645	4.02	1.160	0.92	2.417	6.76
5	1.907	4.40	1.760	4.48	2.588	3.49	1.127	1.38	2.186	6.97
6	1.838	4.12	1.694	4.20	2.543	3.10	1.124	1.16	2.042	6.53
7	1.775	3.89	1.637	3.94	2.504	2.80	1.118	1.04	1.943	5.99
8	1.715	3.7	1.586	3.72	2.474	2.54	1.109	1.01	1.871	5.47
9	1.661	3.53	1.538	3.53	2.447	2.33	1.112	0.85	1.808	5.05
10	1.613	3.37	1.496	3.36	2.423	2.16	1.103	0.83	1.571	5.35

When Table 3 is examined, it becomes clear that sensor sensitivity is high at the resonant frequency of S_{21}, the first resonant frequency of S_{31}, and the second resonant frequency of S_{41}. Notably, the sensor sensitivity at the second resonant frequency of S_{41} is higher than the others.

CONCLUSION

In this study, a new microwave sensor design is proposed using the CST Studio EM simulation tool, which incorporates a directional coupler with a -10 dB coupling factor and CSRR structures. The microwave sensor operates on the principle of resonance frequency shifting due to the interaction between resonator structures and the Material Under Test (MUT) within the 1.5 – 3 GHz range. In resonator-based microwave sensors, sensitivity is a key parameter that determines the accuracy of the sensor's measurements. To enhance sensitivity and evaluate the sensor's performance with several ports, a microwave sensor featuring a directional coupler-based CSRR structure was designed in a simulation environment. Analysis of the data obtained

from the transmission, coupling, and isolation ports revealed that the sensitivity of the isolation port was higher than that of the other ports. The proposed sensor is anticipated to provide significant benefits for commercial applications.

REFERENCES

[1] L. Ali et al., "Design and Optimization of Microwave Sensor for The Non-Contact Measurement of Pure Dielectric Materials," Electronics (Switzerland), vol. 10, no. 24, Dec. 202.

[2] K. Saeed, M. F., M. B., and I. C., "Planar Microwave Sensors for Complex Permittivity Characterization of Materials and Their Applications," in Applied Measurement Systems, InTech, 2012

[3] S. Subbaraj, V. S. Ramalingam, M. Kanagasabai, E. F. Sundarsingh, Y. P. Selvam, and S. Kingsley, "Electromagnetic Nondestructive Material Characterization of Dielectrics Using EBG Based Planar Transmission Line Sensor," IEEE Sens J, vol. 16, no. 19, pp. 7081–7087, Oct. 2016.

[4] M. Saadat-Safa, V. Nayyeri, M. Khanjarian, M. Soleimani, and O. M. Ramahi, "A CSRR-Based Sensor for Full Characterization of Magneto-Dielectric Materials," IEEE Trans Microw Theory Tech, vol. 67, no. 2, pp. 806–814, Feb. 2019.

[5] S. Kiani, P. Rezaei, M. Navaei, and M. S. Abrishamian, "Microwave Sensor for Detection of Solid Material Permittivity in Single/Multilayer Samples with High Quality Factor," IEEE Sens J, vol. 18, no. 24, pp. 9971–9977, Dec. 2018.

[6] M. Yıldırım And M. A. Gözel, "Asimetrik Eş-Düzlemsel Şerit Beslemeli Anten ile Motor Yağ Seviye ve Kullanım Ömrü Tespiti," Mühendislik Bilimleri ve Tasarım Dergisi, Vol. 11, No. 3, Pp. 904–915, Sep. 2023.

RESUME

İbrahim GENÇ

İbrahim GENÇ was born in Artvin in 1978. Since 2008, he has been working as a lecturer at Isparta University of Applied Sciences Yalvaç Vocational School of Technical Sciences Department. He completed his undergraduate education at Selçuk University, Department of Electrical and Electronics Engineering in 2005 and his master's degree at Süleyman Demirel University, Department of Electronics and Communication Engineering in 2015. Genç continues his PhD education at Süleyman Demirel University, Department of Electronics and Communication Engineering.

Assist. Prof. Dr. Mahmut Ahmet GÖZEL

Mahmut Ahmet Gözel was born in Istanbul in 1987. He received his B.Sc. degree in electronics and communication engineering in Suleyman Demirel University, Isparta, Turkey in 2011, and then received M.Sc. and Ph.D. degrees in 2015 and 2020 respectively in the same university. He is currently a assistant professor in Department of Electrical and Electronics Engineering, Suleyman Demirel University, Isparta, Turkey. His research interests include antennas, RF energy harvesting, RFID, microwave active and passive circuits design, broadband technologies and microwave resonator sensor for material characterization.

Prof. Dr. Mesud KAHRİMAN

Mesud Kahriman was born in Germany in 1976. He received his BS, MS and PhD degrees from the Department of Electrical and Electronics Engineering at Sakarya University in 1997, 2001 and 2004 respectively. He worked at the R&D Department of the University of Applied Sciences in Zurich, Switzerland in 2002. Kahriman has studies on Electromagnetic Theory, EMC and Transmission Lines.

ANALYZING THE BOOST CONVERTER CIRCUITS AND SWITCHING TYPES

Seyyid Mustafa KILINÇARSLAN[1], Hasan Erdinç KOÇER[2]

INTRODUCTION

Electric vehicle technology is a technology used by all vehicles with electric motors. With this technology, it is aimed to reduce the negative impact of greenhouse gases on the world. It is aimed to minimize the consumption of fossil resources and increase the use of renewable energy resources. With the spread of electric vehicles, the electrical and electronic equipment used in the vehicle has increased. Therefore, more electricity consumption is required. Charging technology that is not planned in a efficiently way reduces the use of electric vehicles. Considering the studies carried out today, charging technology is as important as electric vehicle technology.

For an efficient charging technology, the charging stations outside the vehicle are as important as the equipment inside the vehicle. The features of the charging stations should vary according to the use of the vehicles. There are AC and DC charging stations as well as wired and wireless charging station technologies.

AC charging technology has lagged behind DC charging technology. Charging stations powered by AC sources have considerably longer charging times. These charging stations are divided into 4 levels according to their source feeding methods[1, 2]. Of these levels, only Level 3 and Level 4 charging stations can provide fast charging. Level 3 charging stations have AC source feeding, expensive installation costs and inadequate infrastructure. Commonly used Level 4 charging stations are DC-powered charging stations. Wireless charging stations are not very common today. Since the efficiency of wireless charging stations is very low, they are still in the research and development phase. Wired charging stations are preferred more due to their fast charging feature. This research has been conducted on DC-wired charging stations.

We need DC-powered charging stations for fast charging of vehicles. The system capacity of electric vehicle charging stations used today is generally between 100kW and 200kW. We need charging stations with a system capacity of approximately 300kW for the minimum charging times of vehicles. The output voltages of charging stations with a system capacity of 300kW are at 1000V levels. In this research, calculations were made for a system capacity of 300kW and an output voltage of 1000V. In order to make these calculations, we need input reference raiting. The necessary calculations were made and analyzed on different circuits to reach the targeted system capacity with the selected input reference raitings.

[1] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0000-0001-7121-4733 , mustafaklncarslan@hotmail.com
[2] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0000-0002-0799-2140 , ekocer@selcuk.edu.tr

The analyzed circuits are divided into two as 1-stage boost converter circuit and 2-stage boost converter circuit. Different mathematical calculations are made for both circuits and the output raitings are observed. The biggest difference between these two circuits is cost, size and higher components rating. In general, 1-stage boost converter circuit is used for systems with low power level and 2-stage boost converter circuit is used for systems with higher power level. In addition to these differences, there are also hard switching and soft switching circuits. The most obvious difference between these two circuits is power losses. Power losses are tried to be minimized with soft switching. In this study, mathematical calculations for hard switching and soft switching continue. Therefore, circuits with various switching types are exemplified in this presentation.

1 Stage Boost Converter Circuit

The 1 stage boost converter circuit is a circuit with a single boost stage. It is a circuit that increases the raitings taken from the input to the output by using a transistor type switching method. The most important advantage of this circuit is that it is low cost. At the same time, the circuit dimensions are not very large. The disadvantage is that this circuit can only be used in systems that work with low power. As the power raitings increase, the component raitings make the physical design of the circuit difficult. There is no component rating problem for this circuit, which is preferred for low power raitings. The 1 stage boost converter circuit analysis made in this study shows why this circuit is not suitable for this system.

Circuit
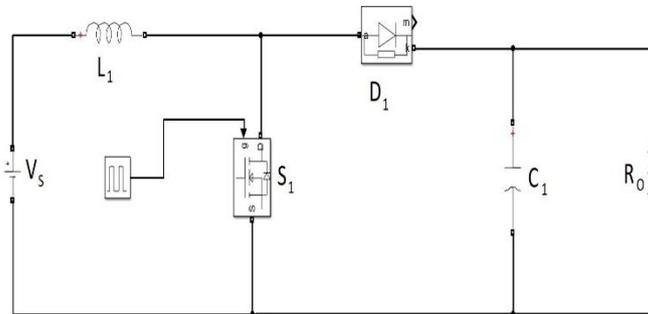
Figure 1: 1 stage boost converter circuit

The 1 stage boost converter circuit analyzed in this study is shown in Figure 1. This circuit operates in 2 time intervals according to the closed and open positions of the S_1 switch.

Calculations

In these circuit analysis calculations, mathematical equations described by Texas Instruments were used [3].

Reference raitings used in this study;

$V_{out} = 1000V$

$I_{out} = 300A$

$V_{in} = 380V$

$f_s = 50000Hz$

$n = 0,95$

The output power raiting is calculated in equation 1.

$$P_{out} = V_{out} \times I_{out}$$
$$P_{out} = 1000V \times 300A \tag{1}$$
$$P_{out} = 300kW$$

The input power raiting is calculated in equation 2.

$$P_{out} = n \times P_{in}$$
$$P_{in} = \frac{300kW}{0,95} \tag{2}$$
$$P_{in} = 315,79kW$$

The input current raiting is calculated in equation 3.

$$I_{in} = \frac{P_{in}}{V_{in}}$$
$$I_{in} = \frac{315,79kW}{380V} \tag{3}$$
$$I_{in} = 831A$$

The duty cycle raiting is calculated in equation 4.

$$\%D = 1 - \frac{V_{in}}{V_{out}} \tag{4}$$
$$\%D = 1 - \frac{380V}{1000V}$$
$$\%D = 0,62$$

The current that will pass through the S_1 transistor is calculated in equation 5.

$$I_{in} = I_L \qquad I_S = I_L \times D$$
$$I_S = 831A \times 0,62 \tag{5}$$
$$I_S = 515,23A$$

The maximum output current is calculated in equation 6.

$$I_{out(max)} = I_D = I_L \times (1 - D)$$
$$I_{out(max)} = 831A \times (1 - 0,62) \tag{6}$$
$$I_{out(max)} = 315,79A$$

In order to calculate the coil raiting, the inductor ripple current is calculated in equation 7. The estimate for the inductor ripple current is between 20% and 40% of the output current. In this study, 20% is taken as reference.

$$\Delta_{IL} = (0,2 \; to \; 0,4) \times I_{out(max)} \times \frac{V_{out}}{V_{in}}$$

$$\Delta_{IL} = 0,2 \; \times 315,79A \times \frac{1000V}{380V} \tag{7}$$

$$\Delta_{IL} = 166,18$$

The coil raiting is calculated in equation 8.

$$L = \frac{V_{in} \times (V_{out} - V_{in})}{\Delta_{IL} \times f_s \times V_{out}}$$

$$L = \frac{380V \times (1000V - 380V)}{166,18 \times 50000Hz \times 1000V} \tag{8}$$

$$L = 29,5 \times 10^{-6} \; H$$

To calculate the capacitor raiting, the output ripple voltage is calculated in equation 9.

$$\Delta_{VO} = \frac{I_{out(max)}}{1 - D} + \frac{\Delta_{IL}}{2}$$

$$\Delta_{VO} = \frac{315,78A}{0,38} + \frac{166,18}{2} \tag{9}$$

$$\Delta_{VO} = 914,09$$

The capacitor raiting is calculated in equation 10.

$$C = \frac{I_{out(max)} \times D}{\Delta_{VO} \times f_s}$$

$$C = \frac{315,79A \times 0,62}{914,09 \times 50000Hz} \tag{10}$$

$$C = 4,28 \times 10^{-6} \; F$$

The load raiting is calculated in equation 11.

$$R = \frac{V_{out}}{I_{out}}$$

$$R = \frac{1000V}{315,79A} \tag{11}$$

$$R = 3,16\Omega$$

Analysis

The circuit with the calculated raitings was analyzed in the matlab program.
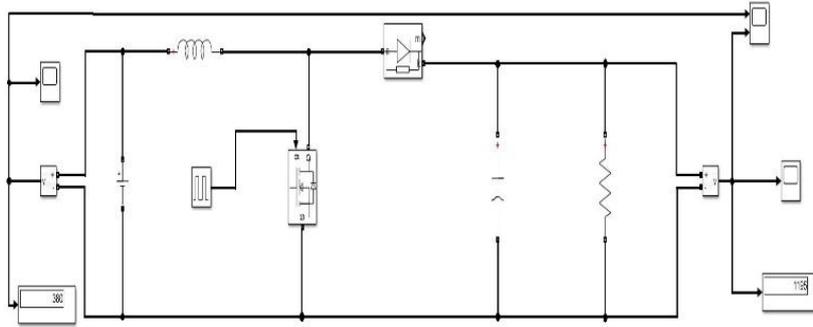
Figure 2: 1 stage boost converter circuit analysis diagram

In order to analyze the circuit shown in Figure 1, parallel measurement instruments were connected to the input and output terminals as shown in Figure 2. The outputs of this analysis are shown in Figure 6.

As a result of the analysis, we observe that the output voltage ripple between 480V and 1200V. This ripple range does not fall into the optimum ripple range during the analyzed time period.

The reason for this ripple is that the 1 stage boost converter circuits are not suitable for the high powers used in this study. This circuit, which works efficiently at low powers, is not preferred for high power systems.

2 Stage Boost Converter Circuit

The 2 stage boost converter circuit is a circuit with a double boost stage. The raitings taken from the input are first boosted in the first stage and then in the second stage to obtain the output raiting. Transistor type switching is also used in this circuit. The most important advantage of this circuit is that it is preferred for high power raitings. Its disadvantages are the increase in cost and the increase in circuit dimensions. This circuit, which is preferred in systems with high power raitings, provides a more efficient result. It ensures that the ripple are minimized. The 2 stage boost converter circuit analysis made in this study shows that this circuit is one of the ideal circuits for this system.
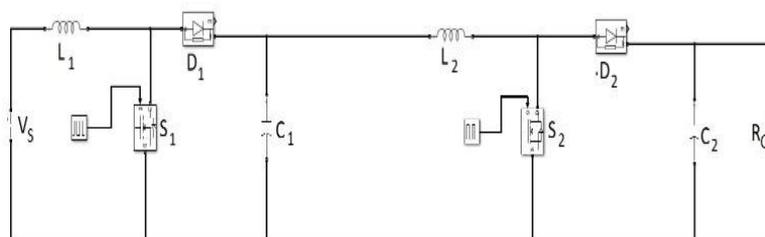
Circuit

Figure 3: 2 stage boost converter circuit

The 2 stage boost converter circuit analyzed in this study is shown in Figure 3. This circuit operates in 2 time intervals according to the closed and open positions of the S_1 and S_2 switches.

Calculations

In the calculation of this circuit analysis, the equations described by Texas Instruments were adopted[3].

Reference raitings used in this study;

$V_{out} = 1000V$

$I_{out} = 300A$

$V_{in} = 380V$

$f_s = 50000Hz$

$n_1 = 0,95$

$n_2 = 0,95$

$\%D_1 = 0,40$

$\%D_2 = 0,40$

The output power raiting is calculated in equation 12.

$$\begin{aligned} P_{out} &= V_{out} \times I_{out} \\ P_{out} &= 1000V \times 300A \\ P_{out} &= 300kW \end{aligned} \tag{12}$$

The input power raiting is calculated in equation 13.

$$\begin{aligned} P_{out} &= n_1 \times n_2 \times P_{in} \\ P_{in} &= \frac{300kW}{0,95 \times 0,95} \\ P_{in} &= 332,5kW \end{aligned} \tag{13}$$

The first stage output power raiting is calculated in equation 14.

$$\begin{aligned} P_{out1} &= n_2 \times P_{in} \\ P_{out1} &= \frac{332,5kW}{0,95} \\ P_{out1} &= 315,87kW \end{aligned} \tag{14}$$

The input current raiting is calculated in equation 15.

$$\begin{aligned} I_{in} &= \frac{P_{in}}{V_{in}} \\ I_{in} &= \frac{332,5kW}{380V} \\ I_{in} &= 875A \end{aligned} \tag{15}$$

The first stage output voltage raiting is calculated in equation 16.

$$V_{out1} = \frac{V_{in}}{1 - D_1} \tag{16}$$

$$V_{out1} = \frac{380V}{1 - 0,40}$$
$$V_{out1} = 633,3V$$

The 2nd stage output voltage raiting is calculated in equation 17.

$$V_{out2} = \frac{V_{out1}}{1 - D_2}$$
$$V_{out1} = \frac{633,3V}{1 - 0,40}$$
$$V_{out1} = 1055,5V$$

(17)

The D_1 and D_2 duty cycle raiting are calculated in equation 18.

$$\%D_1 = 1 - \frac{V_{in}}{V_{out1}}$$
$$\%D_1 = 1 - \frac{380V}{633,3V}$$
$$\%D_1 = 0,399$$
$$\%D_2 = 1 - \frac{V_{out1}}{V_{out2}}$$
$$\%D_2 = 1 - \frac{633,3V}{1055,5V}$$
$$\%D_2 = 0,40$$

(18)

The first stage output current raiting is calculated in equation 19.

$$I_{out1} = \frac{P_{out1}}{V_{out1}}$$
$$I_{out1} = \frac{315,78kW}{633,3V}$$
$$I_{out1} = 498,76A$$

(19)

The current that will pass through the S_1 transistor is calculated in equation 20.

$$I_{in} = I_{L1} \qquad I_{S1} = I_{L1} \times D_1$$
$$I_{s1} = 875A \times 0,4$$
$$I_{s1} = 350A$$

(20)

The maximum 1st stage output current raiting is calculated in equation 21.

$$I_{out1(max)} = I_{D1} = I_{L1} \times (1 - D_1)$$
$$I_{out1(max)} = 875A \times (1 - 0,4)$$
$$I_{out1(max)} = 525A$$

(21)

The current that will pass through the S_2 transistor is calculated in equation 22.

$$I_{out1} = I_{L2} \qquad I_{S2} = I_{L2} \times D_2$$
$$I_{s2} = 498{,}76A \times 0{,}40 \tag{22}$$
$$I_{s2} = 200A$$

The maximum 2nd stage output current raiting is calculated in equation 23.

$$I_{out2(max)} = I_{D2} = I_{L2} \times (1 - D_2)$$
$$I_{out2(max)} = 498{,}76A \times (1 - 0{,}40) \tag{23}$$
$$I_{out2(max)} = 300A$$

To calculate the L_1 coil raiting, the inductor ripple current is calculated in equation 24. The L_1 coil raiting is calculated in equation 25.

$$\Delta_{IL1} = (0{,}2 \; to \; 0{,}4) \times I_{out1(max)} \times \frac{V_{out}}{V_{in}}$$
$$\Delta_{IL1} = 0{,}2 \; \times 325A \times \frac{633{,}3V}{380V} \tag{24}$$
$$\Delta_{IL1} = 175$$
$$L_1 = \frac{V_{in} \times (V_{out} - V_{in})}{\Delta_{IL} \times f_s \times V_{out}}$$
$$L = \frac{380V \times (633{,}3V - 380V)}{175 \times 50000Hz \times 633{,}3V} \tag{25}$$
$$L = 17{,}3 \times 10^{-6} \; H$$

To calculate the L_2 coil raiting, the inductor ripple current is calculated in equation 26. The L_2 coil raiting is calculated in equation 27.

$$\Delta_{IL2} = (0{,}2 \; to \; 0{,}4) \times I_{out2(max)} \times \frac{V_{out}}{V_{in}}$$
$$\Delta_{IL2} = 0{,}2 \; \times 300A \times \frac{1055{,}5V}{633{,}3V} \tag{26}$$
$$\Delta_{IL2} = 100$$
$$L_2 = \frac{V_{in} \times (V_{out} - V_{in})}{\Delta_{IL} \times f_s \times V_{out}}$$
$$L = \frac{633{,}3V \times (1055{,}5V - 633{,}3V)}{100 \times 50000Hz \times 1055{,}5V} \tag{27}$$
$$L = 50{,}6 \times 10^{-6} \; H$$

To calculate the C_1 capacitor raiting, the output ripple voltage is calculated in equation 28. The C_1 capacitor raiting is calculated in equation 29.

$$\Delta_{VO} = \frac{I_{out1(max)}}{1 - D} + \frac{\Delta_{IL1}}{2}$$
$$\Delta_{VO} = \frac{525A}{0{,}60} + \frac{175}{2} \tag{28}$$

$$\Delta_{VO} = 962,5$$
$$C_1 = \frac{I_{out1(max)} \times D}{\Delta_{VO} \times f_s}$$
$$C = \frac{525A \times 0,40}{962,5 \times 50000Hz} \tag{29}$$
$$C = 4,36 \times 10^{-6} \, F$$

To calculate the C_2 capacitor raiting, the output ripple voltage is calculated in equation 30. The C_2 capacitor raiting is calculated in equation 31.

$$\Delta_{VO} = \frac{I_{out2(max)}}{1 - D} + \frac{\Delta_{IL2}}{2}$$
$$\Delta_{VO} = \frac{300A}{0,60} + \frac{100}{2} \tag{30}$$
$$\Delta_{VO} = 550$$
$$C_2 = \frac{I_{out2(max)} \times D}{\Delta_{VO} \times f_s}$$
$$C = \frac{300A \times 0,40}{550 \times 50000Hz} \tag{31}$$
$$C = 4,36 \times 10^{-6} \, F$$

The load raiting is calculated in equation 32.

$$R = \frac{V_{out}}{I_{out}}$$
$$R = \frac{1055,5V}{300A} \tag{32}$$
$$R = 3,51\Omega$$

Analysis

The circuit with the calculated raitings was analyzed in the matlab program.
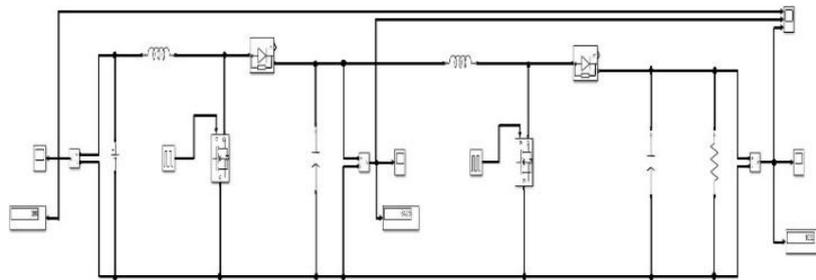

Figure 4: 2 stage boost converter circuit analysis diagram

In order to analyze the circuit shown in Figure 3, parallel measurement instruments were connected to the input terminal, the output terminal of the 1st stage, and the output terminals of the 2nd stage as shown in Figure 4. The outputs of this analysis are shown in Figure 7.

As a result of the analysis, we observe that the first stage output voltage ripple between 610V and 650V. We observe that the second stage output voltage ripple between 1040V and 1060V. These ripple ranges are the optimum ripple ranges obtained during the analyzed time period. The raitings obtained as a result of the analysis are consistent with the calculated raitings. The targeted output raitings for this system were obtained with this circuit.

Hard Switched and Soft Switched Circuit

The process of a semiconductor power element turning on and off is called switching. When switching is done, power losses occur due to the overlapping of the voltage and current of the power element. In addition, power losses can also occur on the diodes and capacitors in the circuit. All these power losses occur as a result of switching without additional regulation of the circuit. This switching process is called hard switching.

Additional arrangements are made to the circuit to prevent these power losses. The switching process that occurs as a result of these arrangements is called soft switching. With soft switching, the efficiency and performance of the circuit are increased.

One of the advantages of soft switching is that it reduces the size and component rating of passive components such as inductors and capacitors used in the circuit. Moreover, the reduction in losses also reduces heat generation. In this way, the size of the cooling unit will be reduced. The reduction in the size of the cooling unit will reduce the cost. Another advantage is that it also reduces the noise generated by the power electronics.

The disadvantages of soft switching are that new components will be added to the circuit because additional arrangements will be made. Therefore, the circuit dimensions will increase. Another disadvantage is that the circuit cost will increase due to the new components to be added to the circuit. The biggest disadvantage is the difficulty of design. The control complexity and calculations of the new components make the design difficult.

A basic soft switching circuit example is shown in Figure 5[4]. In addition to the number of components used in the 2 stage boost converter circuit, there is 1 coil, 1 capacitor and 2 diodes. This soft switching circuit has a total of 12 operating modes. There are separate mathematical calculations for each operating mode. All these situations complicate the design.
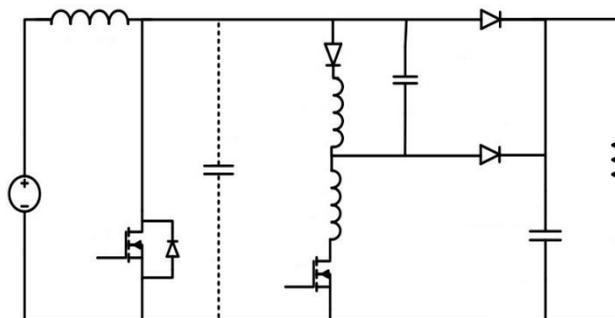
Figure 5: Soft switching circuit example

The circuit design and calculations to be analyzed in the thesis study are ongoing. It is planned to use soft switching technique in this circuit design. Calculations are made using determined reference raitings.
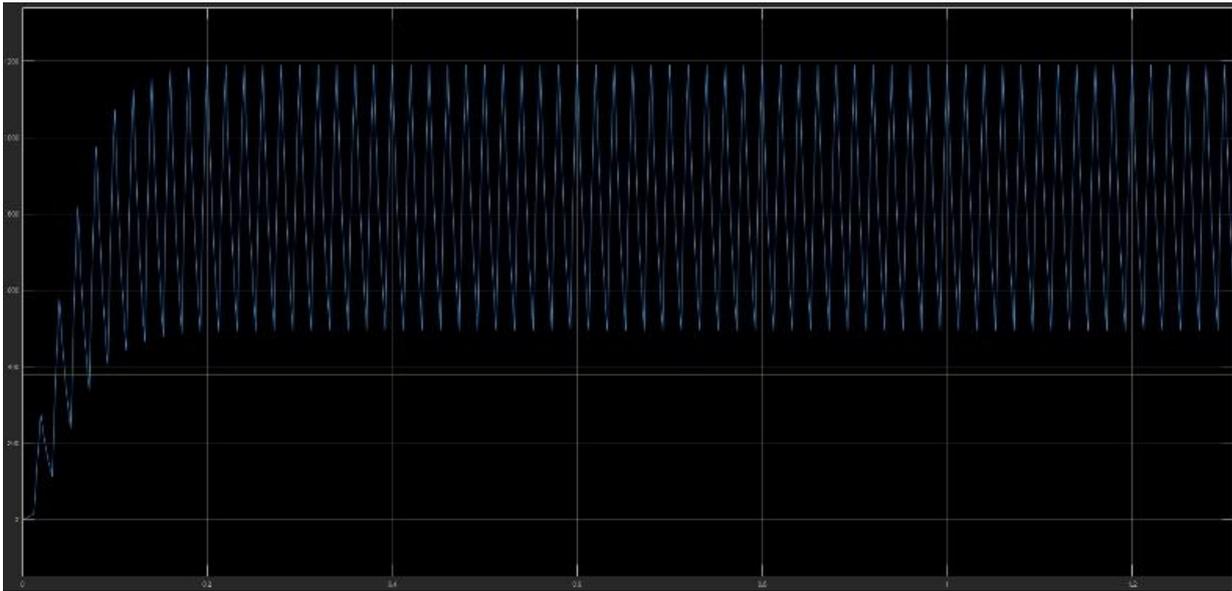
Figure 6: 1 stage boost converter circuit analysis output
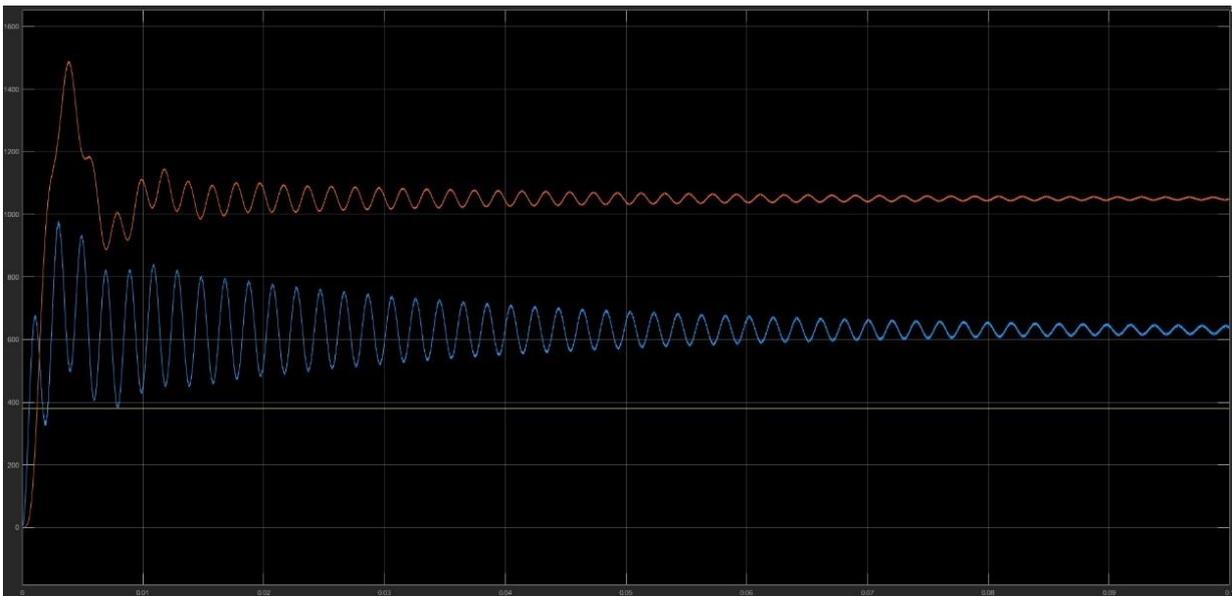
Figure 7: 2 stage boost converter circuit analysis output

CONCLUSION

In this paper, 1 stage boost converter circuit and 2 stage boost converter circuit were analyzed. 300kW power raiting was targeted as output power raiting. Determined reference raitings were used to obtain this power raiting. Calculations were made for both circuits and analysis was performed in matlab program.

It has been observed that 1 stage boost converter circuit is not suitable for the targeted output power raiting. It has been observed that 2 stage boost converter circuit is more efficient for high power raitings.

The advantages and disadvantages of hard switching and soft switching techniques have been research. It has been observed that power losses are high in circuits where hard switching technique is used. It has been understood that circuits designed with soft switching technique operate more efficiently.

In this study, it is planned to design a circuit using the reference raitings and soft switching technique to obtain the targeted output power raitings. Circuit design and calculations are ongoing.

REFERENCES

[1] George Fernandez Savari, M.J.S., L. Anantha Raman Adel El-Shahat, Hany M. Hasanien, Dhafer Almakhles, Shady H.E. Abdel Aleem, Ahmed I. Omar, Assessment of charging technologies, infrastructure and charging station recommendation schemes of electric vehicles: A review. 2022.

[2] Muhammad Shahid Mastoi, S.Z., Hafiz Mudassir Munir, Malik Haris, Mannan Hassan, Muhammad Usman, Syed Sabir Hussain Bukhari, Jong-Suk Ro, An in-depth analysis ofelectric vehicle charging station infrastructure, policy implications, and future trends. 2022.

[3] Hauke, B., Basic Calculation of a Boost Converter's Power Stage. 2022.

[4] Yakup ŞAHİN, İ.A., Naim Süleyman TINĞ, An Improved ZCZVT-PWM DC-DC Boost Converter. 2015.

RESUME

Seyyid Mustafa KILINÇARSLAN

Seyyid Mustafa KILINÇARSLAN graduated from Selçuk University, Faculty of Technology, Department of Electrical and Electronics Engineering in 2021. Kılınçarslan continues his master's degree at Selçuk University, Institute of Science. His main research interests include electric vehicles, charging stations, smart systems, power electronics, and multi-layered card design.

Prof. Dr. Hasan Erdinç KOÇER

Hasan Erdinç KOÇER, received the Ph.D. degree from Department of Electric–Electronic Engineering, Institute of Science, Selçuk University, Konya, Turkey (2007), where he is a professor since 2022. His research interests are in machine vision and pattern recognition, industrial automation on camera, medical image processing and, automation and control.

REMOTE CONTROL OF MICROCONTROLLER BASED ELECTRICAL DEVICES USING DTMF SIGNALING OVER FIXED TELEPHONE

Şerafetdin BALOĞLU[1]

INTRODUCTION

Telephone remote control systems allow control operation over existing telephone lines with any telephone. The advantage of telephone remote control systems is that instead of designing a line or a system to send the control signal required for the control of remote devices and to receive this control signal from the other side, the control process is carried out by sending the control signal over the existing telephone lines. Remote access to devices in the home is gaining momentum and developing rapidly on a global scale [1, 2].

Telephone exchanges use analogue and digital technology according to the way they work. Pulse Amplitude Modulation (PAM) method is used in analogue technology and Pulse Code Modulation (PCM) method is used in digital technology. In both methods, it is generally aimed to serve a large number of subscribers by using a small number of common elements. This is called Time Division Multiplexing (TDM). In PCM technology, samples are taken from the audio signal. For example, if we take a signal in the range of -2.5V to +2.5V, the value of the samples can be any value in the PAM system, while this is not allowed in PCM. The signal range is divided into, for example, 256 ranges and the samples can only take one of these 256 values. These 256 different values are expressed in binary numbers. 1s and 0s can be represented electrically by 0V and -5V. These binary numbers, not analogue signals, circulate in the switchboard. However, these digital signals are converted to analogue when sent to the other subscriber.

Since 0V is perceived as -5V and -5V as 0V and distortion is difficult, the sound quality is high in PCM switchboards. Analogue systems are heavily distorted. Because the most easily distorted feature of a signal in the transmission environment is its amplitude. In order to achieve a certain quality in analogue plants or systems, it is necessary to spend a lot of power and prepare very complex circuits. Signal to Noise (S/N) ratio can be used to determine the quality of a system. In a PCM system a S/N ratio of 20dB is sufficient to minimise error, whereas in PAM a S/N ratio of at least 60dB is required. In contrast, in a PCM system, a 64kHz band is consumed for a 3100Hz audio band. In other words, there is a gain in S/N ratio but a loss in the frequency band.

An important benefit of digital switchboards is that they can be used for switching voice as well as video and data at the same time. Today, efforts are being made to combine various forms of

[1] Seydişehir District Directorate of National Education, Konya/Turkey, Orcid: https://orcid.org/0000-0002-6541-4023 , serefbal@gmail.com

communication, e.g. voice, video and data, under one roof. The resulting network is called the Integrated Services Digital Network (ISDN). It is very difficult for analogue exchanges to adapt to this network. Digital switchboards, on the other hand, can provide these services and adapt to ISDN with minor changes.

Telephones can send pressed keys to switchboards in two ways. In the first method called Dial Pulse Signalling, the signal is sent by interrupting the current between the switchboard and the machine a certain number of times. For example, pressing the "1" key interrupts the current once, pressing the "2" key interrupts the current twice, etc. The current interruption is done either mechanically or by special integrated circuits. The disadvantages of this type of signalling are that it is slow (on average one key information can be sent per second) and the pulses can be distorted by noise on the line, resulting in the wrong number being dialed.

In DTMF, the second and more modern method, the current is not interrupted. For each key, two waves of different frequencies are generated and sent. DTMF was developed in 1963 in the Bell system in the USA [3].

LITERATURE REVIEW

Oluwole et al. designed a GSM based automatic motor control and protection system. In the design, a microcontroller based security gate is proposed using DTMF and GSM technology. [3].

In their study, Bawankule et al. designed a DTMF based home automation system using a microcontroller with a portable power supply. When a user anywhere presses a key on his/her GSM or landline phone, a corresponding DTMF tone is generated which is transmitted to the mobile phone at home. Depending on the frequency of the DTMF tone received at the receiver, the decoder table is consulted and the relevant device is switched on or off depending on the command given by the user. The fact that any unauthorised person can access the devices in the house and the lack of feedback can be shown as disadvantages of the system [4].

In their study, Wakas and Waseem designed a home automation project using DTMF. The disadvantages are that unauthorised persons can access the devices in the home, the uncertainty of whether the task is fulfilled when the relevant key is pressed due to the lack of feedback, and the need for a mobile phone to activate the system. [5].

Choudhari and Rucha presented a paper focusing on the design of a home automation system with DTMF. ATmega8 microcontroller and HT9107B decoder were used. The system decodes the DTMF signals received from the mobile phone headphone jack and controls the light bulbs used as loads. [6].

In their study, Mamunoor and Mehdi controlled household electrical appliances using DTMF with a mobile phone.[7]. A disadvantage of the system is the lack of a password authentication function for unauthorised persons and the lack of feedback so that it is not known whether the given command has been executed or not.

Rangkuti and Simatupang proposed a locking automation system using a smartphone polyphonic tone sensor. The DTMF signal was interpreted with Arduino Uno and the load was

deactivated 3 s after the load was activated. [8]. The lack of password against unauthorised persons and short working interval are disadvantages of the system.

In their study, Shirsath et al. designed an embedded wireless light intensity control using DTMF. In the designed system, the light intensity of the AC lamp load is controlled via a mobile phone. 89C51 was used as microcontroller [9]. The disadvantage of the system is that it can be easily accessed by unauthorised persons.

In this study, a circuit has been designed for the control of automatic irrigation systems over telephone lines, taking into account the agricultural lands in Turkey. The application was developed using a PIC microcontroller and the device was controlled by remote control methods. The project has provided advantages in terms of cost, reliability and ease of use, enabling effective management of irrigation systems. [10].

In his study, Ambarita used a DTMF decoder to control AC loads directly by means of a relay. There is a security vulnerability as no password is required in the study. [11].

MATERIAL AND METHOD

CM8870 DTMF Decoder

Figure 1: CM8870 Integration pin connections.

DTMF signals are generated by the DTMF encoder and converted to binary signals by DTMF decoders such as the CM8870. The main control board and the telephone line control board circuits control each other's operation with line control (RD) and line on/off (HK) signals. RD takes logic 0 value when there is a call signal on the line. This signal is processed in the programme and when the specified value is met, the HK signal is set to logic 1 and the telephone line is opened. The "dso" signal indicates whether a signal is sent to the line from the telephone keys after the line is switched on. When there is a signal on the line, dso becomes logic 1. Q1, Q2, Q3 and Q4 show the binary code of the signal sent to the line. The CM8870 integration performs the conversion of the signal sent to the telephone line into binary code. For example, for the 3 key, two sine waves in the range of 697-1477Hz are generated and sent. These sinusoidal waves are generated by special circuits and decoded by decoders such as CM8870. This method is much more reliable and faster. 10 digits can be sent in one second. Moreover, since the data sent is a wave at certain frequencies, there can be no wrong data transmission.

The DTMF decoder converts the input key frequency into binary data. The table below shows the frequency and binary equivalents of the key information.

Table 3: Key frequencies and binary equivalents [10].

Keypad	Frequency Range		Outputs			
	F (Low)	F (High)	Q0	Q1	Q2	Q3
1	697	1209	0	0	0	1
2	697	1336	0	0	1	0
3	697	1477	0	0	1	1
4	770	1209	0	1	0	0
5	770	1336	0	1	0	1
6	770	1477	0	1	1	0
7	852	1209	0	1	1	1
8	852	1336	1	0	0	0
9	852	1477	1	0	0	1
0	941	1209	1	0	1	0
*	941	1336	1	0	1	1
#	941	1477	1	1	0	0

In the oscilloscope examination, it was observed that the StD pin of the DTMF decoder maintains its state at logic- as long as the telephone key is pressed, and at logic-0 otherwise.

Optical and Magnetic Isolation Circuits

Optocoupler was used to detect the incoming call signal to the telephone. The electrical isolation between the telephone line and the control circuit is provided optically.

In Figure 3, at the optocoupler output used to detect the incoming call, voltage transitions from +5V to 0V occur during the call. The falling edges are counted with the microcontroller. The call presence led (RD) will normally be logic-1 continuously and will flash with a ripple when a call is received. This led is also the stand-by led of the circuit.

The telephone line voltage was measured as -46 V DC when idle and -7.56 V DC when the handset was lifted in the presence of a call signal.

When the handset was lifted when there was no call signal on the telephone line, the current on the telephone line was measured as 40mA and the voltage was measured as 7.56 V. $R = 7.56V/40 \times 10^{-3}$, R= 189 Ω. During the incoming call signal, the phone line will be switched on by using the resistance R with the help of a relay.
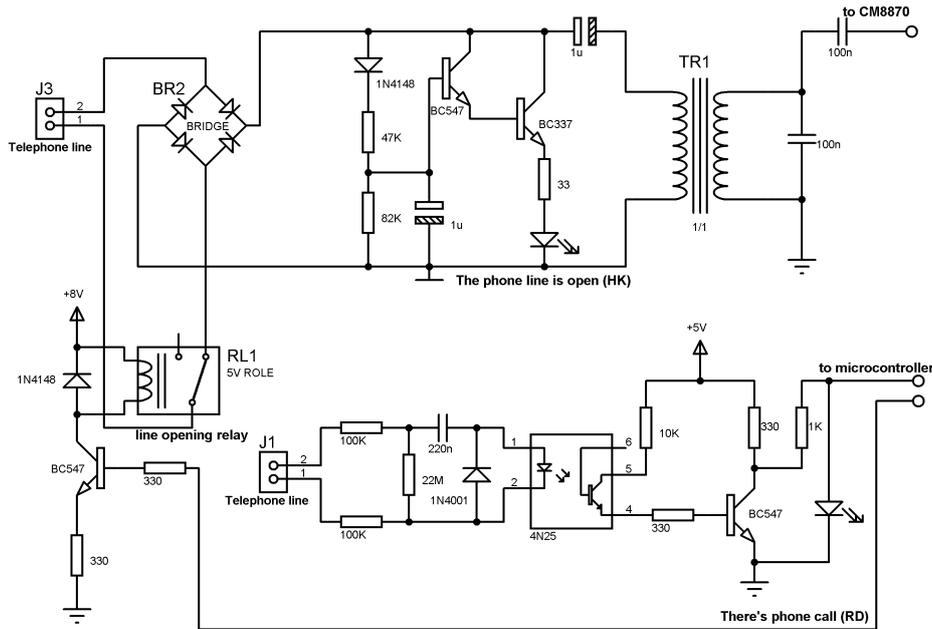
Figure 3: Optical and magnetic isolation circuits.

Key Information Retrieval

CM8870 integration is used to analyse the DTMF signals from the telephone line. If the connection is made directly, there will be an electrical connection between the telephone line and the DTMF integration and the control card is likely to be damaged due to the high voltage on the telephone line. In order not to encounter negative situations, the electrical connection between the telephone line and the DTMF integration is isolated by using the line transformer given in Figute 3 and magnetic isolation is provided. Input and output impedances are equal in the line transformer. The voltage applied to its input is taken exactly from its output.

Microcontroller

The PIC16F84A microcontroller from the widely used PIC family is a RISC architecture controller with 1K Flash program memory, 68-byte RAM memory, 64-byte EEROM memory, 14-bit instructions, interrupt sources, 13 input-output ports, 25 mA port output current, 2 mA at + 5V, 15 nA at 2V low power consumption. It can operate at a clock rate of up to 10MHz, the instruction count time is 400 ns.

System Operation

The programme loaded in the microcontroller waits for the RD signal. During the waiting period, the microcontroller listens to the RD signal. When the call signal comes to the telephone line, the programme in the microcontroller becomes active.

The microcontroller, which enters the operating mode, waits for the phone to ring for the number specified in the programme. In the application, this number is determined as 2. The 2nd ringing phone line is switched on. After this stage, it is checked whether the dso signal is logic 1. This control process is not unlimited. If these signals do not reach the expected value for about 15 seconds, the programme switches off the telephone line. At each logic 1 value of the dso pin, the microcontroller takes the binary values from the CM8870 outputs and converts

them to hexadecimal code and stores them in the RAM memory. With the application, the control of various devices is done by authorised persons in a secure way. The user must enter the given password when the telephone line is switched on. The password is 4 digits and stored in the EEPROM (the password is initially set to "1453"). The person calling the line to which the control card is connected must dial the password. The entered password is checked by the microcontroller. If the password is entered incorrectly, the programme switches off the telephone line. After the correct password entry, the keys related to the device to be controlled must be coded. How to control the devices and how to change the password Table 1 is given in
.

Table 1: Password entry, password change and device control.

Password Login	1st digit	2nd digit	3rd digit	4th digit
Change Password	Phase 1	Press the # key and enter the old password.		
	Phase 2	Enter the new 4-digit password.		
Device control	Opening		Closure	Exit
1. Device Control	1		2	*
2. Device Control	4		5	*
3. Device Control	7		8	*

RESULTS

In the application where microcontroller-based electrical devices connected to the fixed telephone line are remotely controlled, three relays are used on the designed board for the control of AC loads. For the security of the control system, a 4-digit long password is stored in the non-volatile memory (EEPROM) of the microcontroller. The password change and control flow diagram of the system is given in Figure 4 and the initial settings and branching flow diagram of the system is given in Figure 5.

The given scheme was designed and realised as a PCB board in proteus isis-ares environment. The image of the assembled board is given in Figure 6.

The open circuit diagram of the system for remote control of microcontroller based electrical devices via telephone line using DTMF signal is given in Figure 7.
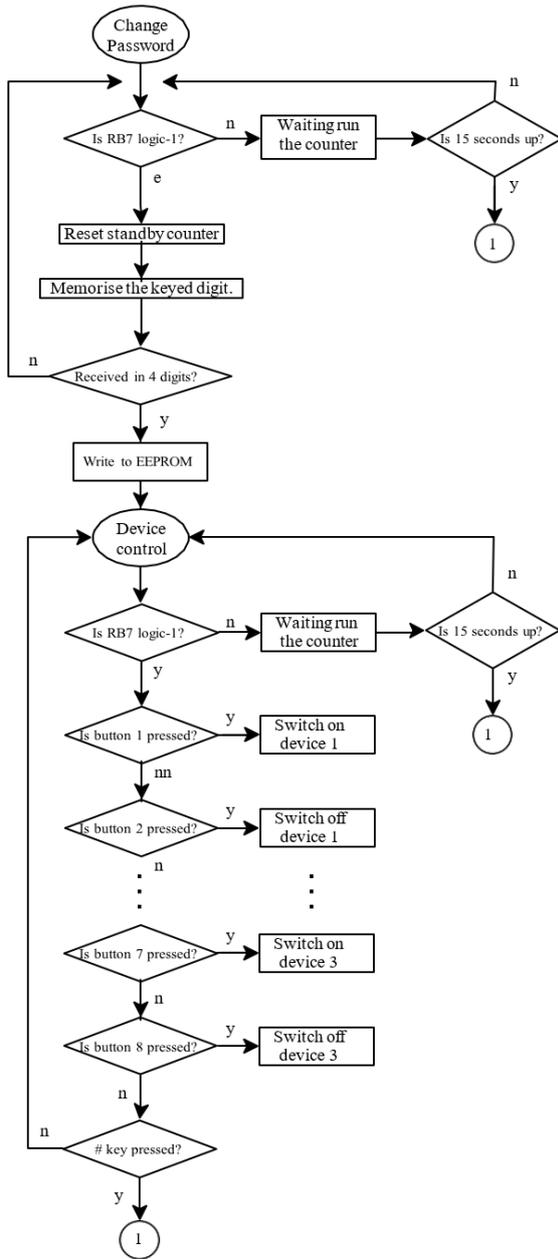
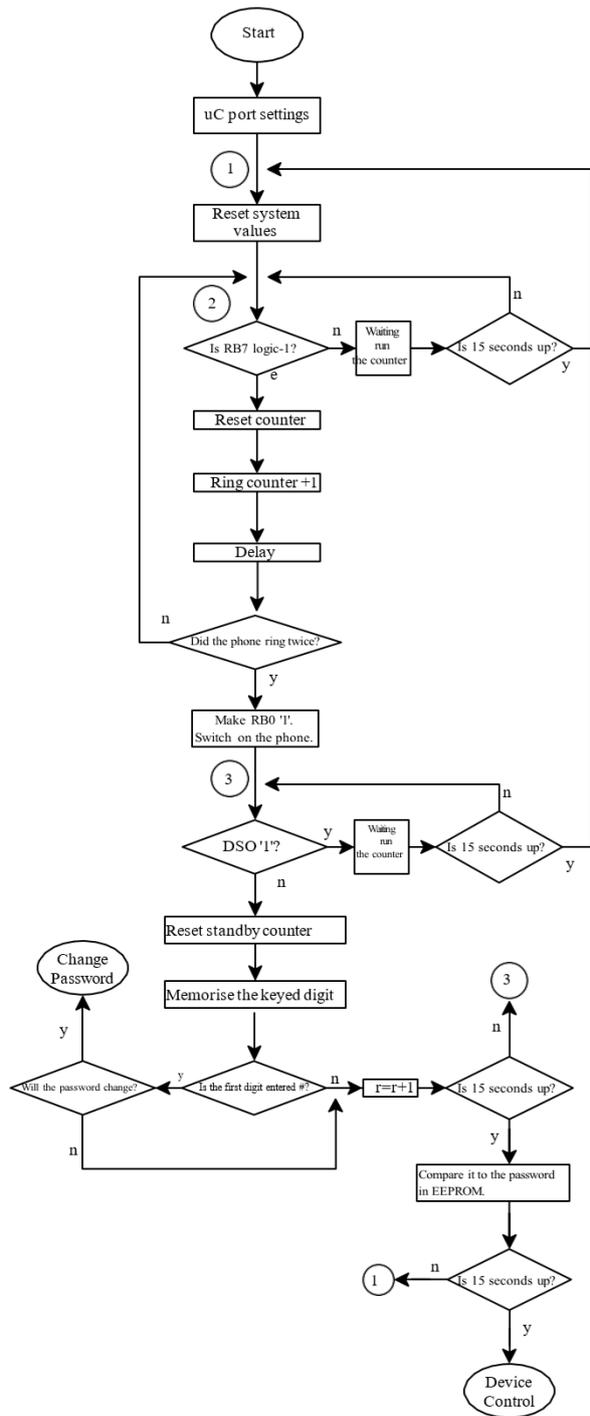Figure 4: Password change and control flow diagram.

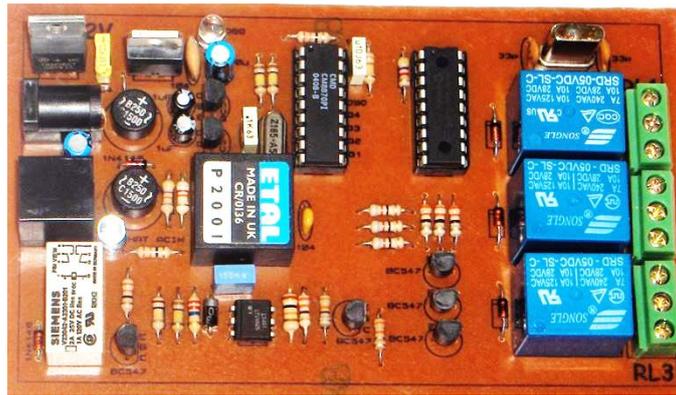Figure 5: Initial settings and branching flow diagram of the system.

Figure 6: Assembled remote control board.

CONCLUSION

The remote control of microcontroller-based electrical devices over the telephone line using the designed DTMF signal has been tested, and it has worked safely and stably. The developed system was tested on a fixed telephone line. The fact that the user can control the devices by entering a password at the beginning and can change the password over the telephone line has eliminated the deficiencies of the applications in the literature that have security vulnerabilities. If desired, this system can be adapted to the headphone jack input of a mobile phone and can be operated both on a fixed telephone line and on mobile phones.

In the future, a voice recording/playback chip can be included in this system and the user can be notified that the given command has been fulfilled audibly. In addition, by including a Caller-ID chip, the caller can control the system without entering a password.

The developed application can be adapted and used on different hardware such as Arduino (nano, mini, promini etc.), Raspberry Pi Zero.
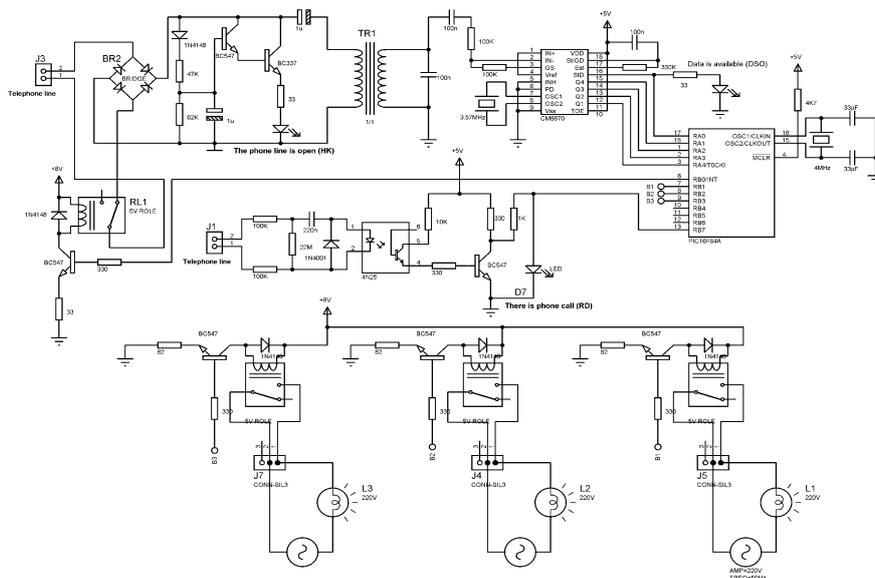
APPENDIX

Figure 7: Open circuit diagram of the developed remote control system.

REFERENCES

[1]   Darji, V., et al., Cell phone operated remote control using DTMF. *International Journal of Computer Applications*, 2018. **179**(35): p. 32-37.

[2]   Dodd, A.Z., *The essential guide to telecommunications*. 2002: Prentice Hall Professional.

[3]   Oluwole, A.S., O. Odekunle, and E. Olubakinde, Applications and recent development of DTMF based technology in home automation. *European Journal of Electrical Engineering and Computer Science*, 2021. **5**(3): p. 60-67.

[4]   L.M. Bawankule, A.V. Bhalerao, and F.K. Ashwini, Home Automation System using DTMF. *International Journal for Scientific Research & Development*, 2019. **7**(01).

[5]   A.Wakas and A. Waseem, DTMF based home automation. 2010: p. 1-2.

[6]   Jogdand, R.R. and B. Choudhari, DTMF based home automation system. *IJESC*, 2017. **7**(2).

[7]   Islam, M.M. and M.H. Chowdhury, DTMF Based Home Appliances Control Using Cell Phone. *Dept. of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology Chittagong-4349*, 2014: p. 1-3.

[8]   Rangkuti, H.A. and J.W. Simatupang. Security lock with DTMF polyphonic tone sensor. In *2015 International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT)*. 2015. IEEE.

[9]   Shirsath, S., et al., Embedded wireless light intensity control using DTMF. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, 2016. **4**(5): p. 57-59.

[10]   Gedikpınar, M. and M. Çavaş, Control of Irrigation Systems via Telephone Lines Using PIC Microcontroller. *Journal of Polytechnic*, 2005. **8**(3): p. 249-254.

[11]   Ambarita, R., Rancang Bangun Pengendali Peralatan Listrik Rumah Tangga Menggunakan Sistem DTMF Handphone. *Ocean Engineering: Jurnal Ilmu Teknik dan Teknologi Maritim*, 2022. **1**(3): p. 92-100.

RESUME

Dr. Şerafetdin BALOĞLU

Dr. Şerafetdin Baloğlu was born in 1978 in Tegelen, Netherlands. He graduated from Fırat University, Technical Education Faculty, Department of Electronics Teaching (2000); Fırat University, Institute of Science, Master's Program in Electronics-Computer Education (2003); Anadolu University, Faculty of Economics, Public Administration (2016, via Open Education); Necmettin Erbakan University, Faculty of Engineering and Architecture, Electrical-Electronics Engineering (2017); and Selçuk University, Institute of Science, Doctoral Program in Electrical-Electronics Engineering (2024). He has published a scientific article in a national journal, papers at international conferences, and a chapter in an international book. Since 2000, Dr. Baloğlu has been working as a teacher at the Ministry of National Education (MEB). He is married and has two daughters.

INNOVATIVE STUDIES IN ADVANCED MATERIALS, COMMUNICATION PROTOCOLS, AND ROBOTIC APPLICATIONS

EXPERIMENTAL WORKS ON POLYMERIC MEMBRANE USED IN PEM FUEL CELLS

Mehmet Fırat DOĞAN[1]

INTRODUCTION

Today's century is very close to being called the hydrogen age. It is obvious that the global warming and destruction on the world caused by the brutal use of fossil fuels after the industrial era, and the crises and wars that countries experience as a result of their dependence on foreign energy will end in the hydrogen age.

The main point that hydrogen should be addressed should be its role as a savior in eliminating the energy deficit rather than eliminating environmental problems. For countries that are dependent on foreign energy, such as Türkiye, the main issue in the use of hydrogen energy should not be zero emission targets, but energy security and reducing foreign dependency [1].

In recent years, the scientific world has started to work on new energy technologies for these reasons. Although there is a variety of fuel cells providing electricity generation from hydrogen, proton exchange membrane fuel cell are the types on which most research and studies are carries out because they are thought to be the most suitable for use in daily life [2].

PEM fuel cells are among the biggest candidates among the energy production devices of the future due to their high efficiency, usability in room conditions and the fact that their waste is not harmful to the environment. However, the biggest reason why PEM fuel cells have not yet been commercialized is because they do not provide the desired life. For this reason, studies on the development of the membrane, the most important element of fuel cells, and even the production of membraneless fuel cells are continuing rapidly.

In this study, considering the operating conditions of the polymeric membrane, which is the main element of PEM Fuel cells, the mass losses and changes in the material heated at 3 different temperature rates in 2 different environments were investigated and interpreted with DTA-TG experimental studies.

FACTORS EFFECTING FUEL CELL LIFE

Damage to fuel cells occurs even in the absence of externally applied mechanical loads. The mechanical response of a stressed membrane is swelling or contraction as a result of hydration and temperature change.

[1] Marmara University, Istanbul/Turkey, Orcid: https://orcid.org/0009-0007-4217-6695 , mfiratdogan@hotmail.com
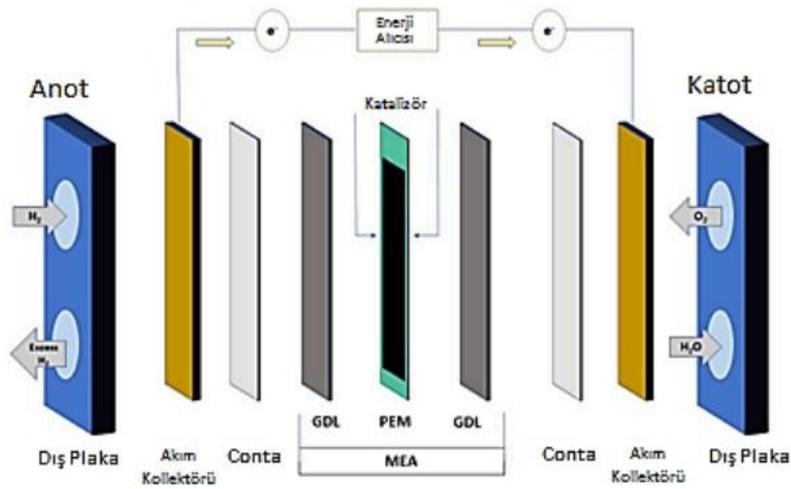
Figure 1: Detailed view of a fuel cell [3]

The main reason for the membrane not being commercially usable is generally due to cracking, tearing, and perforation that cause structural damage due to humidity, temperature, and mechanical effects under operating conditions.

During the operation of the fuel cell, changes in temperature and humidity cause stress and strain in the membrane [4].

It is known that the material undergoes more deformation than other conditions, especially in air-conditioning studies carried out in dry environments without humidity [5].

EXPERIMENTAL WORKS

It is known that Nafion membrane cannot withstand high temperatures and that weight losses occur in the material. The purpose of the DTA-TG experiment we conducted is to see the material losses by heating the Nafion material under PEM fuel cell operating conditions. At the same time, the material was heated in different environments and at different heating rates, and the effect of the heating rate on the material and the mass losses in the material in different operating environments were examined. At the same time, it is to determine the temperatures that the membrane can withstand.

Nafion material was divided into pieces weighing 2.5 mg and a total of 6 experiments were conducted in the DTA-TG device in 2 different environments as Nitrogen and Oxygen, 5, 10 and 15°C/min, 3 different heating rates. To see the behavior of the samples at high temperatures, they were heated up to 400C and the losses of the material according to time, temperature and environment were examined. It was also observed at what temperature the materials changed phase.

Figures 1-6 show the results of DTA-TG experiments of Nafion in two different operating environments, Nitrogen and Oxygen, at heating rates of 5C/min, 10C/min and 15C/min.

The TG curve shows the mass losses in the material due to increasing temperature, and the DTA curve shows the derivative of the TG curve, that is, its change depending on time, and the reaction zones.

As a result of the studies, much more weight loss occurred in heating in an oxygen environment than in heating in a nitrogen environment due to the burning properties of oxygen. In experiments conducted in a nitrogen environment, there is no structural change in endothermic reactions, the structure is not damaged.

From the curves obtained as a result of the experiments, it was seen that the material lost weight very slowly up to around 250C, and much faster after this temperature. If the operating conditions of the PEM fuel cell are taken into account and regular cooling is applied, it was seen that the mass loss can be kept at the desired level in operations below this temperature.

In experiments conducted in oxygen, it was observed that excessive losses occurred due to the burning properties of oxygen. In the experiments, the first part disintegrated up to 250C, and the second part showed a rapid weight loss after 250C. However, in experiments conducted in nitrogen, disintegration of the materials was observed rather than weight loss.

The results of DTA-TG experiments are given in the table below.

Table 1: Experimental Results

	NITROGEN ENVIRONMENT			OXIGEN ENVIRONMENT		
Heating Velocity	5C / min	10C / min	15C / min	5C / min	10C / min	15C / min
Energy required to go from 30C to 50C - Delta H	345,0518 J/g	934,7472 J/g	638,5563 J/g	215,6971 J/g	1151,4036 J/g	960,3826 J/g
Energy required to go from 50C to 70C - Delta H	1275,9387 J/g	3570,7513 J/g	631,8048 J/g	2105,0109 J/g	3260,5708 J/g	3160,4786 J/g
Mass loss in material up to 70C	4,471%	3,956%	4,785%	3,811%	3,201%	3,437%
Mass loss in material up to 270C	5,550%	5,789%	6,830%	5,820%	4,957%	4,998%
Total Mass Loss	24,075%	22,518%	15,406%	57,442%	30,179%	22,233%

Another result obtained from the examination of the curves is that the crystalline temperatures of the material, that is, the phase transition temperatures, are around 50C and 70C.

In addition, the curve evaluations showed that as the heating rate increased, the losses in the material decreased, from which it can be understood that adjusting the initial temperature of the PEM fuel cell will provide an additional contribution to the material life.
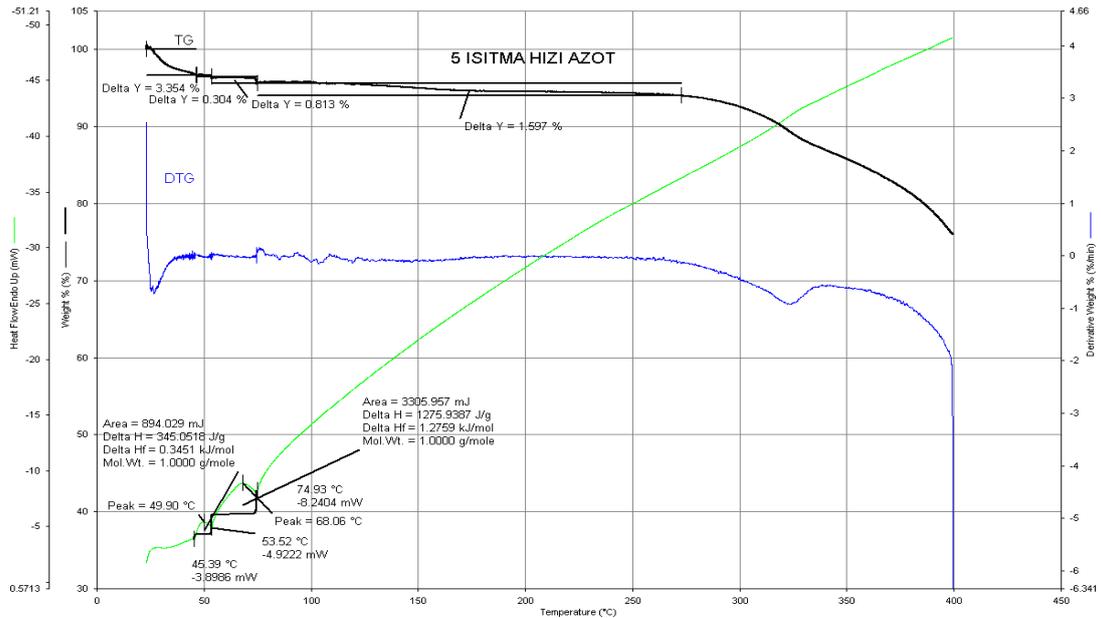
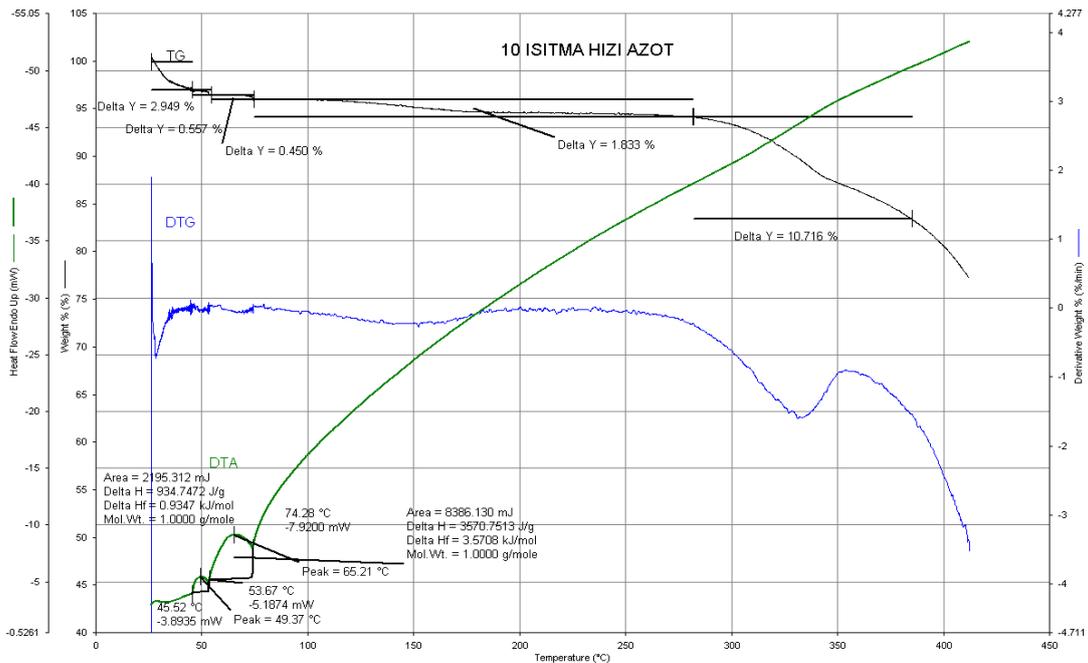Figure 2: DTA-TG Experiment Results of Nafion Material in Nitrogen Environment at 5C/min Heating Rate

Figure 3: DTA-TG Experiment Results of Nafion Material in Nitrogen Environment at 10C/min Heating Rate
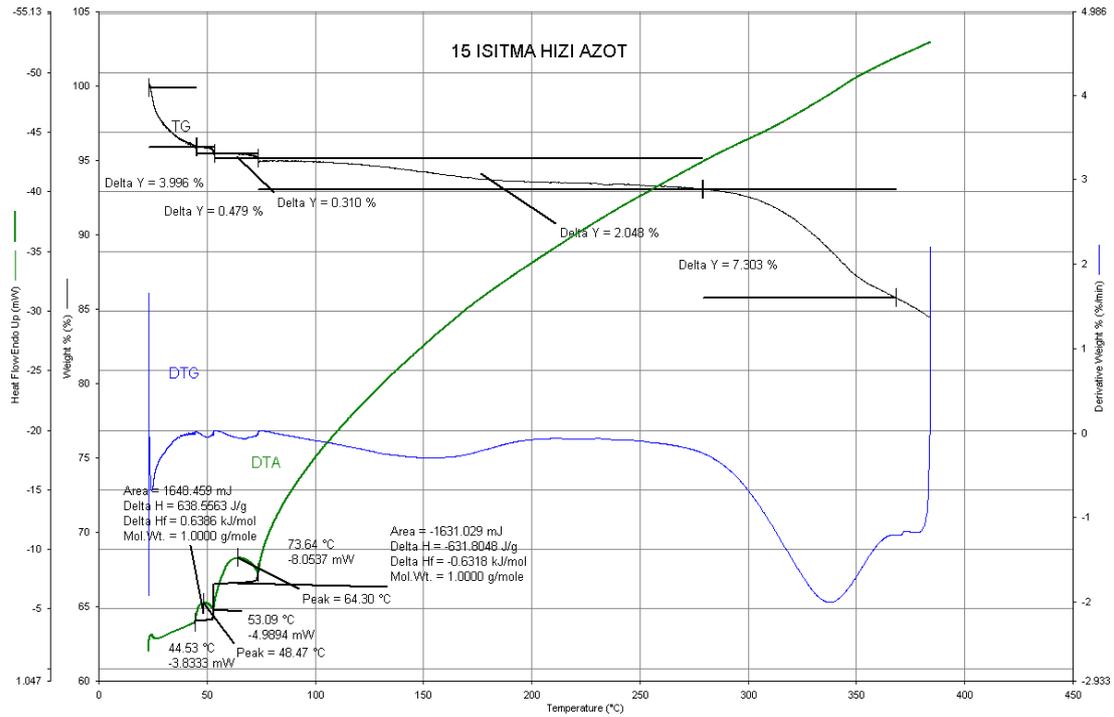
Figure 4: DTA-TG Experiment Results of Nafion Material in Nitrogen Environment at 15C/min Heating Rate

Figure 5: DTA-TG Experiment Results of Nafion Material at 5C/min Heating Rate in Oxygen Environment

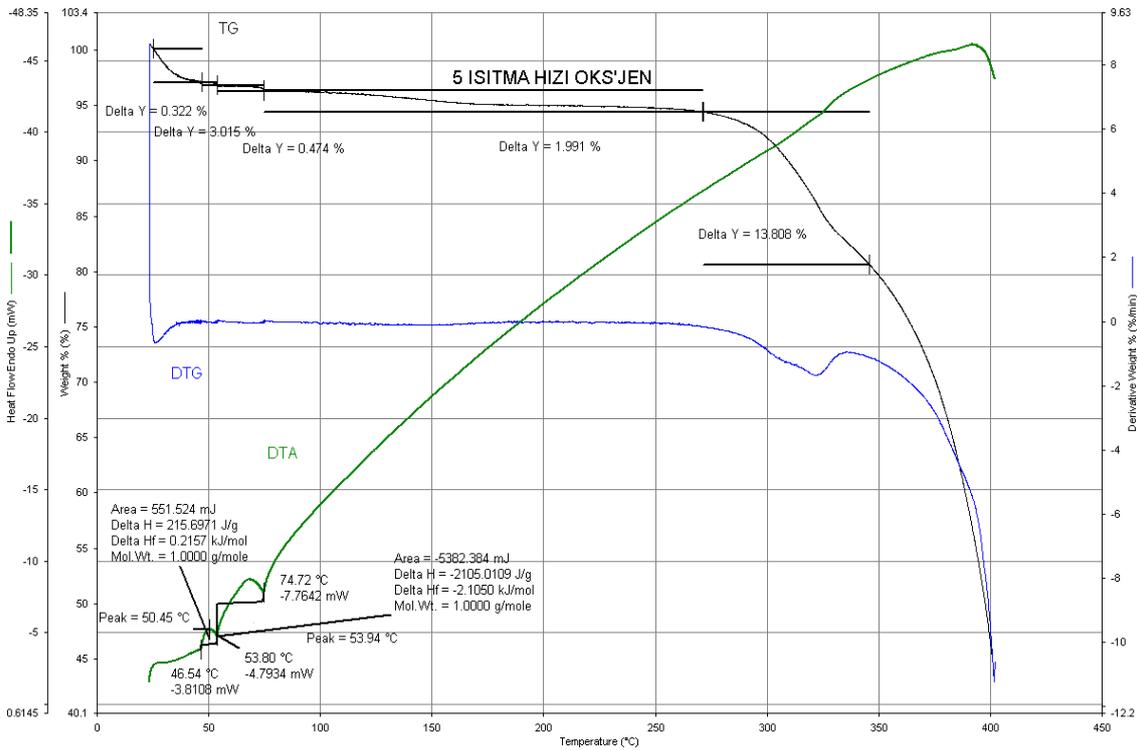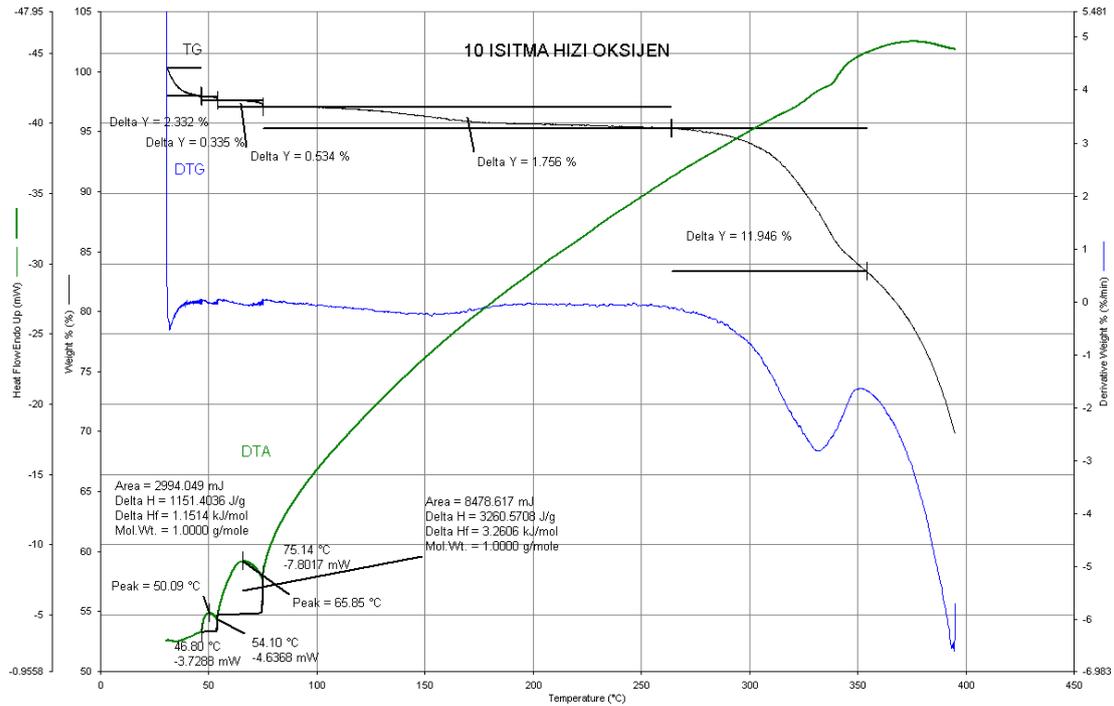Figure 6: DTA-TG Experiment Results of Nafion Material at 10C/min Heating Rate in Oxygen Environment
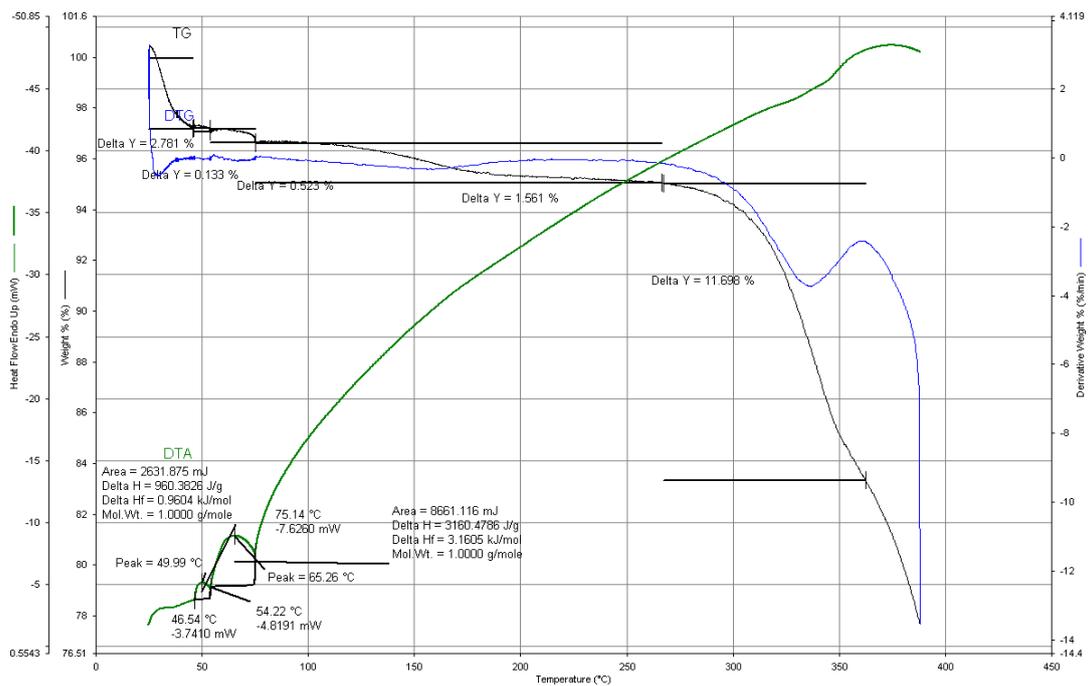
Figure 7: DTA-TG Experiment Results of Nafion Material at 15C/min Heating Rate in Oxygen Environment

RESULT

Each of the equal weight Nafion materials were heated in Oxygen and Nitrogen environments and at 3 different heating rates of 5-10-15C/min and the mass losses in the samples were examined with the DTA-TG device.

In the results, more mass loss was observed in the materials due to the burning feature of oxygen. It was also observed that fragmentation occurred in the materials in the nitrogen environment.

The materials showed normal weight losses up to 250C, and an excessive level of material loss was observed after 250C.

From the obtained data, it was seen that the mass loss decreased with the increase in the heating rate and it was evaluated that the material life could be extended by adjusting the initial operating temperature.

REFERENCES

[1] Mehmet Fırat Doğan, Strategic Strategic Importance of Hydrogen within the Scope of Turkey's New Century Vision, 8th International Hydrogen Congress, Diyarbakır, 2024

[2] Mehmet F. Doğan, vd. Proton Değişim Membran Yakıt Hücrelerinde Kullanılan Polimerik Membranın Modellenmesi, XVI. Ulusal Mekanik Kongresi, Kayseri, 2009

[3] Parekh, Abhi. "Recent developments of proton exchange membranes for PEMFC: A review." Frontiers in Energy Research vol. 10 p.1-13, 2022.

[4] W. Liu , K.Ruth, G.Rusch, J.New Mater, Eletrochem. Syst 4, 227-232, 2001.

[5] M. Firat Doğan, Özgen Çolak, Proton Değişim Membran Yakit Hücrelerinde Kullanilan Polimerik Membranin Deneysel Olarak Incelenmesi 5. Uluslararası İleri Teknolojiler Sempozyumu (İATS'09), 13-15 Mayıs 2009, Karabük, Türkiye

RESUME

Mehmet Fırat DOĞAN

Mehmet Fırat DOĞAN was born in Büyükada, Istanbul, and is the son of an artist mother and a professor father. He has completed his Bachelor, Master's degree and Phd degree in Mechanical Engineering, plus he has MSc degree in Political Science. He has been working on Hydrogen Technology since 2006 as a researcher. While he was studying his Bachelor, Master and Phd degrees, he several times went to USA for his researches. Moreover he has been working as chairman in his group of companies. In Energy, Engineering, Construction, Health, Agriculture, Manufacturing, and Tourism total 7 different sectors, he is managing his own group of companies in more than 30 countries. Plus He is the last descendant of the lineage of Xalti Mir Royal Family, the ancestor and leader of the Yazidis.. He is the selected leader of the Xalti Ezidi People in the World since 2019. After continuing his academic and business life abroad for many years, he returned to Turkey after his father's death to find solutions to the problems of the Yazidi and Haltanli people, whom he led, who were scattered all over the world, to re-establish Yazidi unity in the world, to establish Yazidi representation among world societies, and to benefit the economic, cultural and educational development of each individual of the Haltanli and Yazidi people. Mehmet Fırat Doğan, in addition to being a doctoral scientist, is one of the important people of business life with the group companies he operates in Turkey and abroad, the owner and founder of which he is the founder, and continues to carry out large-scale important projects in various countries of the world and to provide employment. He has made serious contributions to the scientific community with the scientific studies he has carried out in important universities of the world, the international scientific articles and projects he has written. Mehmet Fırat DOĞAN, who is regularly invited as a speaker in international forums and congresses, addresses a wide audience in the world with his business life and academic life, in addition to his social leadership. Doğan, who has been also conducting research and studies on social development, religious history, politics and psychology for over 10 years, continues his business, academic and political life actively. He continues the construction of Xalta Antique City in order to tell the contributions of the Xalti Family, one of the oldest living families in the world, which has an important place in Mesopotamian history, to Mesopotamian history and their 2700-year-old life adventure.

- 241 -

A PRACTICAL EVALUATION OF VANET ROUTING PROTOCOLS IN MODERN TRANSPORTATION SYSTEMS

Joseph Mosengo MOSENGO[1], Mostafa Ayoubi MOBARHAN[2], Atefeh Ahmadniai KHAJEKINI[3], Benie PONTE[4], Kerene TUZOLANA[5]

INTRODUCTION

In the ever-evolving landscape of modern transportation, Vehicular Ad-hoc Networks (VANETs) stand as a transformative force, promising advancements in road safety, traffic efficiency, and overall vehicular communication. The integration of VANETs into this dynamic environment is particularly significant in the context of the Internet of Things (IoT) and Machine-to-Machine (M2M) systems, ushering in a new era of interconnected technologies that are transforming various industries. Driven by seamless communication and data exchange between devices, intelligent and responsive environments are built on the foundational frameworks of IoT and M2M systems. VANETs represent a pivotal component of this interconnected paradigm, focusing on the dynamic and ad-hoc communication among vehicles. VANETs play a crucial role in the evolution of Intelligent Transportation Systems (ITS), contributing to enhanced road safety, traffic efficiency, and an improved overall passenger experience. This research aims to comprehensively understand the fundamentals, challenges, and opportunities within the VANET landscape, particularly focusing on their integration into smart parking systems. As we delve into the multifaceted realm of VANETs, our objective is to shed light on optimal strategies for reliable VANET functioning in diverse traffic scenarios. ITS refer to advanced technologies and applications designed to enhance the efficiency, safety, and sustainability of transportation networks. These systems integrate information and communication technologies to manage and optimize various aspects of transportation, including road traffic, public transit, and infrastructure. VANETs emerge as cornerstones for fostering connectivity and communication among vehicles on the road, providing real-time information, optimizing traffic flow, and mitigating the challenges associated with modern urban mobility.

Expanding the horizon of VANET applications, this research places a specific focus on the integration of VANETs in smart parking systems. The utilization of VANETs in smart parking is envisaged as a solution to the burgeoning problem of limited parking spaces, especially in

[1] Final International University, North Cyprus, via Mersin 10, Turkey, Orcid: https://orcid.org/0009-0003-2038-8313, joseph.mosengo@final.edu.tr
[2] Final International University, North Cyprus, via Mersin 10, Turkey, Orcid: https://orcid.org/0000-0002-0298-3384, mostafa.mobarhan@final.edu.tr
[3] Final International University, North Cyprus, via Mersin 10, Turkey, Orcid: https://orcid.org/0009-0001-5002-1999, atefeh.khajekini@final.edu.tr
[4] Final International University, North Cyprus, via Mersin 10, Turkey, Orcid: https://orcid.org/0009-0003-8785-901X, benie.ponte@final.edu.tr
[5] Final International University, North Cyprus, via Mersin 10, Turkey, Orcid: https://orcid.org/0009-0006-9697-0053, kerene.tuzolana@final.edu.tr

densely populated urban areas. Through the deployment of wireless sensor networks and communication technologies, VANETs contribute to efficient parking space utilization, reducing economic costs, and enhancing the overall urban mobility experience.

An integral aspect of VANETs lies in their routing protocols, which dictate the efficient transmission of data within the dynamic vehicular environment. The Ad Hoc On-Demand Distance Vector (AODV) and Optimized Link State Routing (OLSR) protocols take center stage in this study, as we aim to conduct a comparative analysis based on key performance metrics. The insights derived from this analysis will illuminate optimal routing strategies, essential for the reliable and robust functioning of VANETs in diverse traffic scenarios.

VANET operates as an ad-hoc network, where vehicles serve as nodes engaging in wireless communication. V2V (Vehicle to Vehicle), V2I (Vehicle to Infrastructure), and V2X (Vehicle to Everything) represent distinct variations within the VANET framework.

In the context of V2V communication, vehicles share information encompassing speed, location, stability, braking, and heading. The technology facilitating V2V communication relies on Dedicated Short Range Communication (DSRC) [3], akin to Wi-Fi.

Expanding the communication scope, V2I enables vehicles to interact not only with each other but also with traffic poles and pedestrians, bolstering road safety.

This system establishes a wireless network link allowing vehicles to communicate with the surrounding road infrastructure. The V2I system necessitates essential components, including Vehicle On-Board Units or Equipment (OBU or OBE), Roadside Units or Equipment (RSU or RSE), and a secure communication channel. The OBU, installed within the vehicle, functions as the primary unit for connecting with infrastructure elements.

V2X is an advanced technology that empowers vehicles to communicate comprehensively with traffic and elements in their vicinity through short-range wireless signals. V2V and V2I form the foundational components of the V2X system. The overarching goal of V2X is to heighten safety measures and avert potential collisions. Beyond this, the system provides supplementary information such as real-time weather updates, road conditions, and details on nearby accidents. Notably, these communications play a crucial role even in parking scenarios. Facilitating these operations involves the rapid exchange of substantial data within minimal time frames. The essential Quality of Service (QoS) requirements for effective V2V communication encompass low latency and high bandwidth.
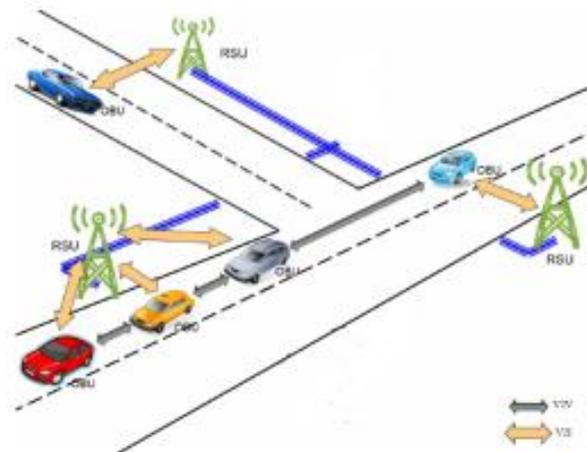
Fig. 1. V2V and V2I communication

RELATED WORKS

In this segment, we examine recent literature concerning smart parking systems employing diverse methodologies and strategies. These systems aim to simplify the vehicle parking experience for drivers. The feasibility of this improvement stems from the extensive research conducted by scholars, focusing on integrating conventional parking systems with innovative technologies and methodologies. This synthesis represents a culmination of sustained efforts to enhance parking solutions and align them with contemporary advancements. The Smart Parking System's reliance on NB-IoT technology ensures energy efficiency and long-range communication capabilities, making it suitable for large-scale urban deployments. The system's scalability allows it to adapt to diverse urban landscapes, from densely populated city centers to suburban areas.

In [5], the researchers presented a more futuristic approach. They explored an automated parking lot management system employing robot-type smart cars with embedded technologies. Utilizing wireless sensor networks, the system streamlines parking processes, contributing to improved efficiency and addressing challenges associated with urban parking congestion. The integration of robotic elements marks a significant advancement, potentially reshaping traditional parking paradigms. Furthermore, the study highlights the potential impact on reducing traffic congestion and improving the overall flow within parking facilities. The autonomous nature of the smart car, coupled with wireless connectivity, contributes to a more dynamic and responsive parking environment.

In [6], this work proposes a smart parking navigation system that prioritizes the protection of user privacy. By providing efficient driving guidance retrieval while implementing privacy-preserving measures, the system strikes a balance between convenience and data security. This approach acknowledges the importance of privacy in smart parking systems, ensuring user trust and compliance with privacy regulations. Furthermore, the research explores the potential impact of the proposed system on traffic optimization, reduced parking search times, and overall improvement in the driving experience. The privacy-preserving approach aligns with

contemporary standards for responsible and ethical deployment of location-based services, particularly in the context of smart cities.

In [10], the researchers focused on developing an intelligent parking system that leverages advanced automation and computing technologies. The system's primary objectives are to reduce the time spent by drivers searching for parking spaces, minimize fuel consumption, and consequently lower CO_2 emissions associated with unnecessary driving in search of parking spots.

Key features of the Smart Parking System include real-time parking space monitoring, automated guidance to available parking spots, and integration with environmental sensors to assess and mitigate the system's ecological footprint. The study explores how the deployment of such a system can lead to more efficient use of urban infrastructure, contributing to a reduction in traffic congestion and the associated environmental impact.

Furthermore, the research investigates the economic and environmental benefits of the proposed smart parking solution. By minimizing the time spent searching for parking, drivers experience fuel savings and reduced carbon emissions. The system's potential positive impact on urban air quality and overall environmental sustainability is a focal point of the study.

VANET ARCHITECTURE

The architecture of VANETs is designed to facilitate the seamless exchange of information among vehicles and infrastructure elements. In what follows, we thoroughly describe the main VANET components and the interactions between them.

Main Components

The European VANET architecture standard diverges slightly from other models, particularly emphasizing the V2X communication system championed by the V2V Communication Consortium. The comprehensive reference architecture of the V2V Communication System is delineated across distinct domains:

In-Vehicle Domain: The in-vehicle domain, a central component of the VANET architecture, is intricately designed with one or multiple Application Units (AUs) and an On-Board Unit (OBU). AUs, whether integrated into the vehicle or functioning as standalone devices like smartphones and laptops, perform diverse functions. Each AU operates applications strategically crafted to utilize the robust communication capabilities of the OBU, encompassing tasks ranging from real-time navigation to entertainment and safety features.

At the core of the In-Vehicle Domain is the OBU, acting as the central hub that coordinates communication and connectivity within the vehicle. The OBU serves as the focal point for all AUs, establishing and maintaining an ongoing connection through either wired or wireless interfaces. It acts as a communication hub, enabling seamless data exchange between AUs and integrating various in-vehicle systems. Beyond facilitating communication within the vehicle, the OBU assumes a pivotal role in linking the vehicle to the broader VANET ecosystem.

The collaboration between AUs and the OBU ensures a comprehensive and integrated approach to in-vehicle communication, effectively transforming the vehicle into a smart and connected entity. From intelligent navigation and adaptive cruise control to infotainment and diagnostics,

the In-Vehicle Domain sets the stage for a new era of interconnected and intelligent vehicular systems. Each AU, whether an integral part of the vehicle or an independent portable device, operates applications that capitalize on the OBU's communication capabilities, fostering seamless connectivity through wired or wireless links.

Ad-hoc Domain: The ad-hoc domain constitutes a dynamic network of vehicles equipped with On-Board Units (OBUs) and strategically positioned stationary Road-Side Units (RSUs) along roadways. Within this dynamic domain, OBUs serve as key facilitators, enabling both direct and multi-hop wireless communication among vehicles, fostering a framework that promotes spontaneous and ad-hoc interactions.

The OBUs play a vital role in establishing and maintaining communication links, allowing vehicles to engage in direct or multi-hop wireless exchanges. This capability is crucial for promoting ad-hoc interactions, where vehicles can communicate with each other flexibly based on the immediate needs of the vehicular environment. The versatility of OBUs in supporting various communication scenarios contributes to the dynamic nature of the ad-hoc domain within VANETs.

Complementing the role of OBUs, stationary Road-Side Units (RSUs) strategically positioned along roadways significantly enhance the ad-hoc network's coverage. These RSUs, stationary devices tethered to both the physical infrastructure and the Internet, play a pivotal role in extending the reach of the ad-hoc domain. Their functions encompass the transmission, reception, and forwarding of data within the ad-hoc network, contributing to an expanded and more robust communication framework.

Moreover, RSUs act as intermediaries within the ad-hoc domain, facilitating seamless communication and data exchange between vehicles. By serving as relay points, RSUs enhance the efficiency and reliability of the ad-hoc network, enabling vehicles to overcome potential communication challenges in various scenarios.

The connectivity provided by RSUs goes beyond the immediate vehicular environment. OBUs, connected to RSUs, gain access to the Internet, allowing for a broader spectrum of communication possibilities. This connectivity may occur through RSUs linked to the infrastructure or by utilizing public/commercial and private wireless Hot Spots (HSs). As a result, the ad-hoc domain becomes intricately woven into the larger network infrastructure, connecting vehicles to the Internet and enabling them to access a wide array of online services and information.

In summary, the ad-hoc domain, characterized by the collaboration between OBUs and RSUs, forms a dynamic and versatile network within VANETs. This network not only facilitates direct and multi-hop wireless communication among vehicles but also extends its reach through strategically positioned RSUs, contributing to a robust and interconnected vehicular communication system.
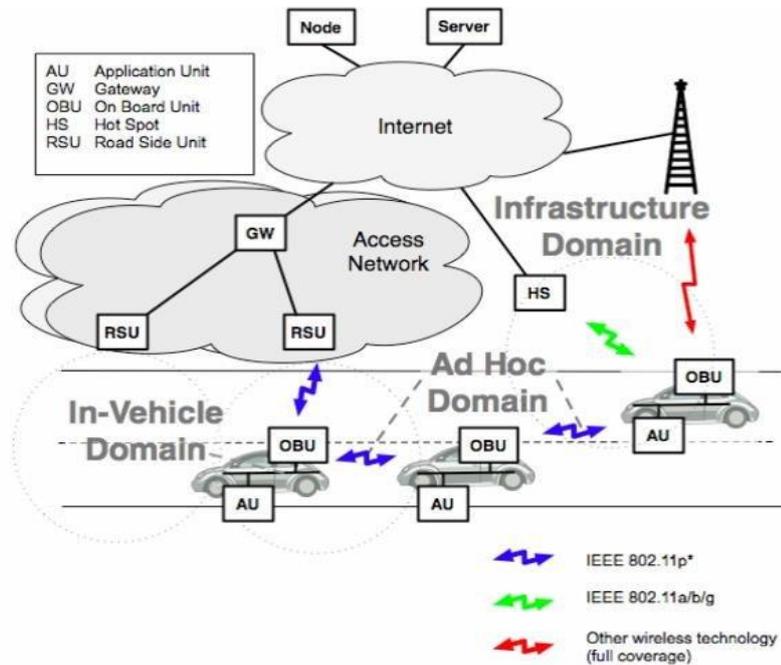
Figure 2. Architecture of VANET systems

Infrastructure Domain: Within the VANET architecture, Infrastructure Domain Access encompasses both Hot Spots (HSs) and Road-Side Units (RSUs), playing crucial roles in facilitating seamless communication and connectivity for vehicles. Hot Spots are specifically designated areas that provide wireless internet access, serving as essential nodes for On-Board Units (OBUs) to establish communication links within the broader VANET ecosystem.

Strategically deployed along roadsides, RSUs serve as pivotal intermediaries for data transmission, significantly enhancing the coverage and reliability of the ad-hoc network. These units are strategically positioned to optimize communication between vehicles and to extend the reach of the VANET infrastructure. The connectivity options for RSUs include integration with both infrastructure networks and the broader internet, contributing to the robustness of the VANET system.

In scenarios where neither RSUs nor HSs offer direct internet access, vehicles equipped with OBUs have the flexibility to utilize cellular radio networks such as HSDPA, WiMax, and 4G. This multi-faceted Infrastructure Domain Access ensures versatile connectivity options for vehicles operating within the VANET framework. The ability to seamlessly switch between different access points enhances overall efficiency in vehicular communication systems, providing vehicles with reliable connectivity regardless of their location.

The integration of HSs and RSUs as part of the Infrastructure Domain Access underscores the adaptability of the VANET architecture to diverse communication scenarios. Whether relying on designated Hot Spots, utilizing strategically positioned RSUs, or seamlessly transitioning to cellular radio networks, vehicles within the VANET framework benefit from a comprehensive and versatile infrastructure. This approach not only ensures effective communication between OBUs but also contributes to the overall enhancement of the VANET system's efficiency and reliability, catering to the dynamic nature of vehicular environments.

Communication Architecture

VANET communications exhibit a nuanced categorization:

• **In-Vehicle Communication:** In this aspect, communication occurs between the On-Board Unit (OBU) of a vehicle and its Application Units (AUs). The OBU serves as a central hub for various applications, fostering seamless communication within the vehicle itself.

• **Vehicle-to-Vehicle (V2V) Wireless Communications:** Vehicles engage in wireless communication with each other through their respective OBUs. This enables direct interaction between vehicles on the road, facilitating real-time data exchange related to speed, location, and other relevant information.

• **Vehicle-to-Infrastructure (V2I):** This involves bidirectional wireless communication unfolding between vehicles and Road-Side Units (RSUs) connected to the infrastructure. RSUs play a crucial role as intermediaries, enabling vehicles to communicate not only with each other but also with the broader road infrastructure.

• **Infrastructure-to-Infrastructure (I2I) Communications:** RSUs participate in wireless communications to extend the network's coverage. This communication mode involves RSUs interacting with each other, contributing to a more expansive and interconnected infrastructure to enhance the overall efficiency of vehicular communication.

• **Vehicle-to-Broadband Cloud (V2B) Communications:** Vehicles establish communication with the broadband cloud using wireless broadband technologies such as 3G/4G. This connection forms a vital link to cloud-based resources and services, enabling vehicles to access a wide array of information and functionalities beyond the immediate vehicular environment.

MAIN ISSUES OF VANET SYSTEMS

Vehicular Ad Hoc Networks (VANETs) confront distinctive challenges that set them apart from conventional wireless networks. In this section, we will talk about the main issues of VANET systems. Since this research focuses specifically on the integration of VANETs into smart parking systems, we will discuss the challenges of VANET systems concerning smart parking.

One challenge lies in developing accurate and real-time parking space availability information. Ensuring that the data provided to drivers is precise requires sophisticated sensor technologies and robust communication protocols. Research focuses on enhancing the reliability and accuracy of such information to optimize parking searches, reduce congestion, and minimize environmental impact.

Smart parking systems also face challenges in achieving efficient navigation.. Integrating VANETs with navigation algorithms to guide drivers to available parking spaces requires addressing issues such as dynamic traffic conditions, real-time updates, and the seamless transition from vehicular networks to localized parking environments. This necessitates the development of adaptive and context-aware navigation strategies.

Interoperability and standardization of communication protocols between vehicles, infrastructure, and smart parking systems present additional challenges. Creating a cohesive

and standardized communication framework ensures seamless integration, promoting efficient data exchange for smart parking applications.

Privacy concerns emerge regarding the collection and dissemination of parking-related data. Balancing the need for accurate information with individual privacy protection requires innovative approaches to data anonymization, encryption, and permission-based access control.

The intricacies arise from the predefined traffic flow, high mobility rates of contemporary transport systems, and the need for large-scale networks to cover extensive geographical areas. VANET applications not only require robust Quality of Service (QoS) but also grapple with the intricacies of secure and reliable message broadcasting.

Navigating the cooperative landscape among neighboring vehicles and sensor networks becomes progressively intricate as the vehicle count rises, posing socio-economic challenges. The dynamic topology of VANETs, influenced by factors like vehicle speed, introduces communication hurdles, particularly in sparsely connected networks. Mobility modeling, incorporating variables such as traffic dynamics and road structure, adds further layers of complexity. On a positive note, the widespread presence of storage and unlimited battery power in modern vehicles enhances the reliability of communication. Additionally, the integration of GPS devices as onboard sensors significantly augments communication effectiveness within VANETs.

VANET ROUTING METHODS

In this section, we will be analyzing the roles of Ad-Hoc On-Demand Distance Vector (AODV) and Optimized Link State Routing (OLSR) in the context of VANETs.

These protocols are instrumental in establishing and maintaining efficient communication among vehicles in dynamically changing vehicular environments. Understanding their features and functionalities provides valuable insights into their contributions to VANET integration.

AODV Protocol: Ad-hoc On-Demand Distance Vector (AODV) is a reactive routing protocol designed to manage the dynamic nature of VANETs. In AODV, routes between source and destination nodes are established only when needed, optimizing network resources. When a vehicle requires a route to communicate with another, AODV initiates a route discovery process, and the route is established on-demand. This approach minimizes the overhead associated with route maintenance in scenarios where constant communication is unnecessary.

AODV relies on distance vector routing principles, where each vehicle maintains a routing table containing information about neighboring vehicles and their distances. This information is updated as the network topology changes. AODV is well-suited for VANETs, as it efficiently adapts to the highly dynamic and unpredictable nature of vehicular environments.

OLSR Protocol: Optimized Link State Routing (OLSR) is a proactive routing protocol that maintains up-to-date routing information for all nodes within the network. Unlike AODV, OLSR continuously updates and refreshes routing tables, ensuring that routes are readily available when needed. OLSR employs a proactive approach by periodically exchanging topology information among vehicles, enabling quick route calculations when data transmission is required.

In OLSR, each vehicle maintains a detailed knowledge of the network's topology, allowing for rapid route determination. This proactive nature of OLSR reduces the latency associated with route discovery, making it advantageous in scenarios where low-latency communication is essential.

The integration of AODV and OLSR protocols within VANETs is pivotal for ensuring reliable and efficient communication among vehicles. The choice between these protocols depends on the specific requirements of the VANET scenario.

AODV, with its reactive nature, is well-suited for VANETs where communication demands are sporadic and on-demand. It minimizes the routing overhead by establishing routes only when necessary, making it efficient in scenarios with varying communication patterns, such as intermittent data exchanges between vehicles.

On the other hand, OLSR's proactive approach is beneficial in VANETs with continuous and frequent communication needs. Its continuous updating of routing tables ensures that routes are readily available, reducing the latency associated with route discovery. This makes OLSR a suitable choice for applications in VANETs where real-time data exchange and low-latency communication are critical.

In summary, the integration of AODV and OLSR protocols in VANETs provides a flexible and adaptive routing framework. The choice between these protocols depends on the specific communication requirements and dynamics of the vehicular environment, ensuring that VANETs can effectively cater to a diverse range of scenarios and applications.

EVALUATION RESULTS

In the pursuit of assessing and comparing the performances of Ad-hoc On-Demand Distance Vector (AODV) and Optimized Link State Routing (OLSR) protocols within Vehicular Ad-Hoc Networks (VANETs), our study leveraged the power of NS3 simulation. This comprehensive evaluation aimed to shed light on the protocols' effectiveness under varying conditions, providing valuable insights into their respective capabilities.

Simulation Inputs: The simulation included various influencing factors that can significantly impact the performance of AODV and OLSR in a VANET environment. The inputs encompassed parameters such as WiFi mode, safety message size, application message size, loss model, transmit power, number of sinks, number of nodes, and speed. Each of these inputs played a crucial role in shaping the network dynamics and influencing the protocols' responses.

- **WiFi Mode:** The chosen WiFi mode set the foundation for the wireless communication environment within the simulation.
- **Safety Message Size:** The size of safety messages directly influences the payload and transmission characteristics crucial for real-time safety applications.
- **Application Message Size:** This parameter determines the size of non-safety-related application messages, contributing to the overall traffic load in the network.
- **Application data rate:** The rate at which application data is transmitted, influencing the overall network load.

- **Loss Model:** The loss model selection played a key role in emulating realistic communication scenarios, influencing packet loss patterns.
- **Transmit Power:** The transmit power setting influenced the range and reliability of communication links between vehicles.
- **Number of Sinks:** The presence of sinks in the network impacted data aggregation and routing strategies.
- **Number of Nodes:** The total number of nodes contributed to the network density, influencing congestion and communication patterns.
- **Node speed:** The speed of nodes simulated realistic vehicular mobility scenarios, affecting the handover and dynamic nature of the network.

Comparison Outputs: To gauge the effectiveness of AODV and OLSR under diverse conditions, we collected a set of key performance metrics for thorough comparison:

- **Average Goodput:** The average rate of successful data packet delivery, reflecting the protocols' efficiency in transmitting useful information.
- **MAC/PHY Overhead:** The overhead associated with Medium Access Control (MAC) and Physical (PHY) layers, indicating the additional resources consumed for communication.
- **Average Throughput:** The average rate of successful data transfer, considering both safety and application messages.
- **Lost Packets:** The number of packets that failed to reach their intended destination, providing insights into the reliability of the protocols.
- **Delay:** The time taken for a packet to travel from the source to the destination, influencing the responsiveness of the network.
- **Jitter:** The variation in delay, highlighting the stability and consistency of data packet delivery.

This comprehensive evaluation, conducted through NS3 simulation, serves as a robust foundation for understanding how AODV and OLSR perform in VANETs across a spectrum of realistic conditions. The gathered outputs facilitate an in-depth comparison, aiding researchers and practitioners in making informed decisions about protocol selection based on specific VANET deployment scenarios and requirements.
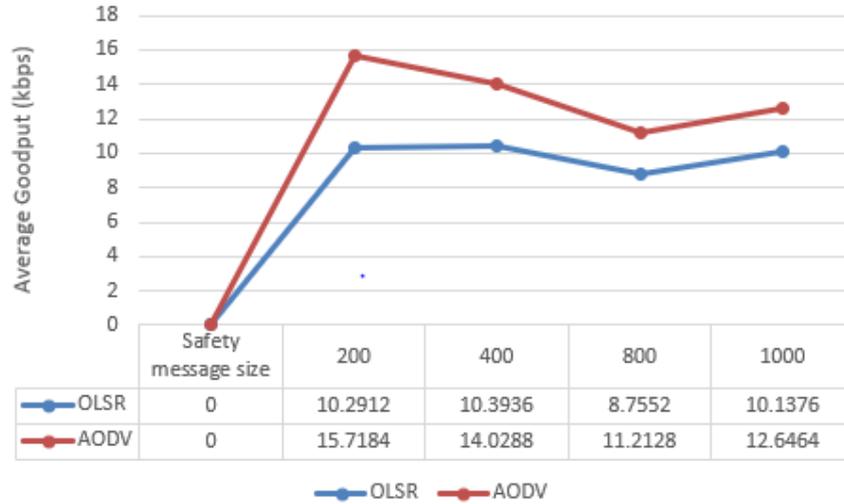
Safety message size	0	200	400	800	1000
OLSR	0	10.2912	10.3936	8.7552	10.1376
AODV	0	15.7184	14.0288	11.2128	12.6464

Figure 3. (a) Performance analysis of Average routing Goodput with varying safety message size

Safety message size	0	200	400	800	1000
OLSR	0	4816182948	2609143813	2508842851	5827011275
AODV	0	1.90435E+11	2.43876E+11	1.64504E+11	1.45454E+11
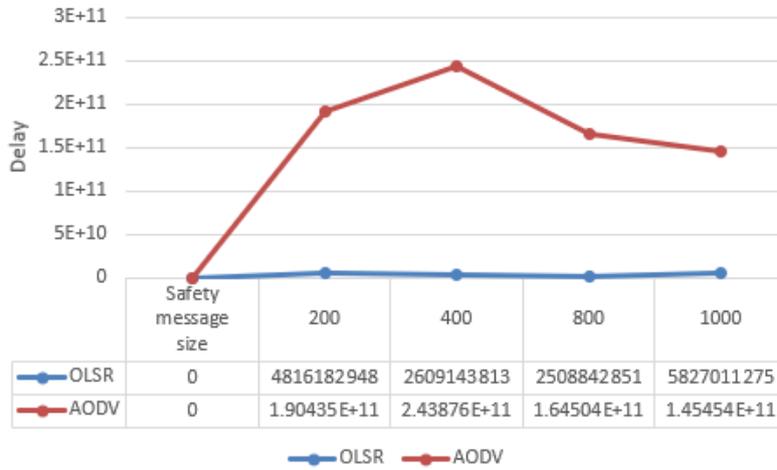
Figure 3. (b) Performance analysis of delay with varying safety message size

Safety message size	0	200	400	800	1000
OLSR	0	3258688102	2360915551	2182790137	2657847504
AODV	0	97529242705	1.04161E+11	73238722173	58898681994
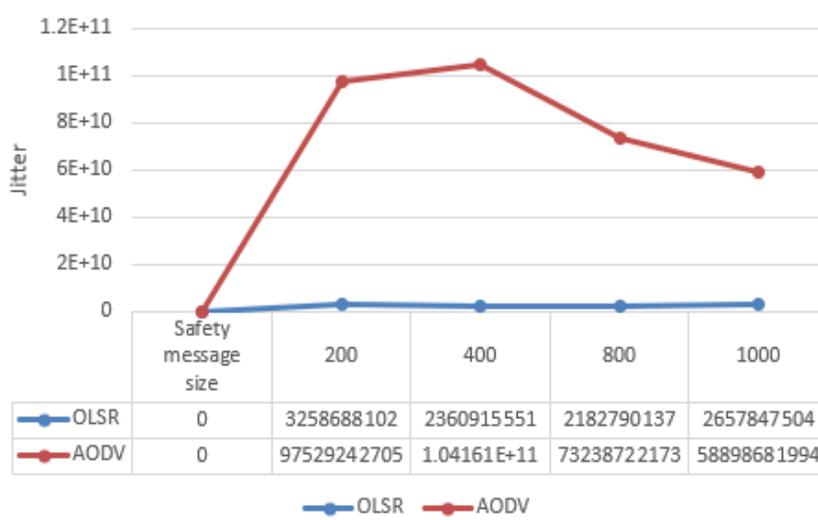
Figure 3. (c) Performance analysis of Jitter with varying safety message size

Safety message size	200	400	800	1000
OLSR	0	41	39	71	44
AODV	0	166	223	193	273
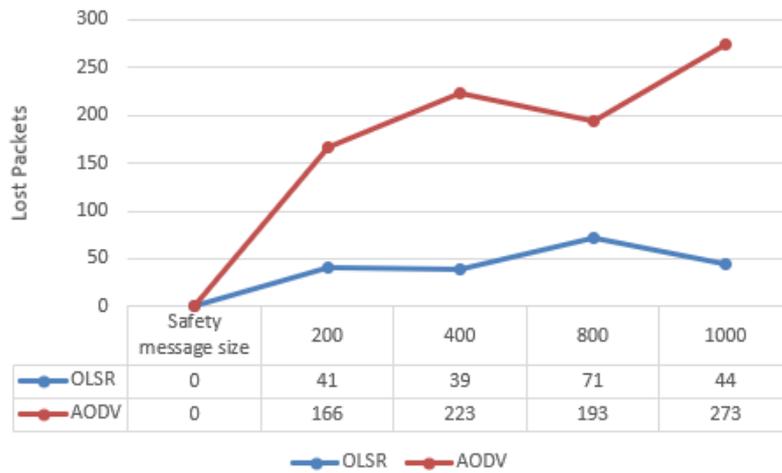
Figure 3. (d) Performance analysis of Lost packets with varying safety message size

Safety message size	200	400	800	1000
OLSR	0	0.304654	0.182495	0.0988055	0.0832917
AODV	0	0.471065	0.302168	0.159942	0.14168
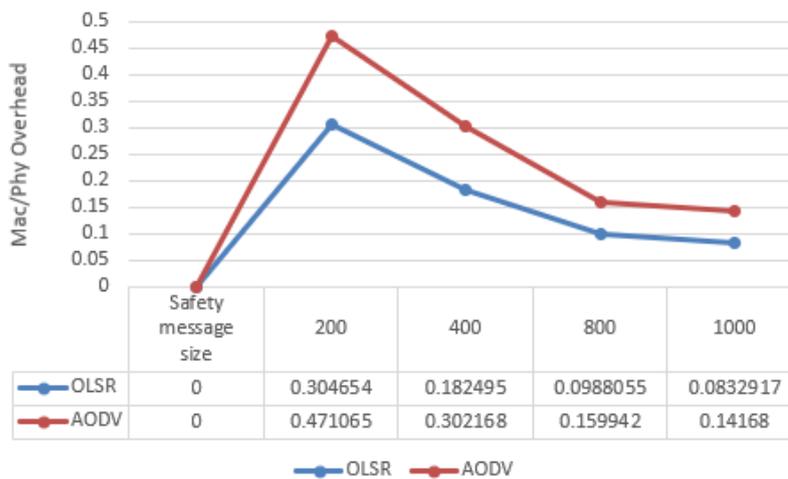
Figure 3. (e) Performance analysis of MAC/PHY Overhead with varying safety message size

Safety message size	200	400	800	1000
OLSR	0	2.45146	2.4721	2.14236	2.42413
AODV	0	15.073	17.22	15.7345	17.0987

Figure 3. (f) Performance analysis of Average Throughput with varying safety message size

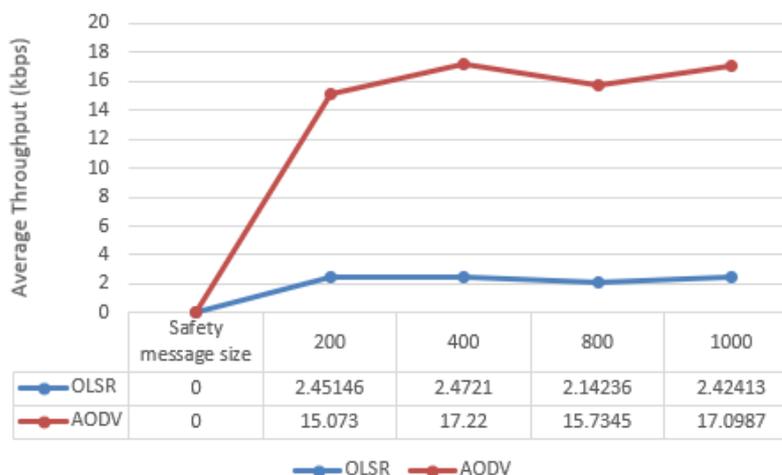- **Safety message size analysis:** In the analysis of safety message size, where the size varied from 200 and doubled up to 1000, distinct performance patterns emerged between the AODV and OLSR protocols. The OLSR protocol exhibited notable

advantages in key metrics, showcasing its efficiency in handling varying safety message sizes. Specifically, OLSR demonstrated significantly lower delay and jitter compared to AODV as shown in Fig 3 (b) and (c). This indicates that OLSR excels in minimizing the time taken for data packets to travel from source to destination and ensures a more stable and consistent delivery pattern.

Moreover, OLSR outperformed AODV in terms of average throughput and lost packets as shown in Fig 3 (f) and (d), further emphasizing its reliability in transmitting data under different safety message size scenarios. The lower average throughput for OLSR could be attributed to its proactive nature, ensuring continuous updates and route refreshments, which might introduce slightly higher control overhead.

In terms of MAC/PHY overhead, we can see in Fig 3 (e) both protocols exhibited an initial increase followed by a gradual decrease. However, throughout this fluctuation, OLSR maintained a lower MAC/PHY overhead value compared to AODV. This indicates that OLSR managed to achieve efficient medium access control and physical layer resource utilization, contributing to a more optimized communication framework. While both protocols experienced a gradual increase in average goodput, the AODV protocol demonstrated a higher average goodput compared to OLSR as shown in Fig 3 (a). Despite this, the lower delay, jitter, average throughput, and lost packets for OLSR suggest that, in the context of safety message size variations, OLSR provides a more stable and reliable performance, prioritizing key aspects such as reduced latency and improved packet delivery.

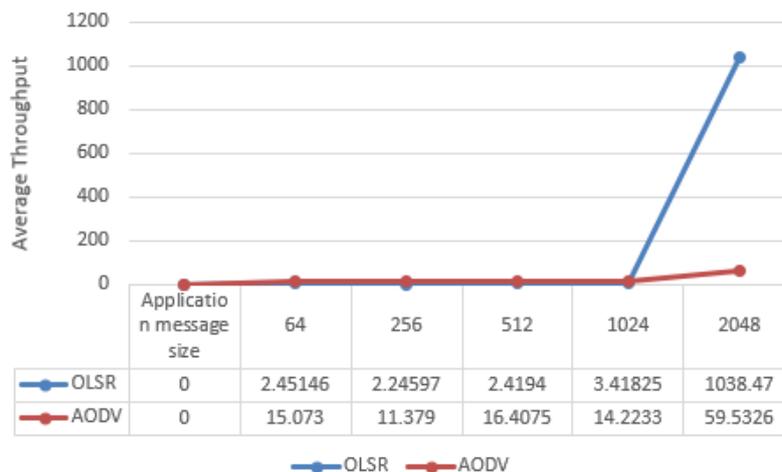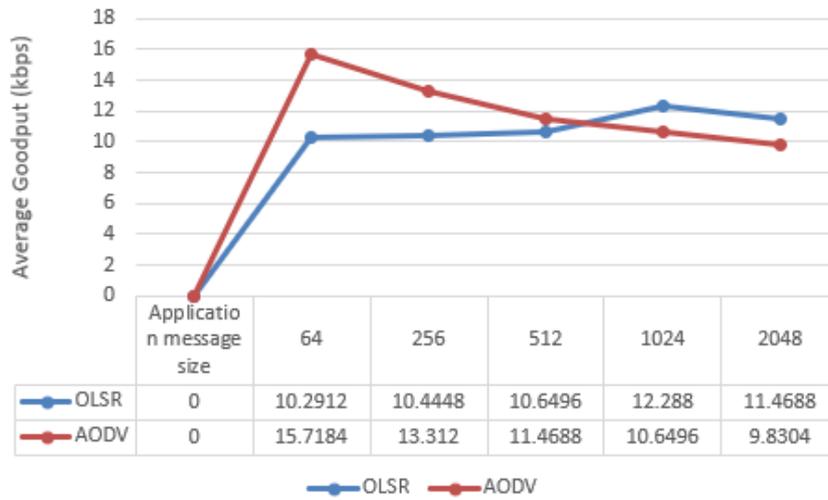| Application message size	64	256	512	1024	2048
OLSR	0	2.45146	2.24597	2.4194	3.41825	1038.47
AODV	0	15.073	11.379	16.4075	14.2233	59.5326

Figure 4. (a) Performance analysis of Average Throughput with varying application message size

Figure 4. (b) Performance analysis of Average goodput with varying application message size
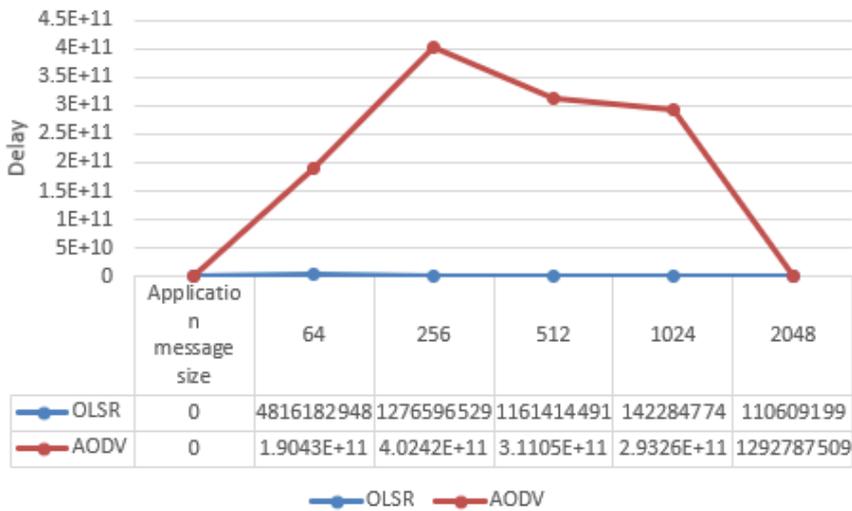
	Application message size	64	256	512	1024	2048
OLSR	0	10.2912	10.4448	10.6496	12.288	11.4688
AODV	0	15.7184	13.312	11.4688	10.6496	9.8304

Figure 4. (c) Performance analysis of delay with varying application message size

	Application message size	64	256	512	1024	2048
OLSR	0	4816182948	1276596529	1161414491	142284774	110609199
AODV	0	1.9043E+11	4.0242E+11	3.1105E+11	2.9326E+11	1292787509
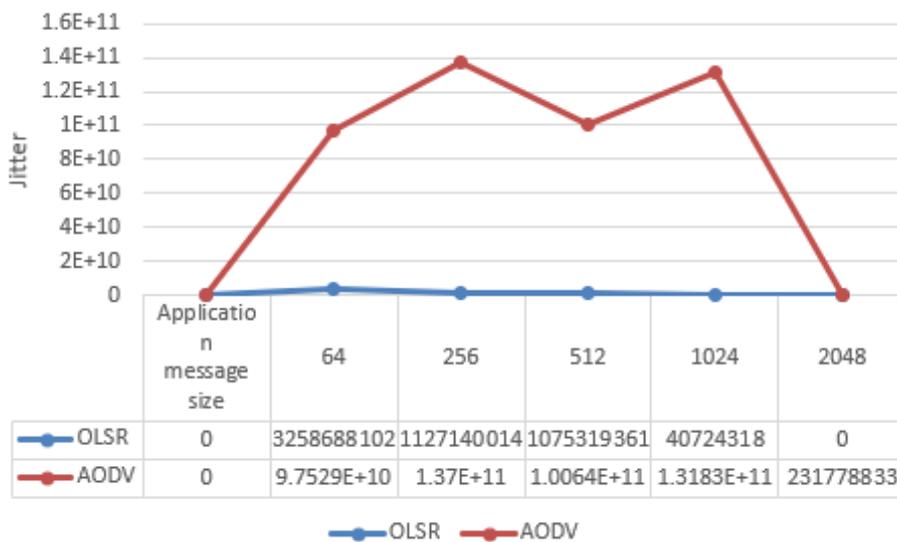
Figure 4. (d) Performance analysis of jitter with varying application message size

	Application message size	64	256	512	1024	2048
OLSR	0	3258688102	1127140014	1075319361	40724318	0
AODV	0	9.7529E+10	1.37E+11	1.0064E+11	1.3183E+11	231778833

Application message size	64	256	512	1024	2048
OLSR	0	41	9	5	3	2
AODV	0	166	338	292	452	95
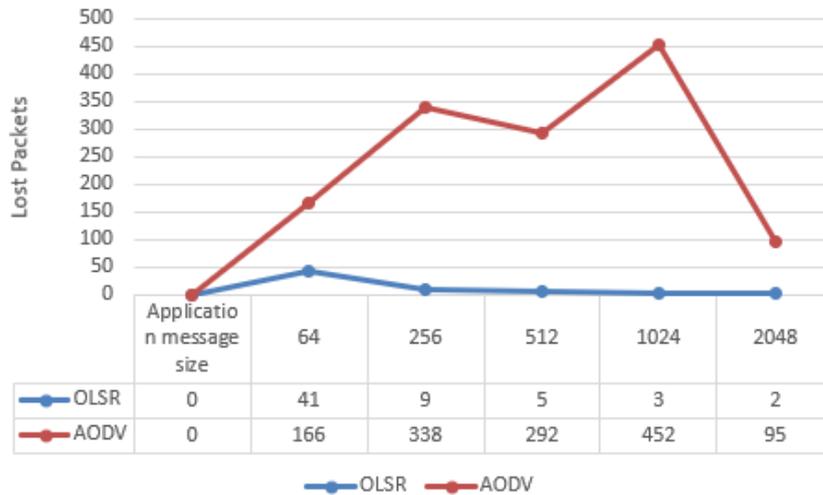
Figure 4. (e) Performance analysis of lost packets with varying application message size

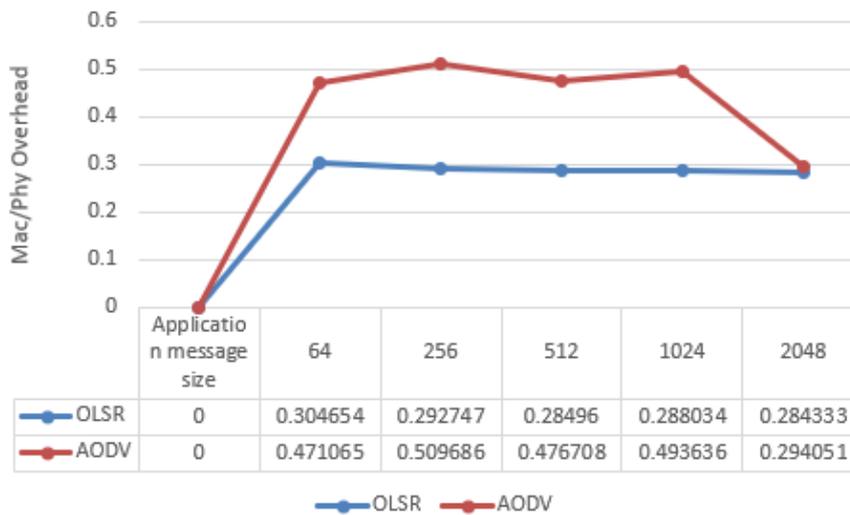| Application message size	64	256	512	1024	2048
OLSR	0	0.304654	0.292747	0.28496	0.288034	0.284333
AODV	0	0.471065	0.509686	0.476708	0.493636	0.294051

Figure 4. (f) Performance analysis of MAC/PHY overhead with varying application message size

- **Application message size analysis:** In the analysis of application message size, where the size ranged from 64 to 256 and then doubled up to 2048, distinct trends in performance metrics were observed for the AODV and OLSR protocols. Notably, OLSR demonstrated noteworthy advantages in several key aspects, highlighting its efficiency in adapting to varying application message sizes.

  In Fig 4 (a), the average throughput for both protocols initially appeared low, but a sudden increase was observed for the OLSR protocol. This indicates that OLSR adapted more effectively to larger application message sizes, showcasing its ability to handle increased data transfer rates efficiently.

  Fig 4 (b) displayed an increase in the average goodput of both protocols, followed by a gradual decrease, with AODV ending up lower than OLSR. This suggests that while both protocols experienced fluctuations in goodput, OLSR maintained a higher average goodput, emphasizing its effectiveness in delivering application-specific data under different size scenarios.

For delay, jitter, and lost packets (Fig 4 c, d, and e), OLSR consistently outperformed AODV, displaying significantly lower values. This indicates that OLSR maintained reduced latency, improved stability, and enhanced reliability in transmitting application messages across varying sizes.

In terms of MAC/PHY overhead (Fig 4 f), both protocols initially increased, but OLSR's values remained consistent throughout the analysis, while AODV's started to decrease. However, even with this trend, OLSR maintained a lower MAC/PHY overhead compared to AODV. This suggests that OLSR achieved more efficient medium access control and physical layer resource utilization, contributing to a robust and stable communication framework.

In summary, the analysis of application message size reveals that OLSR consistently outperformed AODV across various metrics, emphasizing its suitability for scenarios involving different application message sizes, where it showcased higher throughput, goodput, and lower delay, jitter, lost packets, and MAC/PHY overhead.
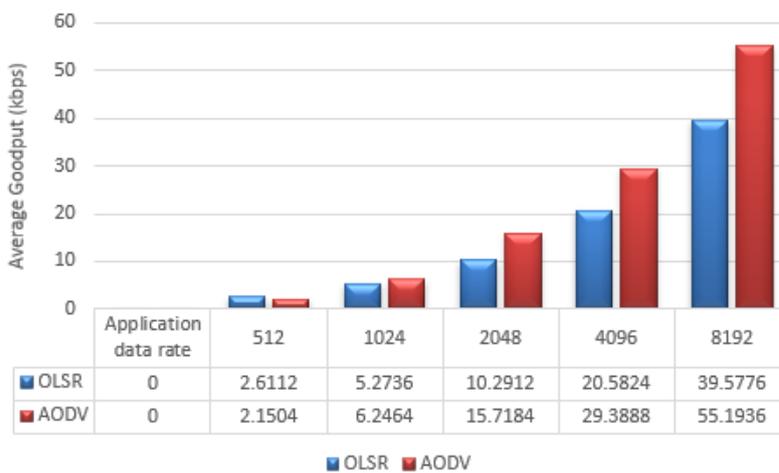

Application data rate	0	512	1024	2048	4096	8192
OLSR	0	2.6112	5.2736	10.2912	20.5824	39.5776
AODV	0	2.1504	6.2464	15.7184	29.3888	55.1936

Figure 5. (a) Performance analysis of Average Goodput with varying application data rate

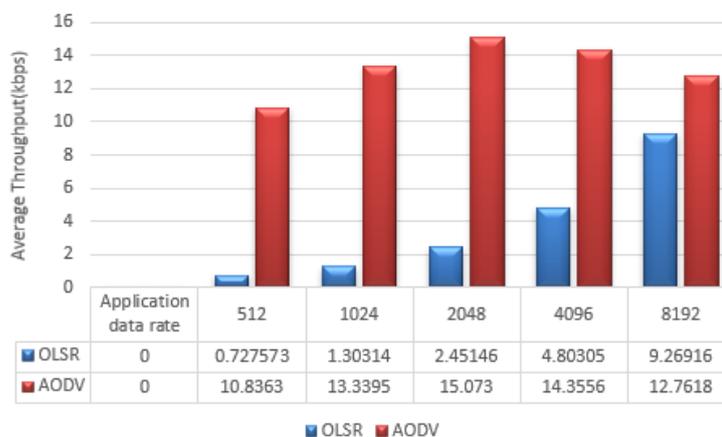| Application data rate	0	512	1024	2048	4096	8192
OLSR	0	0.727573	1.30314	2.45146	4.80305	9.26916
AODV	0	10.8363	13.3395	15.073	14.3556	12.7618

Figure 5. (b) Performance analysis of Average Throughput with varying application data rate
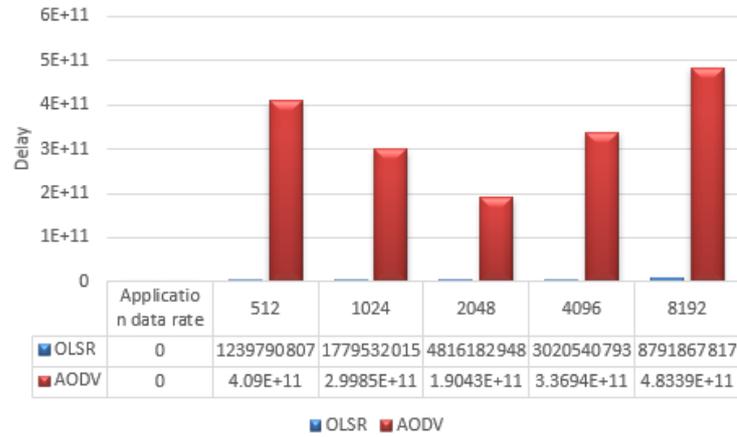
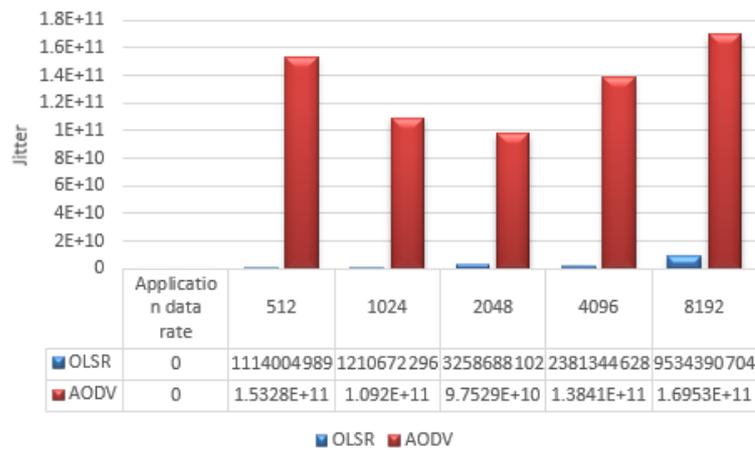Figure 5. (c) Performance analysis of Delay with varying application data rate

Figure 5. (d) Performance analysis of Jitter with varying application data rate
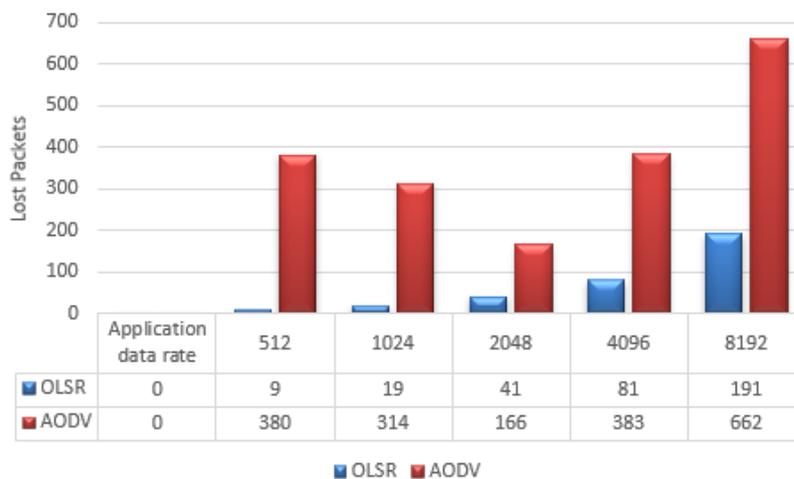

Figure 5. (e) Performance analysis of Lost Packets with varying application data rate
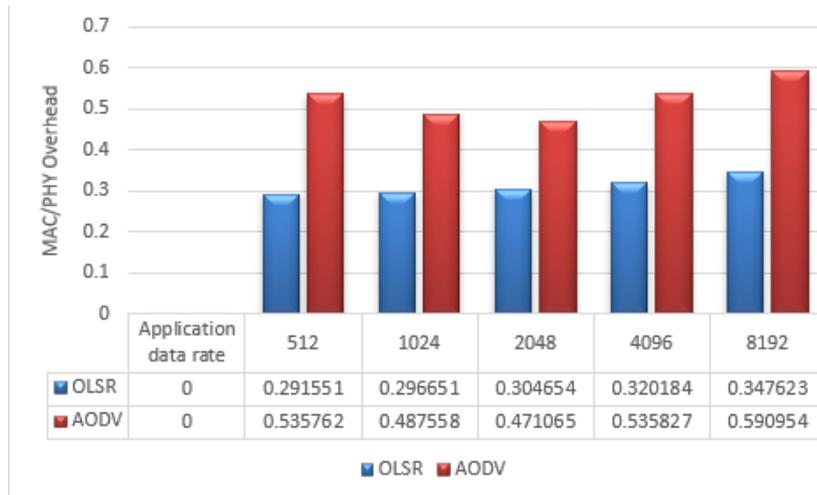
Application data rate		512	1024	2048	4096	8192
OLSR	0	0.291551	0.296651	0.304654	0.320184	0.347623
AODV	0	0.535762	0.487558	0.471065	0.535827	0.590954

Figure 5. (f) Performance analysis of MAC/PHY Overhead with varying application data rate

- **Application data rate analysis:** In the analysis of application data rate, where the value ranged from 512 and doubled to 8192, distinct performance trends emerged for the AODV and OLSR protocols. Examining the various metrics provides valuable insights into the protocols' responsiveness to different application data rates.

  Fig 5 (a) showcased a gradual increase in the average goodput of both protocols, with the OLSR protocol managing to remain consistently lower than the AODV protocol. This suggests that while both protocols experienced improvements in goodput with increasing application data rates, AODV achieved a higher rate of successful data packet delivery compared to OLSR.

  Fig 5 (b) displayed an increase in the average throughput of both protocols, but the AODV protocol exhibited a higher rate of increase compared to OLSR. The final throughput values indicated that AODV maintained a higher average throughput than OLSR, emphasizing its efficiency in achieving successful data transfer with varying application data rates.

  For delay, jitter, and lost packets (Fig 5 c, d, and e), OLSR consistently demonstrated significantly lower values compared to AODV. This indicates that OLSR maintained reduced latency, improved stability, and enhanced reliability in transmitting application data at different rates.

  In terms of MAC/PHY overhead (Fig 5 f), both protocols experienced an increase, but the increase in the AODV protocol was higher, and OLSR's values remained consistently lower. This suggests that OLSR achieved more efficient medium access control and physical layer resource utilization, contributing to a stable communication framework even with varying application data rates.

  In summary, the analysis of application data rate indicates that while both protocols exhibited improvements in various metrics with increasing data rates, AODV demonstrated higher average throughput, and goodput. However, OLSR consistently outperformed AODV in terms of delay, jitter, lost packets, and MAC/PHY overhead, showcasing its suitability for scenarios involving different application data rates where reliability, stability, and efficiency are crucial.
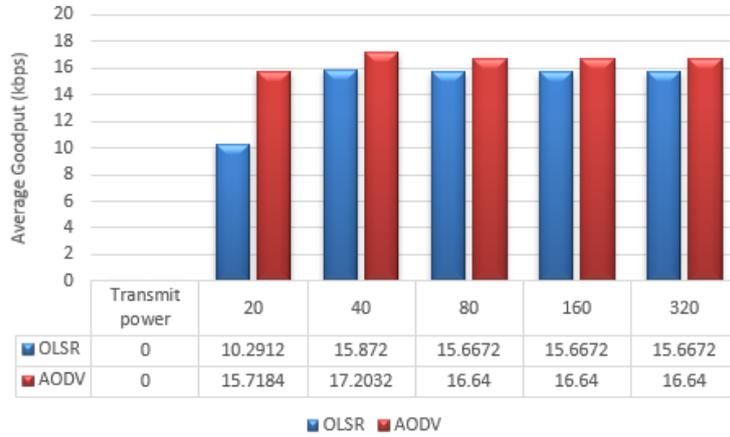
Figure 6. (a) Performance analysis of Average routing goodput with varying transmit power
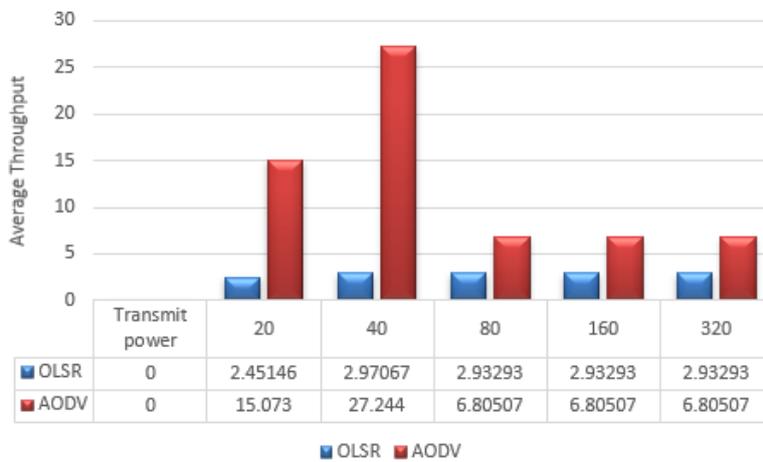

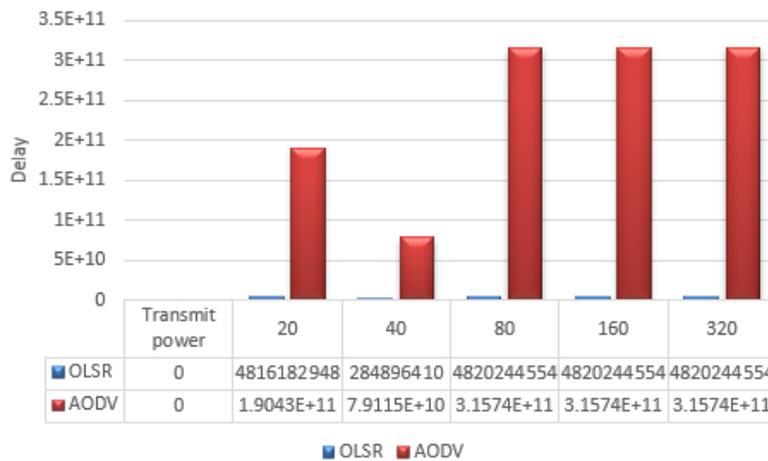Figure 6. (b) Performance analysis of Average throughput with varying transmit power

Figure 6. (c) Performance analysis of delay with varying transmit power

	Transmit power	20	40	80	160	320
OLSR	0	3258688102	357526004	2355811390	2355811390	2355811390
AODV	0	9.7529E+10	3.1619E+10	1.0566E+11	1.0566E+11	1.0566E+11
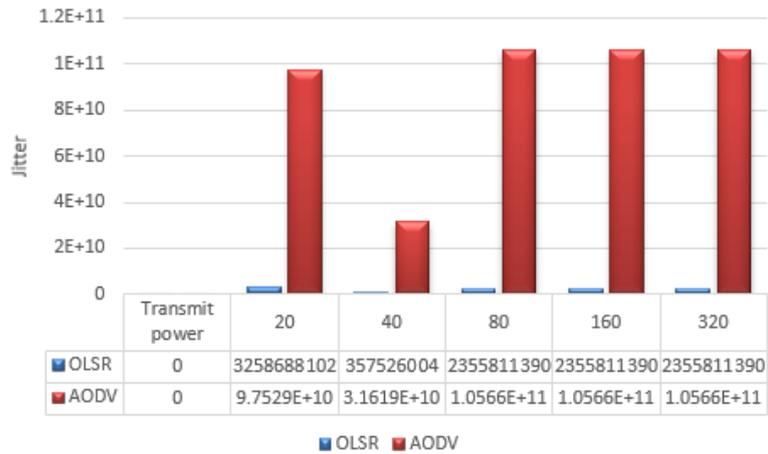
Figure 6. (d) Performance analysis of jitter with varying transmit power

	Transmit power	20	40	80	160	320
OLSR	0	41	0	4	4	4
AODV	0	166	72	255	255	255
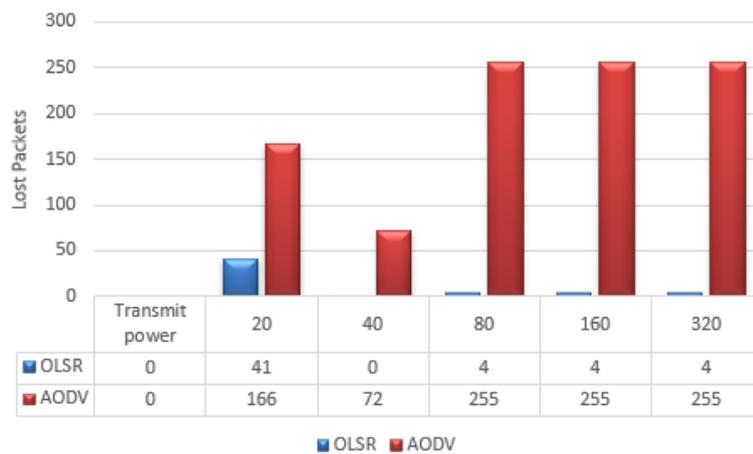
Figure 6. (e) Performance analysis of lost packets with varying transmit power

	Transmit power	20	40	80	160	320
OLSR	0	0.304654	0.310147	0.309913	0.309913	0.309913
AODV	0	0.471065	0.364841	0.493837	0.493837	0.493837

Figure 6. (f) Performance analysis of MAC/PHY overhead with varying transmit power

- **Transmit power analysis:** In the analysis of transmit power, where the power varied from 20 and doubled to 320, distinctive performance characteristics were observed

between the AODV and OLSR protocols. The examination of key metrics sheds light on how the protocols responded to changes in transmit power.

Fig 6 (a) illustrates a gradual increase in the average routing goodput for both protocols. Although the values were very close, the goodput of the AODV protocol ended up slightly higher than that of OLSR. This suggests that, as transmit power increased, both protocols demonstrated improved routing efficiency, but AODV achieved a slightly higher rate of successful data packet delivery.

Fig 6 (b), (c), (d), and (e) presented significant differences in Jitter, Delay, Lost Packets, and Average Throughput between the two protocols. OLSR consistently exhibited considerably lower values in these metrics compared to AODV. This indicates that, with varying transmit power levels, OLSR maintained lower latency, improved stability, reduced packet loss, and achieved higher throughput, highlighting its superior performance in providing reliable and efficient communication.

In Fig 6 (f), which represents MAC/PHY overhead, both protocols experienced an increase, but the MAC/PHY overhead of the OLSR protocol remained consistently lower. Notably, after the transmit power reached 40, the values for OLSR remained the same, showcasing its ability to efficiently utilize medium access control and physical layer resources, even as transmit power increased.

In summary, the transmit power analysis demonstrates that, while AODV achieved a slightly higher average routing goodput, OLSR consistently outperformed AODV in terms of Jitter, Delay, Lost Packets, Average Throughput, and MAC/PHY overhead. OLSR showcased its robustness in adapting to changes in transmit power, ensuring lower latency, higher stability, reduced packet loss, and efficient resource utilization, making it a favorable choice in scenarios where varying transmit power levels are encountered.

	Loss Model	Friis	ItuR1411Los	TwoRayGround
OLSR	0	10.2912	5.5808	10.2912
AODV	0	15.7184	16.5168	15.7184

Figure 7. (a) Performance analysis of Average goodput with varying loss model

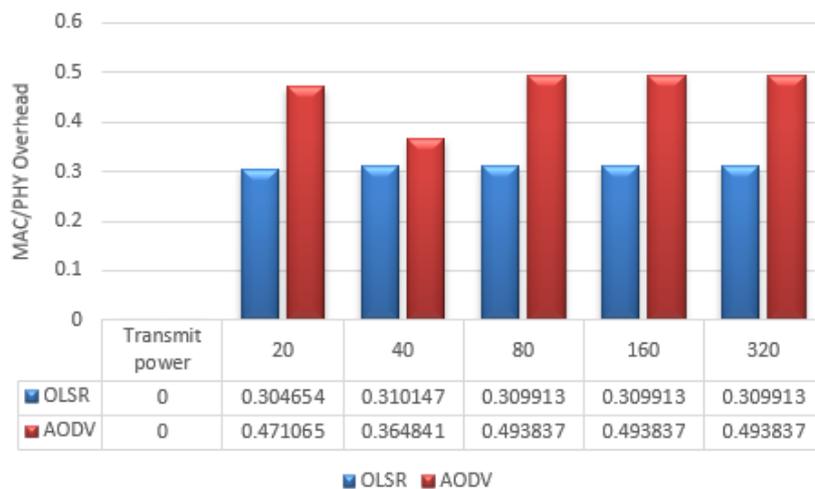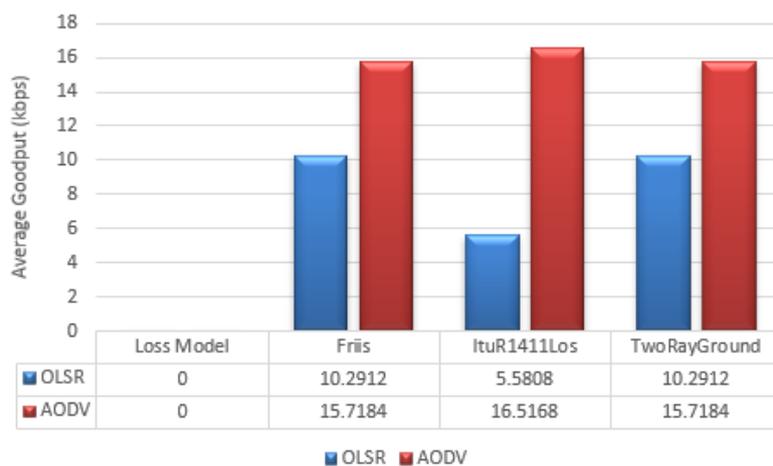Figure 7. (b) Performance analysis of Average throughput with varying loss model

Loss Model		Friis	ItuR1411Los	TwoRayGround
■OLSR	0	2.45146	2.77625	2.45146
■AODV	0	15.073	21.6552	15.073

■OLSR ■AODV
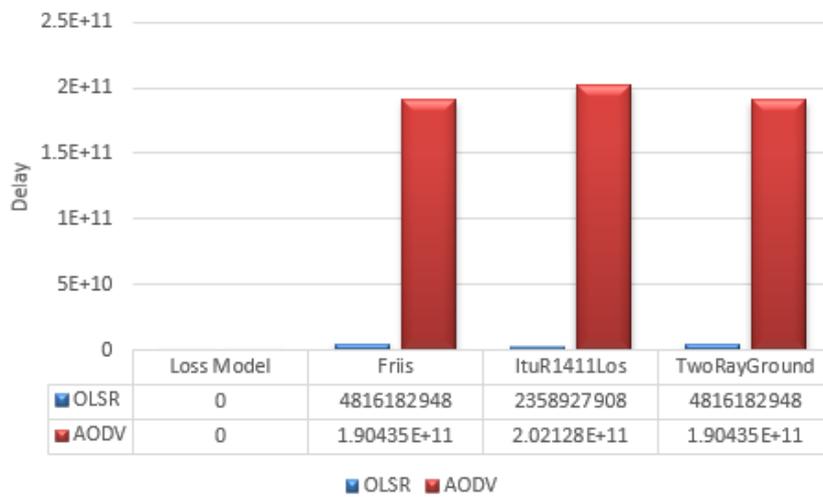
Figure 7. (c) Performance analysis of delay with varying loss model

Loss Model		Friis	ItuR1411Los	TwoRayGround
■OLSR	0	4816182948	2358927908	4816182948
■AODV	0	1.90435E+11	2.02128E+11	1.90435E+11

■OLSR ■AODV

Figure 7. (d) Performance analysis of jitter with varying loss model

Loss Model		Friis	ItuR1411Los	TwoRayGround
■OLSR	0	3258688102	1081087043	3258688102
■AODV	0	97529242705	75700554128	97529242705

■OLSR ■AODV

Figure 7. (e) Performance analysis of Lost Packets with varying loss model

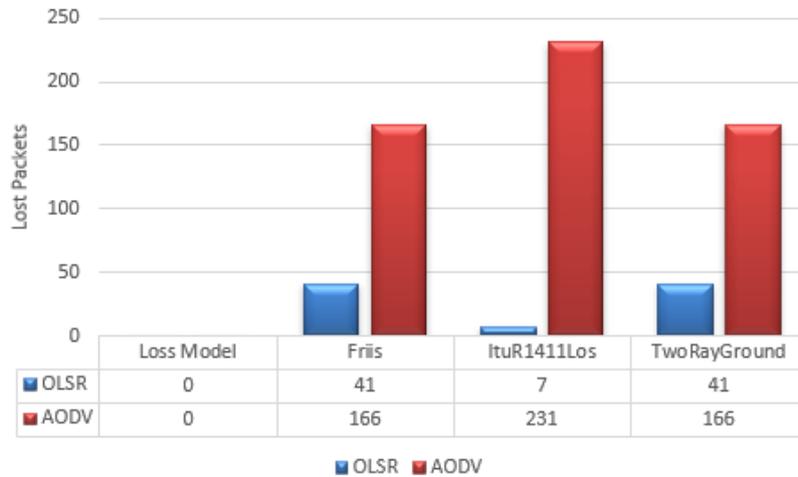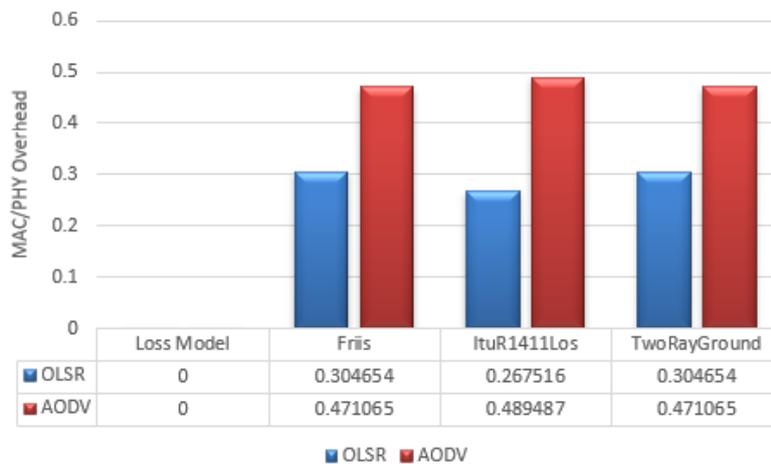Figure 7. (f) Performance analysis of MAC/PHY overhead with varying loss model

- **Loss Model analysis:** In the analysis of the loss model, where the model varies from Friis to ItuR1411Los to TwoRayGround, distinct performance trends emerged between the AODV and OLSR protocols. The examination of key metrics provides insights into how the protocols responded to different loss models.

  Fig 7 (a) depicted the Average Routing Goodput, revealing that AODV experienced a higher increase compared to OLSR as the loss model varied. This suggests that, under changing loss models, AODV demonstrated a relatively higher rate of successful data packet delivery compared to OLSR.

  Fig 7 (b), (c), (d), and (e) demonstrated significant differences in Jitter, Delay, and Average Throughput and Lost packets between the two protocols. OLSR consistently exhibited considerably lower values in these metrics across various loss models. This indicates that, regardless of the loss model used, OLSR maintained lower latency, improved stability, and achieved higher throughput, highlighting its superior performance in providing reliable and efficient communication.

Fig 7 (f), representing MAC/PHY overhead, showcased that both protocols experienced an increase, but the MAC/PHY overhead of the OLSR protocol remained consistently lower. This suggests that OLSR efficiently utilized medium access control and physical layer resources across different loss models.

In summary, the loss model analysis demonstrates that AODV experienced a higher increase in Average Routing Goodput, but OLSR consistently outperformed AODV in terms of Jitter, Delay, Average Throughput, and MAC/PHY overhead across various loss models. OLSR's robust performance under different loss conditions, ensuring lower latency, improved stability, and efficient resource utilization, reinforces its suitability for diverse VANET deployment scenarios.

Number of sinks	0	5	10	20
OLSR	0	6.8608	10.2912	24.9856
AODV	0	7.5264	15.7184	28.9792
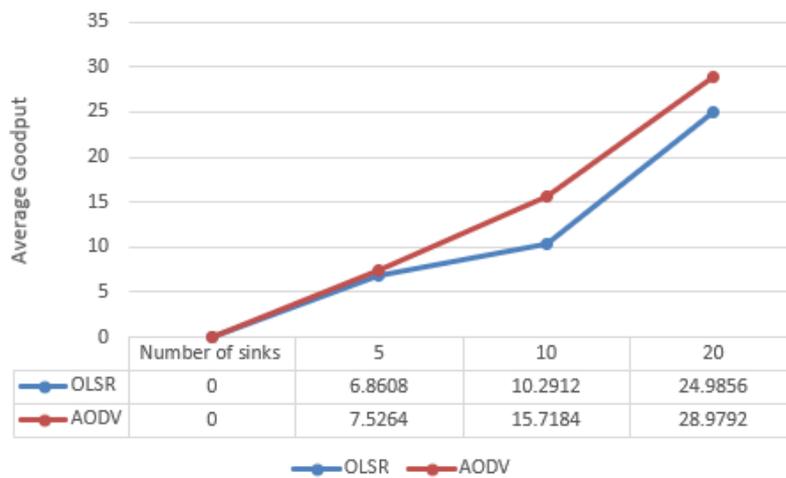
Figure 8. (a) Performance analysis of Average goodput with varying number of sinks

Number of sinks	0	5	10	20
OLSR	0	2.98823	2.45146	2.72521
AODV	0	19.8627	15.073	16.7798
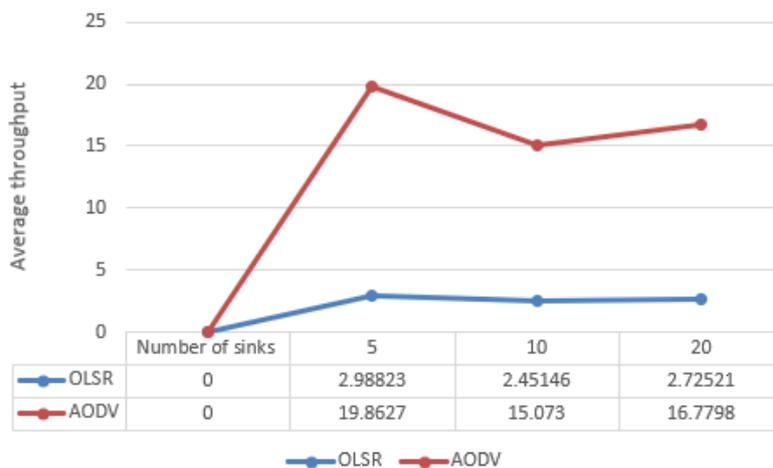
Figure 8. (b) Performance analysis of Average throughput with varying number of sinks

Number of sinks		5	10	20
OLSR	0	1000108992	4816182948	11427069116
AODV	0	1.76466E+11	1.90435E+11	4.39617E+11
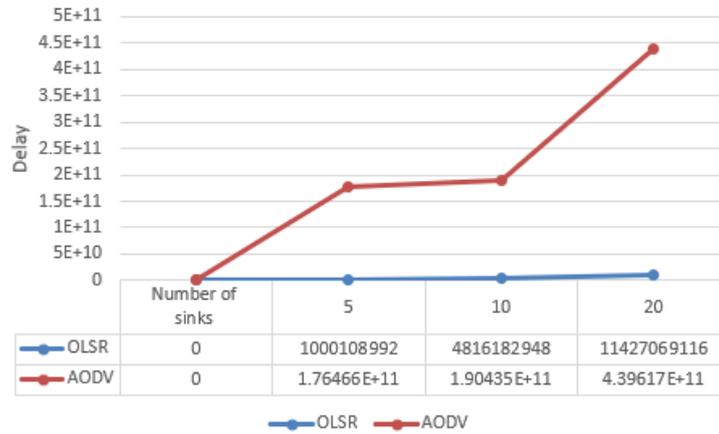
Figure 8. (c) Performance analysis of Delay with varying number of sinks

Number of sinks		5	10	20
OLSR	0	1799689932	3258688102	10545239026
AODV	0	64500411720	97529242705	1.31416E+11
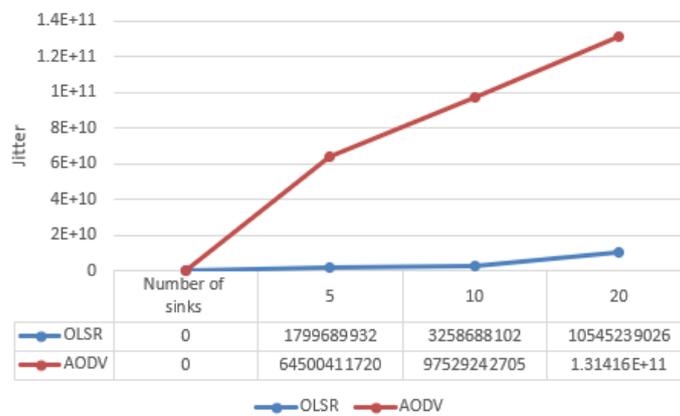
Figure 8. (d) Performance analysis of Jitter with varying number of sinks

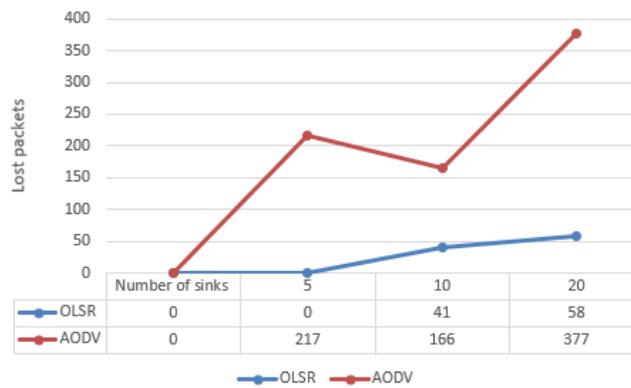| Number of sinks		5	10	20
OLSR	0	0	41	58
AODV	0	217	166	377

Fig. 8. (e) Performance analysis of Lost Packets with varying number of sinks
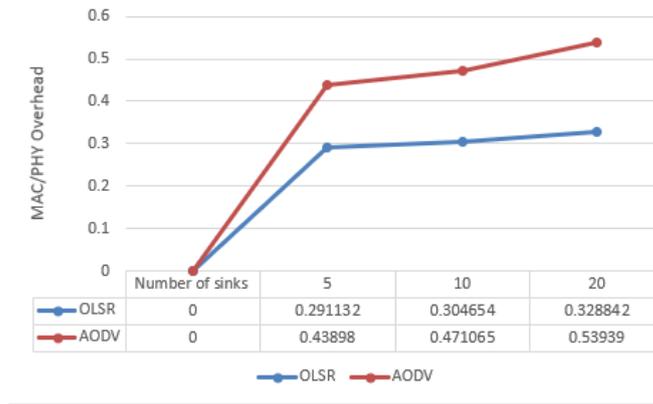
Fig. 8. (f) Performance analysis of MAC/PHY Overhead with varying number of sinks

- **Number of sinks analysis:** In the analysis of the number of sinks, where the number varies from 5 and doubles to 20, key metrics shed light on the comparative performance of AODV and OLSR protocols in response to changes in the number of sinks.

  Fig 8 (a) portrayed the Average Routing Goodput, revealing a gradual increase for both AODV and OLSR protocols. Despite the close values, the average routing goodput of the AODV protocol was slightly higher than OLSR. This suggests that, with an increasing number of sinks, AODV demonstrated a marginally higher rate of successful data packet delivery compared to OLSR.

  Fig 8 (b), (c), (d), and (e) demonstrated substantial differences in Jitter, Delay, Average Throughput, and Lost Packets between the two protocols. OLSR consistently exhibited significantly lower values in these metrics as the number of sinks increased. This indicates that irrespective of the number of sinks, OLSR maintained lower latency, improved stability, achieved higher throughput, and experienced fewer lost packets, highlighting its superior performance in providing reliable and efficient communication in scenarios with varying sink densities.

  Fig 8 (f), representing MAC/PHY overhead, showcased an increase for both protocols with the growing number of sinks. However, the MAC/PHY overhead of the AODV protocol was consistently higher than that of OLSR. This suggests that OLSR efficiently utilized medium access control and physical layer resources across different sink densities.

  In summary, the number of sinks analysis indicates that AODV exhibited a slightly higher average routing goodput, but OLSR consistently outperformed AODV in terms of Jitter, Delay, Average Throughput, Lost Packets, and MAC/PHY overhead across different sink densities. OLSR's robust performance under varying sink conditions reinforces its suitability for diverse VANET deployment scenarios, providing lower latency, improved stability, and efficient resource utilization.

	Number of nodes	50	100	156	200	250	300	350
OLSR	0	2.45146	2.64292	2.45146	2.36211	2.58349	2.2557	2.87922
AODV	0	15.073	8.53484	15.073	5.98313	8.90491	11.3066	11.3066
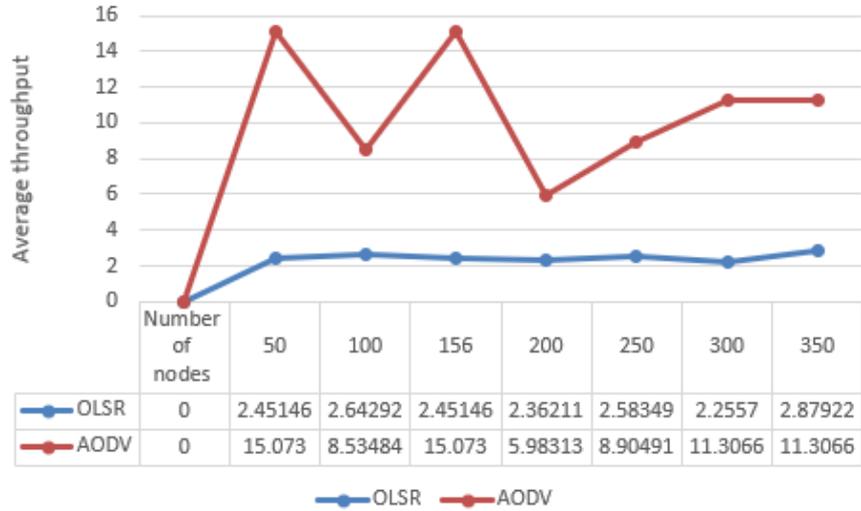
Figure 9. (a) Performance analysis of Average throughput with varying number of nodes

	Number of nodes	50	100	156	200	250	300	350
OLSR	0	10.2912	8.6528	10.2912	7.168	7.8848	4.864	7.6288
AODV	0	15.7184	9.728	15.7184	10.0352	11.3152	9.728	9.728
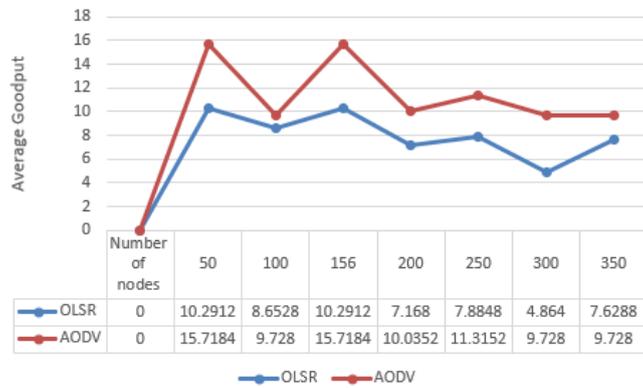
Figure 9. (b) Performance analysis of Average goodput with varying numbers of nodes

	Number of nodes	50	100	156	200	250	300	350
OLSR	0	4.8E+09	3.1E+09	4.8E+09	1E+10	1.3E+10	1E+10	8.8E+09
AODV	0	1.9E+11	8.1E+11	1.9E+11	1.1E+12	7.3E+11	3.5E+11	3.5E+11
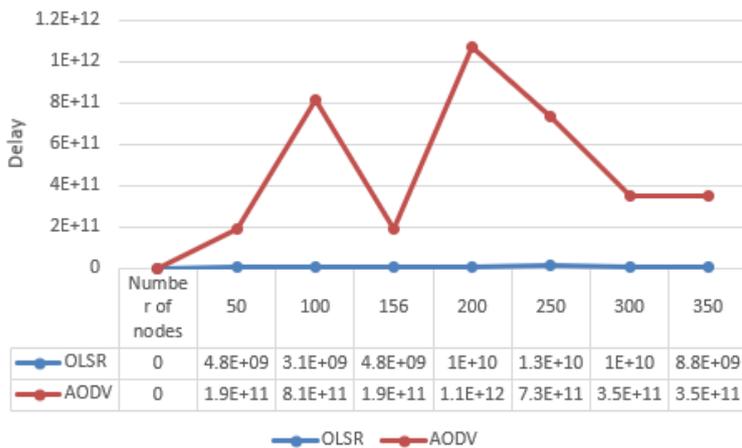
Figure 9. (c) Performance analysis of Delay with varying numbers of nodes

Number of nodes		50	100	156	200	250	300	350
OLSR	0	3.3E+09	4.9E+09	3.3E+09	1.5E+10	1.6E+10	1E+10	1.1E+10
AODV	0	9.8E+10	2.5E+11	9.8E+10	2.7E+11	2E+11	6.8E+10	6.8E+10
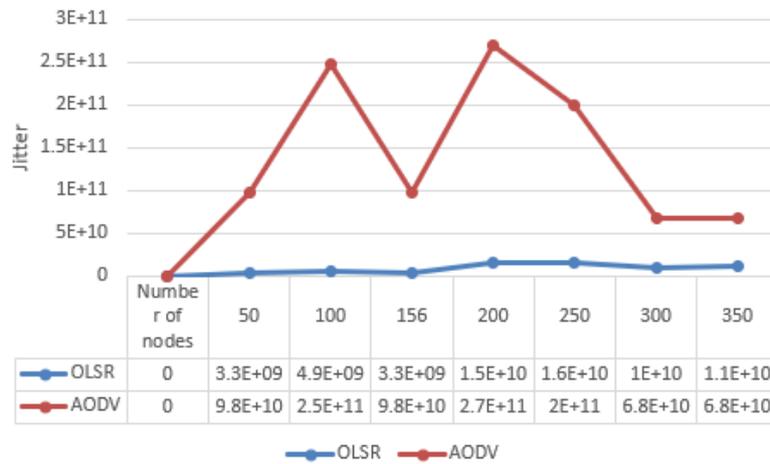
Figure 9. (d) Performance analysis of Jitter with varying numbers of nodes

Number of nodes		50	100	156	200	250	300	350
OLSR	0	41	26	41	41	35	38	6
AODV	0	166	1430	166	2387	1558	1104	1104
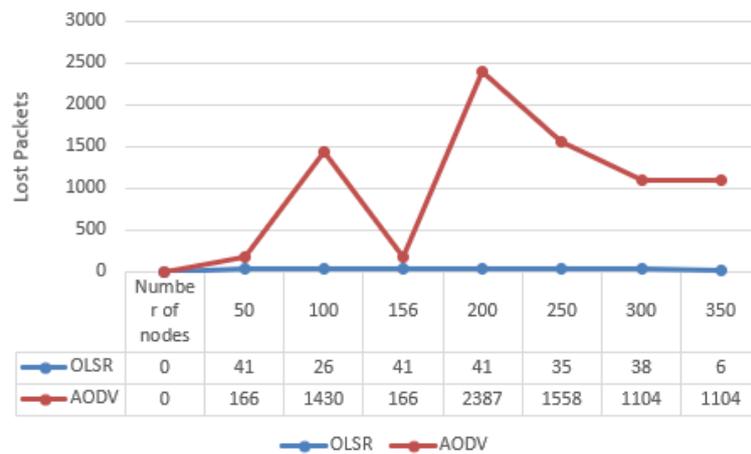
Figure 9. (e) Performance analysis of Lost packets with varying numbers of nodes

Number of nodes		50	100	156	200	250	300	350
OLSR	0	0.304654	0.31376	0.304654	0.355401	0.372852	0.386516	0.397185
AODV	0	0.471065	0.56916	0.471065	0.498428	0.470949	0.384783	0.384783

Fig. 9. (f) Performance analysis of MAC/PHY Overhead with varying numbers of nodes

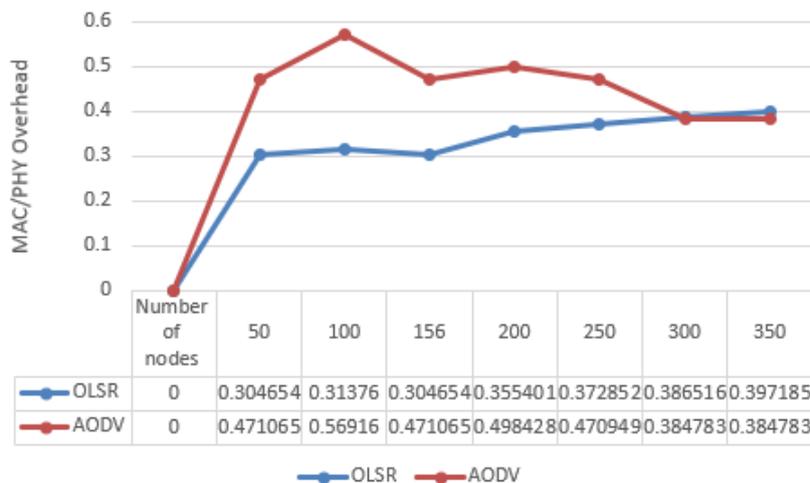- **Numbers of nodes analysis:** In the analysis of the number of nodes, where the number varies from 50 to 100 to 156 to 200 to 250 to 300 to 350, key metrics provide insights

into the comparative performance of AODV and OLSR protocols under different node densities.

Fig 9 (a), (c), (d), and (e) illustrated substantial differences in Jitter, Delay, Average Throughput, and Lost Packets between the two protocols. OLSR consistently exhibited significantly lower values in these metrics as the number of nodes increased. This indicates that, regardless of the number of nodes, OLSR maintained lower latency, improved stability, achieved higher throughput, and experienced fewer lost packets, highlighting its superior performance in providing reliable and efficient communication in scenarios with varying node densities.

Fig 9 (b) portrayed the Average Routing Goodput, revealing that the difference between AODV and OLSR was not drastic, and the AODV protocol's values remained slightly higher than OLSR. This suggests that, with an increasing number of nodes, AODV demonstrated a slightly higher rate of successful data packet delivery compared to OLSR.

Fig 9 (f), representing MAC/PHY overhead, showcased an increase for both protocols with the growing number of nodes. Interestingly, the MAC/PHY overhead of the AODV protocol initially increased but then decreased to become lower than OLSR. This suggests that AODV optimized resource utilization more effectively as the node density increased.

In summary, the number of nodes analysis indicates that OLSR consistently outperformed AODV in terms of Jitter, Delay, Average Throughput, and Lost Packets across different node densities. While AODV exhibited slightly higher average routing goodput initially, the gap was not substantial. Moreover, AODV managed to decrease its MAC/PHY overhead as node density increased, indicating its adaptability to varying network sizes. The choice between AODV and OLSR may depend on specific deployment scenarios and priorities, considering their nuanced performance characteristics.

	Node speed	5	10	40	80	120	180
OLSR	0	10.3936	10.8032	11.8784	12.0832	12.6976	13.2096
AODV	0	12.5952	12.8	11.3664	13.1072	16.0256	15.2064
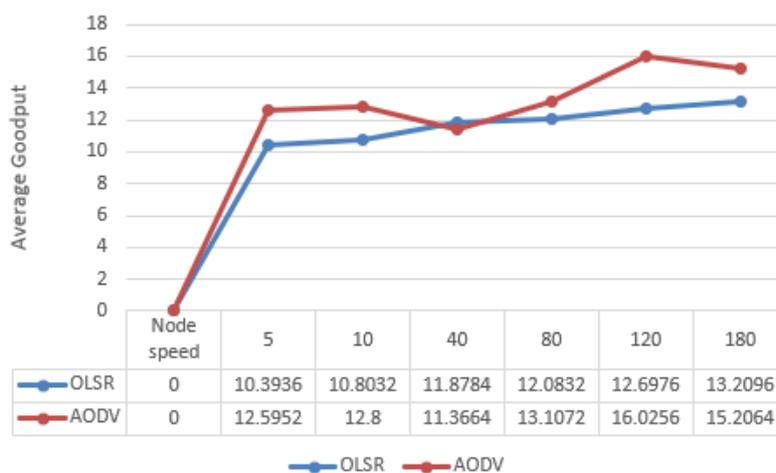
Figure 10. (a) Performance analysis of Average Goodput with varying node speed

Node speed		5	10	40	80	120	180
OLSR	0	0.294933	0.304283	0.300954	0.304897	0.301386	0.311004
AODV	0	0.543154	0.52023	0.501706	0.45048	0.505961	0.502564
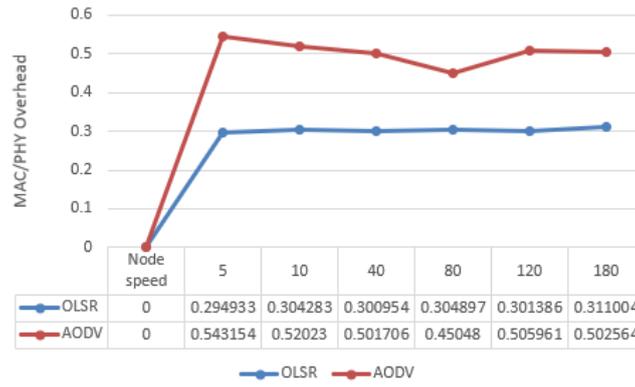
Figure 10. (b) Performance analysis of MAC/PHY Overhead with varying node speed

Node speed		5	10	40	80	120	180
OLSR	0	16	22	17	12	2	16
AODV	0	407	481	481	241	465	471
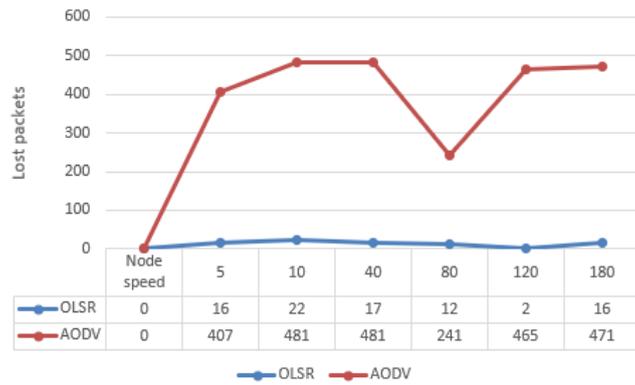
Figure 10. (c) Performance analysis of Lost packets with varying node speed

Node speed		5	10	40	80	120	180
OLSR	0	2.99915	2.81787	2.81116	2.85444	2.97821	2.95511
AODV	0	16.0394	11.763	18.4361	17.2993	19.8662	16.6133
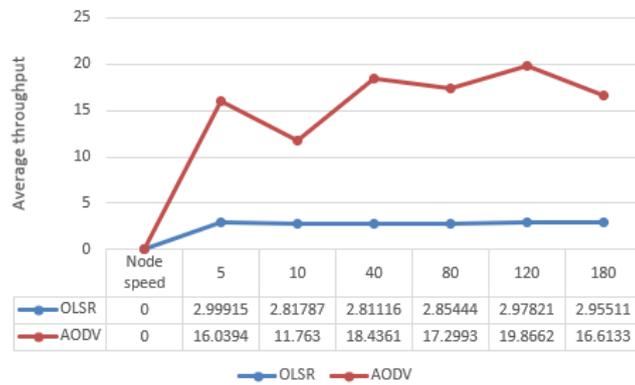
Figure 10. (d) Performance analysis of Average throughput with varying node speed

Node speed		5	10	40	80	120	180
OLSR	0	2.55E+09	2.67E+09	2.42E+09	2.46E+09	4.39E+09	3.29E+09
AODV	0	3.4E+12	3.49E+11	3.26E+11	2.15E+11	2.58E+11	3.37E+11
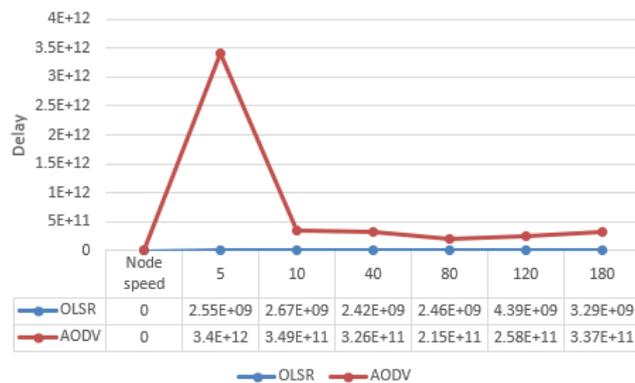
Figure 10. (e) Performance analysis of Delay with varying node speed

	Node speed	50	100	156	200	250	300
OLSR	0	2.24E+09	2.34E+09	1.24E+09	1.31E+09	6.17E+09	6.27E+09
AODV	0	1.34E+11	1.4E+11	1.13E+11	7.78E+10	9.07E+10	1.22E+11

Figure 10. (f) Performance analysis of Jitter with varying node speed

- **Speed node analysis:** In the node speed analysis, where the speed of nodes varies from 5 to 10 to 40 to 80 to 120 to 180, the performance metrics shed light on the comparative behavior of AODV and OLSR protocols under different node speeds.

  Fig 10 (c), (d), (e), and (f) demonstrated significant differences in Jitter, Delay, Average Throughput, and Lost Packets, with OLSR consistently exhibiting notably lower values than AODV. This indicates that, irrespective of node speed, OLSR maintained lower latency, improved stability, achieved higher throughput, and experienced fewer lost packets, emphasizing its superior performance in scenarios with varying node speeds.

  Fig 10 (a) portrayed the Average Routing Goodput, revealing that the difference between AODV and OLSR was not substantial, and the AODV protocol's values remained slightly higher than OLSR throughout the analysis. This suggests that, with varying node speeds, AODV demonstrated a slightly higher rate of successful data packet delivery compared to OLSR.

  Fig 10 (b), representing MAC/PHY overhead, showcased an increase for both protocols with the escalating node speeds. However, the MAC/PHY overhead of the AODV protocol was consistently higher than OLSR. This suggests that OLSR optimized resource utilization more effectively, even in scenarios with higher node speeds.

  In summary, the node speed analysis indicates that OLSR consistently outperformed AODV in terms of Jitter, Delay, Average Throughput, and Lost Packets across different node speeds. While AODV exhibited slightly higher average routing goodput, the difference was not substantial. Moreover, OLSR managed to maintain lower MAC/PHY overhead, indicating its efficiency in resource utilization, especially in scenarios with varying node speeds. The choice between AODV and OLSR may depend on specific deployment scenarios and priorities, considering their nuanced performance characteristics.

Impact on Vanet Applications

The observed performance variations between AODV and OLSR protocols carry profound implications for the effectiveness of VANET applications. OLSR consistently outperforms AODV in key metrics such as delay, jitter, lost packets, and average throughput, making it particularly well-suited for safety-critical applications within VANETs. In scenarios where

split-second decision-making is imperative, such as collision avoidance or emergency message dissemination, OLSR's lower delay and jitter, coupled with minimal packet loss, can significantly enhance the reliability and responsiveness of these applications.

Moreover, the advantages demonstrated by OLSR contribute to improved overall vehicular communication, which is pivotal for the success of Intelligent Transportation Systems (ITS) in real-world environments. The ability to maintain low latency and high throughput is crucial for applications beyond safety, such as traffic management, intelligent navigation, and cooperative driving experiences. The impact on VANET applications underscores the critical role that communication protocols play in shaping the efficiency and reliability of modern vehicular networks.

Recommendations and Implications

The evaluation results present clear insights into the comparative performance of AODV and OLSR protocols within VANETs. Based on these findings, several recommendations and implications arise for the deployment and optimization of vehicular communication systems.

Protocol Selection Based on Application Requirements: The choice between AODV and OLSR should be guided by the specific requirements of VANET applications. For safety-critical applications demanding low latency and minimal jitter, OLSR emerges as a preferred choice. However, AODV might find its niche in scenarios where sporadic communication and on-demand routing are more suitable.

Consideration of Safety-Critical Applications: Safety applications, which demand real-time and reliable communication, benefit significantly from the superior performance of OLSR. VANETs designed for collision avoidance, emergency braking, and rapid response to dynamic road conditions can leverage OLSR to ensure timely and accurate data transmission.

Network Scalability and Density: The evaluation under varying numbers of nodes and sinks highlights the scalability of both protocols. However, AODV exhibits higher MAC/PHY overhead in densely populated networks. Implementers should consider the expected density of vehicular networks when choosing a protocol to ensure efficient communication.

Adaptability to Changing Network Conditions: OLSR's proactive nature, continuously updating routing tables, makes it adept at handling changing network conditions. For applications requiring adaptability to dynamic environments, such as urban traffic with frequent stops and starts, OLSR may provide more consistent and reliable performance.

Integration with Emerging Technologies: As VANETs continue to evolve, integrating these protocols with emerging technologies like 5G and beyond becomes essential. Future work should explore how AODV and OLSR integrate with advanced communication infrastructures to meet the demands of next-generation vehicular networks.

CONCLUSION

In conclusion, this exploration into Vehicular Ad-Hoc Networks (VANETs) within the broader context of Intelligent Transportation Systems (ITS), Internet of Things (IoT), and Machine-to-

Machine (M2M) systems reveals the transformative potential of these technologies in reshaping modern transportation. The comprehensive analysis of VANET fundamentals, challenges, and opportunities, with a specific emphasis on their integration into smart parking systems, underscores the pivotal role they play in enhancing road safety, traffic efficiency, and urban mobility.

By comparing routing protocols, particularly the Ad Hoc On-Demand Distance Vector (AODV) and Optimized Link State Routing (OLSR), we gain critical insights into optimal strategies for reliable VANET functioning in diverse traffic scenarios. The examination of communication architectures, including Vehicle-to-Vehicle (V2V), Vehicle-to-Infrastructure (V2I), and Vehicle-to-Everything (V2X), emphasizes the significance of wireless connectivity in sharing real-time information, thereby contributing to enhanced safety measures and overall passenger experience.

For this work, we aimed to present to you the value of this system and its impact when it comes to solving rising problems, specifically problems related to smart parking. To do that, not only did we do thorough research on this topic, but we also did some analysis with the help of NS3 simulation with the chosen protocols to analyze their performances regarding the establishment and maintenance of communication in dynamic vehicular environments, but we did not yet propose any other solutions to the main challenges encountered by the other scholars who studied this topic.

For our next work, we will aim to, not only provide a wider explanation of this system, but we will also propose solutions that, we believe, will solve various issues and challenges encountered by other researchers and will help improve the performance of this system. As we look toward the future, our research will extend beyond the current horizons to address emerging challenges and explore innovative solutions within the realm of VANET systems. Future work will focus on:

Investigating and developing advanced routing protocols to further enhance the efficiency and reliability of data transmission in VANETs.

Exploring robust security measures and privacy-preserving techniques to address concerns related to the collection and dissemination of sensitive data within VANETs.

Addressing the challenges associated with the scalability of VANET applications and promoting interoperability between vehicles, infrastructure, and smart parking systems.

Continuously optimizing VANET applications, particularly in the context of smart parking, to ensure efficient space utilization and improved overall urban mobility.

Exploring the integration of VANETs with emerging technologies such as 5G networks, edge computing, and artificial intelligence to unlock new potentials and capabilities.

By delving deeper into these avenues, our future work aims to contribute to the ongoing evolution of VANET systems, ensuring their resilience, adaptability, and effectiveness in addressing the dynamic challenges of modern transportation and smart city initiatives.

In essence, the journey into VANETs is not only a reflection of our commitment to advancing transportation technologies but also an invitation to continuously explore and innovate for a safer, more efficient, and interconnected future.

REFERENCES

[1] P. K. Singh, S. K. Nandi, and S. Nandi, "A tutorial survey on vehicular communication state of the art, and future research directions," Vehicular Communications, vol. 18, p. 100164, 2019.

[2] Z. Sheng, A. Pressas, V. Ocheri, F. Ali, R. Rudd, and M. Nekovee, *"Intelligent 5g vehicular networks: An integration of dsrc and mmwave communications,"* in 2018 International Conference on Information and Communication Technology Convergence (ICTC), 2018, pp. 571–576.

[3] W. Qi, B. Landfeldt, Q. Song, L. Guo, and A. Jamalipour, "Traffic differentiated clustering routing in dsrc and c-v2x hybrid vehicular networks," IEEE Transactions on Vehicular Technology, vol. 69, no. 7, pp. 7723–7734, 2020

[4] J. Shi, L. Jin, J. Li, and Z. Fang, "A Smart Parking System Based on NB-IoT and Third-party Payment Platform," 2017.

[5] Y. Kang, D. Jung, and I. Doh, "Automated Parking Lot Management System using Embedded Robot Type Smart Car based on Wireless Sensors," 2017.

[6] J. Ni, K. Zhang, Y. Yu, X. Lin, and X. S. Shen, "Privacy-preserving Smart Parking Navigation Supporting Efficient Driving Guidance Retrieval," IEEE Trans. Veh. Technol., vol. 9545, no. c, pp. 1–14, 2018.

[7] R. K. Lenka, "PSPS : An IoT Based Predictive Smart Parking System," pp. 311–317, 2017.

[8] B. M. Mahendra, S. Sonoli, N. Bhat, Raju, and T. Raghu, "IoT based sensor-enabled smart car parking for advanced driver assistance system," 2017 2nd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol., pp. 2188–2193, 2017.

[9] Nursalim, H. M. Saputra, N. Ismail, and T. S. Gunawan, "Kinematic analysis of rotary car parking system mechanism," in 2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA), 2017, pp. 1–5.

[10] M. ur Rehman and M. A. Shah, "A smart parking system to minimize searching time, fuel consumption and CO2 emission," in 2017 23rd International Conference on Automation and Computing (ICAC), 2017, pp. 1–6.

[11] P. Nguyen, H. Nguyen, D. Nguyen, T. N. Dinh, H. M. La, and T. Vu, "ParkSense: Automatic Parking Positioning by Leveraging In Vehicle Magnetic Field Variation," IEEE Access, vol. 5, pp. 25021–25033, 2017.

[12] Ahmad, R. M. Noor, I. Ahmedy, S. A. A. Shah, I. Yaqoob, E. Ahmed, and M. Imran, "Vanet–lte based heterogeneous vehicular clustering for driving assistance and route planning applications," Computer Networks, vol. 145, pp. 128–140, 2018.

[13] S. Hamad and T. Yeferny, "A smart data dissemination protocol for vehicular ad-hoc networks," International Journal of Computer Science and Network Security, vol. 19, p. 176, 2019

[14] D. Praveen and D. P. Raj, "Smart traffic management system in metropolitan cities," Journal of Ambient Intelligence and Humanized Computing, pp. 1–13, 2020

[15] B. Mukhopadhyay and T. Samanta, "A Smart Parking-lot Occupancy Model in 5G V2V and V2I Wireless Communication," in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–6, IEEE, 2021.

[16] S. Yadav, N. K. Rajput, Dr. A. K. Sagar, D. Maheshwari, et al., "Secure and Reliable Routing Protocols for VANETs," in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, pp. 1–6, 2018.

RESUME

Joseph MOSENGO

Joseph Mosengo is currently a student at Final International University, located in Girne, North Cyprus, where he is pursuing a degree in Computer Engineering at the Faculty of Engineering. He began his studies in 2019 and is expected to graduate in January 2025. Prior to this, he completed his high school education at Sainte Famille in the Democratic Republic of Congo, earning his diploma in 2018. Joseph is conducting research on "A Practical Evaluation of VANET Routing Protocols in Modern Transportation Systems," which is expected to be published soon. He presented his work at the International Conference on Engineering Technologies in November 2024. Joseph has developed strong skills in both frontend and backend development. He is proficient in using modern frontend technologies such as HTML, CSS, JavaScript, React, and Bootstrap to create responsive and user-friendly web applications. On the backend, he is experienced with platforms like Node.js and Django, utilizing them to build robust, scalable server-side solutions. His deep understanding of both ends of development allows him to create full-stack applications that are well-integrated and efficient. In addition to his academic pursuits, Joseph is a simultaneous interpreter, particularly involved in church activities. He is passionate about language and communication, further honing his skills through his interpretation work.

Asist. Prof. Dr. Mostafa Ayoubi MOBARHAN

Mostafa Ayoubi Mobarhan is an Assistant Professor in the Engineering Faculty at Final International University, where he has been a faculty member since 2022. He received his BSc degree in Electrical Engineering (Electronics) from Islamic Azad University of Lahijan, Iran, in 2008. In 2009, he joined the Department of Computer Engineering at the University of Guilan, Iran, as a full-time MSc student and earned his MSc degree in 2011. He later pursued his Ph.D. in the Computer Engineering Department of Eastern Mediterranean University (EMU), focusing on advanced topics in wireless networks and network security.

Dr. Ayoubi Mobarhan's research interests include wireless networks, network security, cryptographic algorithms, and mobile communications. His work contributes to enhancing secure and efficient communication in modern networked environments. Since joining Final International University, he has been actively involved in both teaching and research, striving to inspire the next generation of engineers and researchers.

Sen. Ins. Atefeh Ahmadniai KHAJEKINI

Atefeh Ahmadniai Khajekini received the BSc degree in Electrical and Electronic Engineering from the Islamic Azad University of Lahijan (Iran) in 2008. In October 2013, she joined the Department of Computer Engineering at the University of Guilan, Iran, as a full-time MSc student. She received the MSc degree in 2015. She is a Ph.D. student under advisor Ali Hakan Ulusoy in the Electrical and Electronic Engineering Department of Eastern Mediterranean University (EMU). She is a senior instructor with the Electrical and Electronic Engineering

Department at the Final International University in Northern Cyprus. Her research interests include computer networks, wireless networks, network security, cryptography algorithms, and mobile communications.

Benie PONTE

Benie Ponte is currently a student at Final International University in Girne, North Cyprus, where she is pursuing a degree in Computer Engineering at the Faculty of Engineering. She started her studies in 2020 and is expected to graduate in January 2025. Benie has collaborated on research titled "A Practical Evaluation of VANET Routing Protocols in Modern Transportation Systems," which is set to be published soon. She also presented her work at the International Conference on Engineering Technologies in November 2024, where she received a certificate of participation.

Keren TUZOLANA

Keren Matondo Tuzolana is a Computer Engineering student at Final International University in North Cyprus. She began her academic journey in 2021 and is now on the verge of graduating, equipped with a wealth of knowledge and skills in her field. Passionate about technology and innovation, Keren chose this path to explore digital solutions that transform daily life and improve efficiency. Beyond her technical studies, Keren has a strong passion for organization and planning, which reflects her natural inclination for structure and efficiency. She is also deeply interested in management and personal development, believing in the importance of continuous growth and self-improvement. These interests, combined with her technical expertise, allow her to envision innovative solutions while implementing effective strategies. During her studies, Keren completed an internship where she developed a web application designed to streamline her university's academic registration process. The application automated the display of student grades using a color-coded system, demonstrating her skills in programming, problem-solving, and system development. Keren is also committed to inspiring and empowering young women to pursue careers in technology and break barriers in traditionally male-dominated fields. She aspires to make a meaningful impact in education and promote inclusivity in tech, using her experiences to advocate for greater opportunities for women. With determination, creativity, and leadership, Keren is poised to graduate as a skilled engineer, ready to tackle the challenges of tomorrow and leave a lasting mark in the fields of technology and education.

INNOVATIVE END EFFECTOR DESIGN FOR ROBOTIC DNA EXTRACTION APPLICATIONS

Burak YILMAZ[1], Veysel FIRAT[2]

INTRODUCTION

The field of robotics has seen a growing demand for precise and efficient DNA extraction techniques, particularly in the context of genomic research and diagnostics. Conventional DNA extraction methods, while effective, can be labor-intensive, time-consuming, and require specialized equipment, limiting their scalability and accessibility.

To address these challenges, researchers have explored the potential of microfluidic technology to develop innovative end effector designs for robotic DNA extraction applications. These microfluidic-based systems offer several advantages, including reduced sample volume, faster processing times, and the ability to integrate multiple steps (e.g., lysis, purification, elution) into a single, automated platform. [1] [2]

One such approach is the use of paramagnetic beads for DNA extraction, as described in the literature. In this method, a mixture of blood lysate, paramagnetic beads, and binding buffer is placed into the input well of a microfluidic device. The DNA-bound paramagnetic beads are then pulled through a central channel containing a wash buffer to the output well, which contains the elution buffer. This streamlined process eliminates the need for repetitive pipetting and centrifugation steps, significantly reducing the labor and time required for DNA extraction. Additionally, the suitability of microfluidic technology in performing standard DNA extraction and purification protocols with integration into downstream applications, such as PCR amplification and sample preparation, has been investigated by numerous researchers [1 - 3]. These studies have demonstrated the potential of microfluidic-based systems to provide a sensitive, robust, and reliable integrated method of DNA extraction that is rapid, cost-effective, and space-efficient.

In modern biotechnology laboratories, DNA isolation and liquid transfer processes require high precision and efficiency. Traditional robotic systems are often limited to custom-made dosing units, which increases costs and limits flexibility of use. In this study, a modular end effector design suitable for standard hand-held dosing units was developed [4]. The design allows users to easily integrate existing dosing equipment into the robot and quickly replace it with different dosing units when necessary. This innovative approach reduces costs in DNA isolation and liquid transfer processes while increasing operational flexibility.

Additionally, a simple firmware model design example is given in this study.

[1] Konya Technical University, Konya/Turkey, Orcid: https://orcid.org/0000-0001-5549-8385, byilmaz@ktun.edu.tr
[2] Selcuk University, Konya/Turkey, Orcid: https://orcid.org/0000-0002-5232-8898, vfirat@gmail.com

END EFFECTOR DESIGN

Mechanical Hardware

The robot developed within the scope of the project uses magnetic bead technology. DNA extracted with magnetic bead technology is stuck to the beads in a liquid containing magnetic beads and waits [5]. Later, when exposed to a magnetic field, the DNA is protected by sticking to the beads and the remaining material is removed from the DNA. In order for this procedure to work properly, the end effector must apply a magnetic field at the necessary times during the process and maintain the DNA in the pipette and release the magnetic field at the necessary times to remove the DNA from the pipette [6]. The original design of the end effector that will meet the requirements appropriate to the procedure is shown in Figure 1.
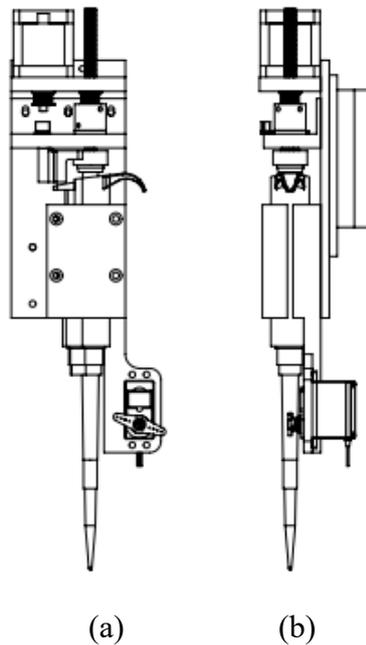

(a)   (b)

Figure 1**:** a) Front view of the end effector b) Side view of the end effector

The end-effector design is shown in Figure 1. The first of these details is that the end-effector used in the design is designed to be compatible with classical pipetting devices used in laboratories. The advantage of this design is that the pipetting device can be easily changed, and its capacity can be adjusted when desired. The second important design detail is that the magnet system of the device can move thanks to the mechanism attached to the device. Thus, the magnetic system can be activated when necessary and apply a magnetic field to the liquid in the pipette. In addition, the magnets in this system can be changed at any time and the field strength can be adjusted. The components of the end effector are represented in Figure 2.
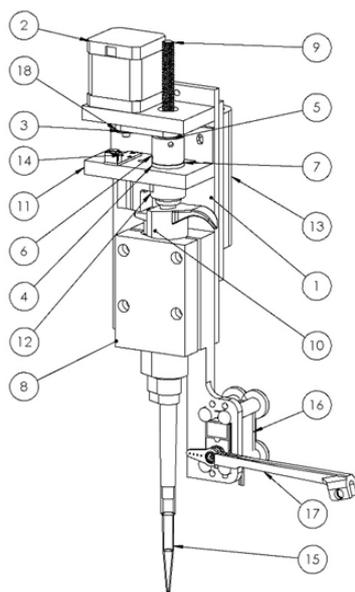
Figure 2: Components Layout of End Effector Design

In Figure 2 the assembled end effector shown with all components uniquely numbered. The list of the components in Figure 2 is listed in Table 1

Table 1: The list of the components

ITEM NO.	PART NUMBER	AMOUNT
1	Dispenser Bed	1
2	Nema 17 stepper motor	1
3	Pulley Gt2_20t-sh5	1
4	Bearing 608	2
5	Pulley Gt2_20t-sh5	1
6	Joining Piece	1
7	Trapezoidal Nut	1
8	Dispenser Cover	1
9	Trapezoidal Screw	1
10	Dispenser	1
11	Lower Support	1
12	Pushing Part	1
13	Z Axis Bed	1
14	bearing shaft	2
15	Optical switch	1
16	Analogue Servo Engine	1
17	Magnetic Tip	1
18	Belt Gt2	1

The magnetic tip used in the study was designed to ensure effective application of paramagnetic bead technology in the DNA isolation process.

This device needs to be powered and controlled with electronic hardware and special software must be installed to the electronic hardware.

Electronic Hardware

The system is driven by a 400-watt switching power supply. Extra regulator circuits are used to adjust the various voltage values in the system. Energy distribution is done by a busbar and terminal system. Energy is distributed to drivers, sensors and control circuits in the system. Control signals in the system are generated by the controller in the panel, and the energy to drive the motors is channeled by the drivers.

According to the power calculations, it was determined that a drive value of 1.6 Newtons of force is required for the Z axis to which the End Effector is connected. In this case NEMA 17 motors were selected. For the end effector. Therefore, a motor with a lower energy requirement was selected.

M542 and TB6600 model stepper motor drivers were used to drive the motors. During the experiments, the "STM Discovery" model control card was used to control the drivers [7]. The devices mentioned are shown in Figure 3.
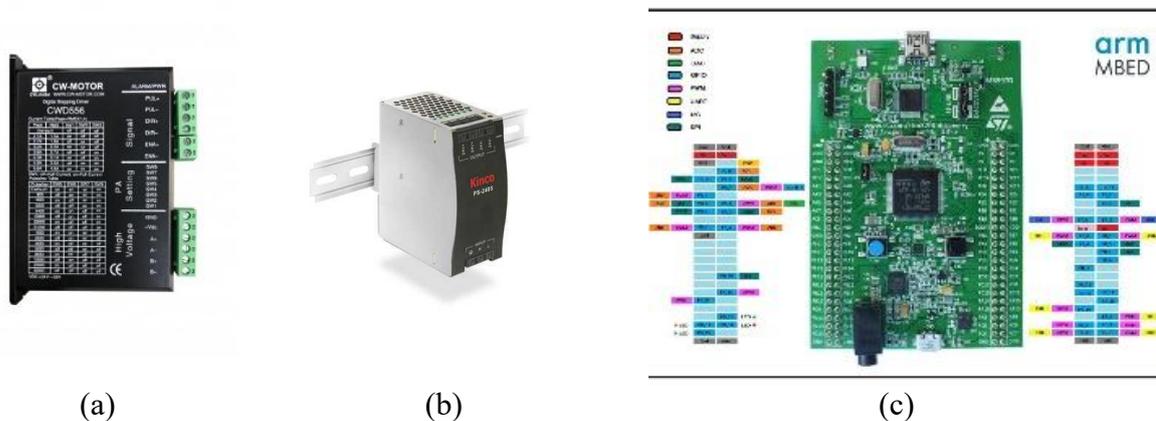

| (a) | (b) | (c) |

Figure 3: Electronic hardware components a) Stepper motor driver, b) Switch mode power supply, c) STM Discovery control card

Software Design

Software design will be according to the user's preference, but a sample software design is also presented in this study. The first part is loaded into the control card, the task of the program is to provide direction and movement information to the motor driver circuits by generating control signals at the electronic level [8]. The program starts working as soon as the system is energized, and first controls the system. Then bring the axes of the robot to the starting point. Finally, it goes into the command waiting state. It performs operations such as taking the pipette, releasing it, applying a magnetic field, moving to the desired position and stopping according to the given command. The flow diagram of the program is as in Figure 4.
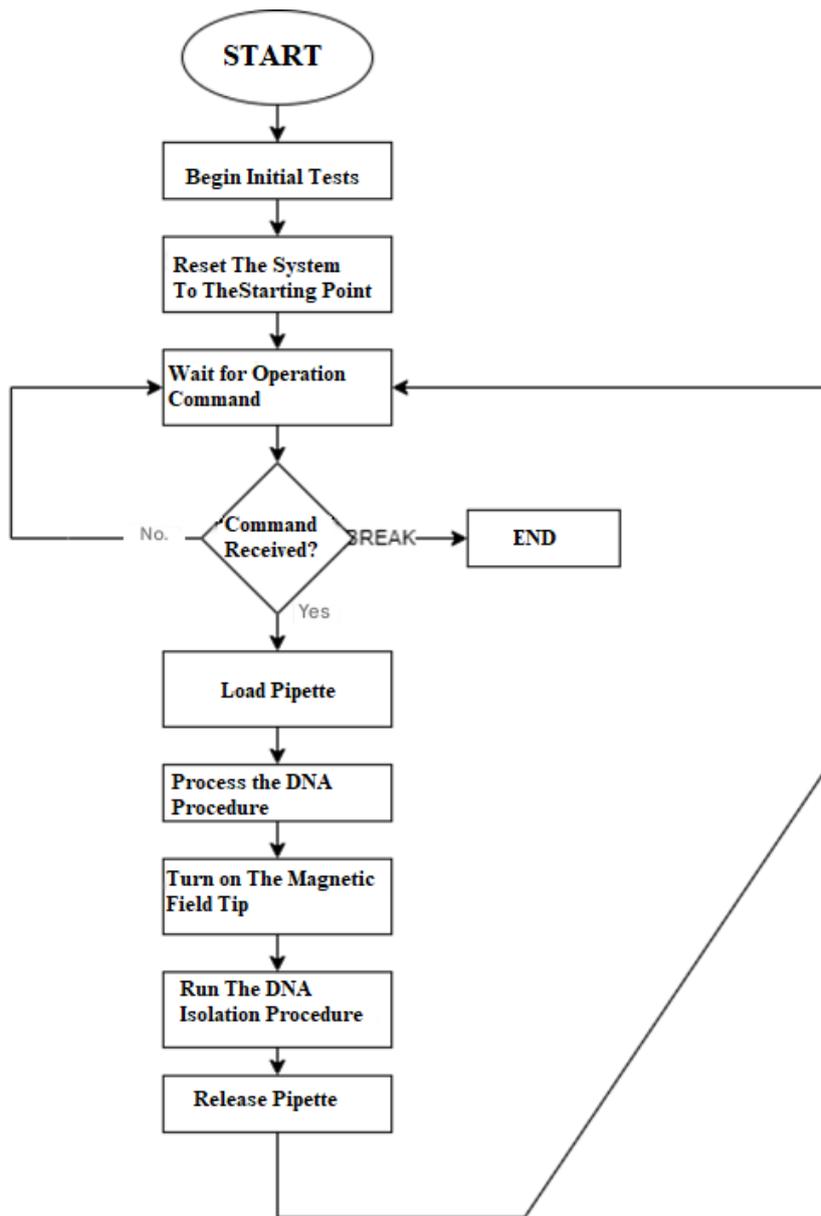
Figure 4: Flowchart of firmware software

Figure 4 represents basic operation firmware, on the other hand, it is highly recommended to develop a interface software according to the robot, which is used with the end effector.
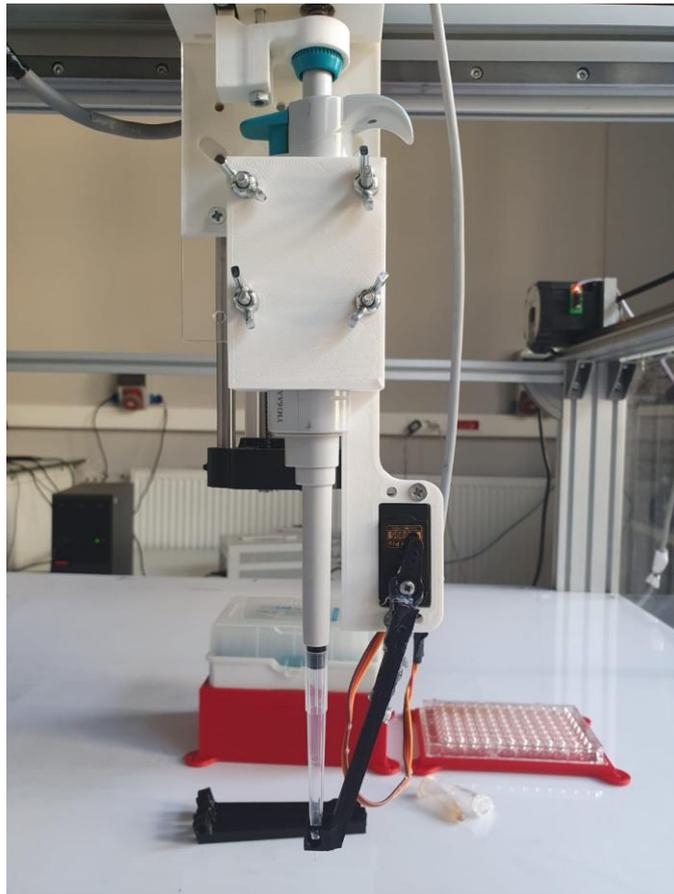
Figure 5: A real-life photograph of the developed end effector system integrated into the Cartesian robotic platform

A real-life photograph of the developed end effector system integrated into the Cartesian robotic platform represented in Figure 5. The image illustrates the modular design and its compatibility with standard pipetting devices, showcasing the system's operational setup during DNA extraction experiments. The photograph highlights key components, including the movable magnetic system and the pipette holder, demonstrating the system's adaptability and functionality in a laboratory environment.

DNA ISOLATION PROCESS

As mentioned before, the silica-based isolation method was taken as the basis for DNA isolation [9]. This isolation method is used by different companies in the market in both microcentrifuge tube-based and magnetic bead-based DNA isolation kits. In tube-based isolation kits, silica membrane is placed in the centrifuge tube as a filter, while in magnetic bead-based kits, the magnetic beads are covered with silica. In tube-based methods, liquid flow and isolation are provided with the help of a centrifuge. In methods where magnetic beads are used, the magnetic beads and therefore the DNA attached to them are released into the liquid medium or removed from the medium by bringing the magnets closer or further away [10].

Buffer solutions prepared and used for DNA isolation with silica-coated magnetic beads Table 2

Table 2: Buffer contents used in DNA isolation.

Buffer Name	Buffer Content
Buffer-1 (cell lysis buffer)	50 mM Tris-HCl, pH 7.5 50mM EDTA 1% Tween 20 1% Triton X-100 %1 SDS
Buffer-2 (neutralization buffer)	3.0 M potassium acetate (pH 4.8) 50% acetic acid
Buffer-3 (binding buffer)	.0 M guanidium thiocyanide 95% ethanol
Buffer-4 (wash buffer)	20 mM Tris-HCl (pH 7.5) 400mM NaCl 70% ethanol

In the study a DNA isolation procedure is also defined for a 96-well microplates and standard tubes (in user-definable sizes).

96-Well Plate Isolation Process

In order to ensure correct addressing during the isolation process, the base plate addresses are defined as shown in Table 3.

Table 3: 96-Well Plate Address Definitions
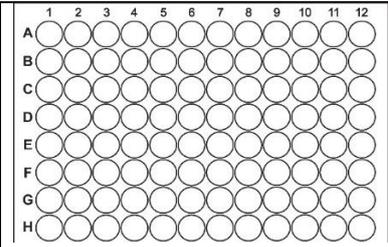
Example of a standard 96-well plate (left) and addressing (bottom).

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12

At the same time, for the use of some high-volume buffers, places for 1.5- or 2.0-ml volume microcentrifuge tubes were designated within the robot work area. Table 4 Represents the steps of the 96-Well Plate Isolation Process

Table 4: The steps of the 96-Well Plate Isolation Process

Step	Address	Process	Magnet	Amount	Repeat
		Tubes / Tube Areas / Work Areas:			
		- Tube-1 >>> supernatant, bacteria			
		- Tube-2 >>> buffer-3 + magnetic beads			
		- Tube-3 >>> buffer-4			
		- Tube-4 >>> sterile H_2O			
		- Trash >>> throw straw			
		- Pipette >>> load pipette			
1	Tube-4	Take it	-	20 ul	-
2	1A6	Leave	-	20 ul	-
3	Tube-1	Take it	-	50 ul	-
4	1A1	Leave	-	40 ul	-
5	Rubbish	Throw a straw	-	-	-
6	Pipette	Take a pipette	-	-	-
7	Tube-2	Take-Leave	-	70 ul	10
8	Tube-2	Take it	-	70 ul	-
9	1A1	Leave	-	60 ul	-
10	Rubbish	Throw a straw	-	-	-
11	Pipette	Take a pipette	-	-	-
12	Tube-3	Take it	-	100 ul	-
13	1A2	Leave	-	100 ul	-
14	Tube-3	Take it	-	100 ul	-
15	1A3	Leave	-	100 ul	-
16	1A1	Take-Leave	-	100 ul	10
17	1A1	Take it	Open	100 ul	-
18	1A4	Wait	Open	-	2 minutes
19	1A4	Leave	Open	100 ul	-
20	1A2	Take it	Open	100 ul	-
21	1A2	Drop-and-Take	-	100 ul	10
22	1A2	Take it	Open	100 ul	-
23	1A2	Wait	Open	-	2 minutes
24	1A2	Leave	Open	100 ul	-
25	1A3	Take it	Open	100 ul	-
26	1A3	Drop-and-Take	-	100 ul	10
27	1A3	Take it	Open	100 ul	-
28	1A5	Leave	Open	100 ul	-
29	1A6	Take it	Open	20 ul	-
30	1A6	Drop-and-Take	-	20 ul	10
31	1A6	Take it	Open	20 ul	-
32	1A6	Wait	Open	20 ul	2 minutes
33	1A7	Leave	Open	20 ul	-

RESULTS

In this study, the laboratory strain Escherichia coli DH5α was used. As a standard, this strain was cultured from -80 °C stocks onto LB solid medium and incubated overnight at 37 °C. Subsequently, colonies were transferred to liquid LB medium and grown for approximately 16

hours at 37 °C with 250 rpm shaking. After growth, approximately 1 ml of culture was subjected to DNA isolation.

Using the basic tube workflow, a yield of 2,450 ng of genomic DNA was obtained from a total volume of 50 µl per 1 ml of sample. This measurement, performed using fluorescent dyes sensitive to double-stranded DNA, directly reflects the amount of pure, double-stranded genomic DNA. According to previous studies, genomic DNA yields from E. coli strains under similar conditions typically range between 2,000–3,000 ng per ml of culture. The value obtained in this study falls within this range, indicating that the DNA isolation and quantification process was both effective and reliable [11].

The integrity and purity of the isolated DNA were further confirmed using agarose gel electrophoresis analysis, as shown in Figure 6. The gel includes a molecular weight marker (DNA ladder) in the first lane on the left, which demonstrates that the extracted DNA was intact and of high quality.

This result was achieved using a Cartesian robotic system with the end-effector described in this study.
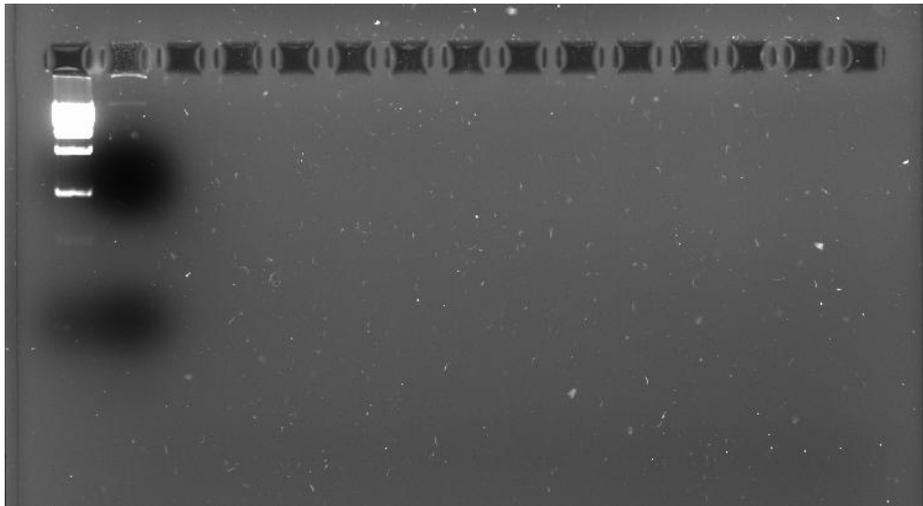

Figure 6: Agarose gel electrophoresis analysis.

CONCLUSION

The study has brought a new approach to the studies on the subject and has been produced at minimum cost using existing resources. It has been demonstrated that such a device can be produced within the scope of the project.

In this study, a reliable and efficient workflow for genomic DNA isolation from *Escherichia coli* DH5α was demonstrated, yielding 2,450 ng of pure, double-stranded DNA per ml of culture. This result falls within the expected range reported in the literature (2,000–3,000 ng/ml), affirming the effectiveness of the methodology employed. Furthermore, the integrity and quality of the extracted DNA were validated using agarose gel electrophoresis, confirming its suitability for downstream applications. The use of a Cartesian robotic system equipped with the end-effector developed in this study played a critical role in achieving reproducible and

high-quality results, showcasing the potential of automated systems in molecular biology workflows. Future studies could expand upon this work by exploring additional robotic optimizations or applying this workflow to other bacterial strains.

Program development processes are ongoing, and it is aimed to provide image processing capabilities to the system in the future.

ACKNOWLEDGMENT

REFERENCES

[1] K. Lee and A. Tripathi, "Parallel DNA Extraction From Whole Blood for Rapid Sample Generation in Genetic Epidemiological Studies," Apr. 29, 2020, Frontiers Media. doi: 10.3389/fgene.2020.00374.

[2] M. N. A. Uda, N. A. Parmin, A. B. Jambek, U. Hashim, M. N. A. Uda, and S. N. A. Shaharuddin, "Evaluation and Optimization of Genomic DNA Extraction from Food Sample for Microfluidic Purpose," Feb. 01, 2020, IOP Publishing. doi: 10.1088/1757-899x/743/1/012031.

[3] F. RB, "Hybrid Integrated Microfluidic Device for Sample Preparation and qPCR on an EWD Platform," Jun. 21, 2019, Biomedical Research Network. doi: 10.26717/bjstr.2019.19.003247.

[4] T. M. Tran, S. Kim, C. Modavi, and A. R. Abate, "Robotic automation of droplet microfluidics," Jan. 01, 2022, American Institute of Physics. doi: 10.1063/5.0064265.

[5] C.-Y. Park, Y.-H. Park, Y.-S. Kim, H.-J. Song, and J.-D. Kim, "Permanent magnet actuation for magnetic bead-based DNA extraction," Oct. 29, 2018, BioMed Central. doi: 10.1186/s12938-018-0572-7.

[6] C. M. Hale and J. Darabi, "Magnetophoretic-based microfluidic device for DNA isolation," Jul. 01, 2014, American Institute of Physics. doi: 10.1063/1.4893772.

[7] A. W. Wardhana and D. T. Nugroho, "Stepper motor control with DRV 8825 driver based on square wave signal from AVR microcontroller timer," Jan. 01, 2019, American Institute of Physics. doi: 10.1063/1.5097484.

[8] N. Q. Le and J. W. Jeon, "An open-loop stepper motor driver based on FPGA," Jan. 01, 2007. doi: 10.1109/iccas.2007.4406542.

[9] J. I. Taylor, C. D. Hurst, M. J. Davies, N. Sachsinger, and I. J. Bruce, "Application of magnetite and silica-magnetite composites to the isolation of genomic DNA." Accessed: Nov. 18, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0021967300001072

[10] S. Berensmeier, "Magnetic particles for the separation and purification of nucleic acids," Applied Microbiology and Biotechnology, vol. 73, no. 3. Springer Science+Business Media, p. 495, Oct. 24, 2006. doi: 10.1007/s00253-006-0675-0.

[11] Mesapogu, S., Jillepalli, C.M., Arora, D.K. (2013). Microbial DNA Extraction, Purification, and Quantitation. In: Arora, D., Das, S., Sukumar, M. (eds) Analyzing Microbes. Springer Protocols Handbooks. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34410-7_1.

RESUME

Asist. Prof. Dr. Burak YILMAZ

Burak Yılmaz received his bachelor's degree from Fırat University, Department of Electrical and Electronics Engineering in 2001. Yılmaz completed his master's degree in computer engineering at Selçuk University in 2007 and received his doctorate degree from Selçuk University, Department of Electrical and Electronics Engineering in 2017.

His research interests include biomedical signal processing, image processing, biomedical image analysis, pattern recognition and artificial intelligence.

Lecturer Dr. Veysel FIRAT

Veysel Fırat received his BSc degree from Selcuk University, Department of Mechanical Engineering in 1999 and started working at Akören Ali Rıza Ercan Vocational School in the same year. Fırat completed his MSc degree in Selcuk University, Department of Mechanical Engineering in 2005 and received his PhD degree from Selcuk University, Department of Mechanical Engineering in 2016. I am currently working as an academician at Selcuk University. His research interests include machine design, wear and finite element methods.

- 292 -

COST EFFICIENCY AND SCHEDULE PREDICTABILITY IN CONSTRUCTION PROJECTS

CHAPTER 20:
Implementation of Earned Value Analysis for Constructıon Cost and Duration Control
  Ayşe Nur ŞENGÜL, Nezir DOĞAN, Önder Halis BETTEMİR

CHAPTER 21:
Assessment of Schedule Variances by Monte Carlo Simulation
  Merve KAYA, Önder Halis BETTEMİR, Mehmet Fatih ACAR

IMPLEMENTATION OF EARNED VALUE ANALYSIS FOR CONSTRUCTION COST AND DURATION CONTROL

Ayşe Nur ŞENGÜL[1], Nezir DOĞAN[2], Önder Halis BETTEMİR[3]

INTRODUCTION

Earned Value Analysis (EVA) is a method that allows to evaluate the cost and time performance of a project at an early stage and gives managers the opportunity to take proactive corrective measures [1]. This analysis method integrates time and cost performance within the project scope and helps the project manager to recognize the current performance indices and make future forecasts [2]. In the cost management process during the conbstruction project management, direct, indirect and overhead costs are aggregated to create a cost baseline. This cost baseline is a cumulative budget used to measure and monitor current and future project cost performance and is usually graphically represented in the form of an S-curve. However, traditional approaches can be misleading as they do not take into account the value of work completed, and looking only at expenditures may not accurately reflect project progress. This is where EVA provides a more comprehensive project monitoring.

Unplanned expenditures can lead to cost overruns, while inflation in commodity prices can affect the cost of inputs. Construction projects carry risks and these risks can delay project completion or increase costs. Therefore, systematic monitoring and control of projects is necessary. BEA allows assessing the cost performance of projects through the calculation of performance variances and indices [3].

LITERATURE REVİEW

EVA uses three main types of project information: Planned Value (PV), Actual Costs (AC) and Earned Value (EV). PV is the budgeted cost of the work planned to be done, i.e. the Budgeted Cost of Work Planned (BCWS). AC is the total cost spent on the completed work, called the Actual Cost of Work (ACWP). EV is the percentage of the budgeted cost of work completed at a given time and is referred to as the Budgeted Cost of Work Performed (BCWP). EV is calculated by multiplying the budget of an activity by the percentage of progress of that activity as given in Eq. 1.

$$EV = \% \text{ completed} \times \text{budget} \tag{1}$$

[1] İnönü University, Malatya/Turkey, Orcid: https://orcid.org/0009-0008-4968-7153, aysenursengul1307@gmail.com
[2] İnönü University, Malatya/Turkey, Orcid: https://orcid.org/0009-0005-9685-0740, nezirdogan22@gmail.com
[3] İnönü University, Malatya/Turkey, Orcid: https://orcid.org/0000-0002-5692-7708, onder.bettemir@inonu.edu.tr

If a project has poor cost performance, the CPI is less than 1. If the SPI is greater than 1, more work has been done than planned. An SPI value greater than 1 indicates that more work has been done, while an SPI less than 1 indicates that less work has been done within the analyzed time interval. However, SPI does not indicate progress on the critical path, so the team can achieve high SPI by making progress on non-critical work, but the project may still be delayed. The whole project process should be reviewed to understand the overall status of the program. Another evaluation tool, the TCPI, shows the performance required to achieve financial targets in the remaining works. TCPI can be calculated based on the remaining budget or completion estimate and gives project managers an idea of whether this performance can be achieved. Generally, TCPI calculation which bases on the completion estimate is more meaningful [4].

Research shows that the cumulative CPI usually stabilizes when 20% of a project is completed. That is, the CPI has not changed by more than 10% after this stage. At 30% completion of the project, the CPI is also expected to change by no more than 10%. For example, if the CPI at 30% completion is 0.80, the final CPI is expected to be 0.88 at best, indicating that the budget will be exceeded by at least 13.6% [4].

Traditional cost monitoring only compares budget to actual expenditures and does not predict future performance. Due to these limitations, it is stated that EVA is an effective tool for monitoring project cash flows by relating them to time and cost [3]. Project cost analysis and forecasting is important for evaluating cost performance. According to PMI PMBOK®, some key parameters related to EVA are as follows [3].

Cost estimates are made by calculating performance metrics to predict future performance trends. Reporting data is generated by dividing the total budget of the project by the performance index (CPI) for each quarter. The Estimate at Completion (EACc) curve shows how resource utilization efficiency affects the final cost. The remaining cost is determined by the Estimated Completion (ETC) factor. Resource utilization efficiency (CPI) has a direct impact on the Variance at Completion (VAC). High CPI values tend to result in higher VAC and under-budget completion, while low CPI values reduce VAC. A positive VAC trend indicates that the project will be completed under budget [3].

Although decisions in the EVA method are usually based on SPI and CPI values, a numerical and linguistic control mechanism is missing, which makes it difficult to fully understand the current situation. A fuzzy control scheme with α-cutting to control earned value performance indices in linguistic terms is presented in [2]. The manager can use classical "in control" or "out of control" judgments as well as intermediate judgments. This includes intermediate decisions such as "more in control" or "more out of control".

In the case study by Noori et al. the CPI on the state day is equal to 0.832, assuming a = 0.5, and the value required to make a decision about the process condition is 0.82. The SPI in the current state is equal to 0.752 and the required value is 0.02 and the program performance status is "very poor" [2].

The applicability and contribution to success of the Earned Value Analysis (EVA) in the construction industry is examined in [5]. The study examines a construction company that integrated EVA into its processes. At the first construction site, EVA was used as the main tool

of construction control and successful results were achieved with full support. According to the research, for EVA to be successful, it is important that the project objectives are clear and the scope is clear. While simple and straightforward projects work better with EVA, it is difficult to apply it in projects with unfinished products or unclear objectives. Clear definition of the project scope, EVA experience of the teams and organizational support increase the effectiveness of EVA [5].

For effective implementation of EVA, the design and operational aspects should be properly addressed. Using agency and organizational justice theories, the relationships between project participants and the contribution of EVA to success were examined in [6]. In a construction project, EVA was effectively applied to achieve time and cost targets, and the project was successfully completed with strong communication and trust. Organizational justice strengthened collaboration. However, in the clinical research project, lack of communication and low perception of fairness weakened trust, decreased collaboration and led to failure. It is stated that attention should be paid to design and operational aspects for the successful implementation of EVA and it is emphasized that organizational justice contributes to success by increasing trust [6].

To apply EVA, confidence in the cost monitoring system is required. Most systems only compare planned and actual cost, but the Earned Value method also requires the concept of "earned value". Much software does not support this element, making EVA difficult to implement. In addition, system change costs and management issues also hinder adoption. Another barrier to EVA adoption is the overuse and complexity of formulas. Project management often focuses on planning, which leads to insufficient attention to execution. An effective EVA system requires up-to-date and detailed reporting, but many companies have difficulty providing this information. Although project managers are responsible for the progress of projects, their limited access to cost data makes EVA difficult. For widespread implementation of EVA, a time-phased baseline schedule based on schedule and cost estimates should be developed. It should be known how much should have been spent at any given moment in the project. There should also be a detailed project plan and a cost management system that allows the manager to continuously monitor expenditures [7].

Earned Value Management Systems (EVMS) set out 32 criteria for management systems that fall into five main categories. The categories are listed as organization, planning and budgeting, accounting, analysis and revisions. Over time, however, these criteria have become a burden for management and have been found to be of little use. Compliance activities for EVMS increase costs [8]. This study examined the costs of Department of Defense (DOD) regulations, especially EVMS, in the industry. The results show that the cost of EVMS is the third of the ten highest cost drivers and accounts for about 0.9% of the value added cost. The majority of EVMS costs come from engineering and finance. A DOD report stated that two-thirds of these costs are unnecessary. Removing unnecessary EVMS requirements could reduce costs, but not eliminate them entirely. It is estimated that reforms could save the industry one billion dollars annually. The output of the EVMS process, the earned value report, should be reliable for decision-making. The predictive value and timeliness of the report are important. Studies show that managers find these reports useful [8].

Studies on the improvement and application of EVA in coal mine projects in China are reviewed in [9]. Traditional EVA is a common method to evaluate the cost and time performance of projects; however, it has some limitations in schedule and cost measurement. To overcome these limitations, the weighted earned value analysis (WEVA) has been proposed [9].

The traditional EVA is limited by its inability to distinguish between critical and non-critical activities and its statistical errors. WEVA overcomes these shortcomings and provides more accurate results in projects. It is stated that WEVA is a more effective method in project management processes, especially in large projects. In this method, a weighting coefficient (ki) is used to measure the impact of each activity on the project. Critical activities have a higher weight coefficient due to their importance on the project [9]. Weight coefficient formula is given in Eq. 2.

$$k_i = e(TF_i)_{i=1}^n e(TF_i) \tag{2}$$

Here TF (Total Float) refers to the total time difference of each activity. Critical activities have a TF value of zero and these activities are vital for the timely completion of the project. The traditional EVA and WEVA which is proposed in [9] are compared on a coal mine project in China. At the end of the fourth month, the schedule performance index (SPI) calculated by the traditional EVA was 0.99, while the SPI calculated by the WEVA was 1.04. This shows that WEVA is able to predict project delays more accurately. WEVA provides a more reliable approach to project management by estimating project deviations more accurately.

A control mechanism was developed in [10] by integrating EVA into uncertain production environments. The effect of uncertain factors on activity durations was modeled with fuzzy logic and triangular fuzzy numbers. For each activity, optimistic, probable and pessimistic values for the uncertainty in the duration of activities were determined. In addition, it is recognized that costs are directly related to activity durations and it is proposed to calculate costs as optimistic, probable and pessimistic costs for different values of these durations.

A CSI close to 1 indicates that the project is in line with the objectives; greater than 1 indicates that the project is exceeding expectations in terms of both time and cost, and less than 1 indicates poor performance. CSI provides early warnings to managers by monitoring the cost and time balance simultaneously, which makes it possible to take the necessary corrective measures in a timely manner. Thus, the project can be intervened before budget overruns or time loss [10].

The control mechanism developed for a multi-period and multi-product production planning problem is tested in [10]. The results show that the control mechanism effectively monitors cost and time performance in uncertain production processes. The mechanism detects delays and cost overruns in advance and provides the opportunity to intervene. Deviations in cost and schedule were identified by CPI and SPI values, giving managers the opportunity to make necessary interventions at critical moments. Production conditions were categorized as "super star", "good", "poor" and "worst".

For project management and control under uncertainty, [11] aims to analyze project deviations using Monte Carlo simulation. This method uses anomaly detection algorithms to determine whether deviations are due to natural variation or unexpected events. It helps the project

manager to understand the causes of deviations. When analyzing the observed deviations, Bayes' theorem is used to classify them. This classification is based on probability calculations to determine which class the observations belong to, and one of the most common methods as a Bayesian classifier is criminant analysis. Given an observation x, these algorithms calculate the probabilities $Pr(Y=k|X=x)$ for each possible class k, assigning the observation to the class with the higher probability.

Simulation results can also be treated as a regression problem. For example, one can estimate the expected final cost and duration of a project once it reaches a certain level of earned value. The regression models used here allow the predictions to be more flexible. This is where Generalized Additive Model (GAM) comes into play. GAM estimates the final cost and duration of the project by modeling the non-linear relationships between the independent variables in the project and the outcomes.

This allows project managers to see more clearly whether deviations are within the planned stochastic variation or are due to the structural problems. Monte Carlo simulations and anomaly detection algorithms allow for close monitoring of project progress. Simulations and algorithms provide more detailed information on when and at what cost the project will be completed [11].

The study conducted by [1], examines the Earned Value Analysis (EVA) method in construction projects. The study also defines the main parameters of EVA and emphasizes its importance in cost management. The developed software was created using Visual Studio 2008, SQL Server 2005 and .Net (C#) and compared with Microsoft Project 2007 and Primavera P6. In the case study, the planned duration was 159 days and the actual duration was 166 days. The cost variance was -76,800 Indian Rupees, the cost performance index was 0.93 and the time variance was 8 days. The software provides 99.5% accuracy with other programs and also a new parameter, time variance, helps to better evaluate the time performance of projects.

In the study in [12] applied EVA to the construction of Monica Park in Rio de Janeiro. The construction cost is 5 million US dollars and the construction is planned to be completed in 10 months. The use of EVM has brought significant benefits despite some challenges. In the initial phase, some people resisted because they could not see the benefits of EVM. However, with EVM, cost and time management were integrated, early warning systems were developed and problems were quickly identified. For successful EVA implementation, there needs to be high-level commitment, the project team needs to be trained in EVA, and the project scope needs to be well defined and detailed. The schedule and budget should be organized in a Work Breakdown Structure (WBS), responsibilities should be clarified, relationships between activities and participants should be clearly defined, data collection procedures and cost/schedule control system should be established, and EVA reports should be well planned and distributed.

The study in [13], presented a more transparent, flexible and practical approach to project management by simplifying the complex calculations and terms used in traditional EVA methods. Formulas are developed that clearly define the impact of resource utilization on the project and the cost linkage, and new notations are proposed for performance measurements based on the differences in cost and time saved. In addition, a linear approach is developed that

allows past performance to be used in future forecasts, and with these innovations, project managers can perform both past analysis and future performance forecasts more effectively [13]. Similarly, in [14], it was found that fixed control thresholds are not always suitable for the dynamic nature of projects and are insufficient to provide early warnings, and to overcome this problem, the "The To-Complete Schedule Compression Ratio" (TCSCR) scale was developed. In this way, the weaknesses of the classical EVM are overcomed by providing time-based measurements with the earned schedule method (ESM) and dynamic performance measures focusing on project duration are proposed. With this approach, deteriorations in project performance and potential delays are recognized at earlier stages thanks to the dynamic control thresholds [14].

High fluctuations of the CPI and SPI especially in the beginning of the construction might mislead the project managers [15]. Moreover, at the final phase of the construction CPI and SPI indexes saturates and they do not response to any adverse situation which can also be dangerous [15]. The study proposes some modifications to eliminate the aforementioned deficiencies but the overall improvement cannot overcome the whole deficiencies.

METHOD

EVA uses three main parameters to measure cost and time performances in projects:

1. BCWS (Budgeted Cost for Work Scheduled): The budgeted cost of the work to be performed at a given stage according to the project plan.

2. ACWP (Actual Cost for Work Performed): Refers to the actual costs of the work performed.

3. BCWP (Budgeted Cost for Work Completed): This is the budgeted cost of work completed during the project and is also referred to as earned value.

Two performance indices are used to assess project performance:

CPI (Cost Performance Index): It is the ratio of BCWP to ACWP and is expressed as presented in Equation 3.

$$CPI = BCWP/ACWPC \tag{3}$$

If CPI<1, the cost performance of the project is worse than planned, if CPI>1, it is better than planned, and if CPI=1, the cost performance is exactly as planned.

SPI (Schedule Performance Index): It is the ratio of BCWP to BCWS. It is calculated by the expression presented in Eq. 4.

$$SPI = BCWP/BCWS \tag{4}$$

If SPI<1, the project is delayed, if SPI>1, the project is on going ahead of schedule, and if SPI=1, the project is on time [9].

Estimate to Completion (ETC): This provides an estimate of the remaining cost required to complete the project.

Estimate at Completion (EAC): This is estimation for the total expected cost when the project scope is completed.

Budget at Completion (BAC): The total budgeted project cost to complete the whole construction which includes overhead and contingencies.

Contingencies allocate additional budget and time to the project as a measure to cover unplanned events that occur in projects. It is important to consider the remaining contingencies when calculating ETC and EAC. This should include adjustments for 'known unknowns' [4].

Equation 5 is used for the most pessimistic EAC estimate where both cost and schedule impact are taken into account and where project performance is generally poor [4]. EAC is computed as given in Eq. 5.

$$\text{EAC} = \text{BAC} / (\text{CPI x SPI}) \tag{5}$$

Estimate of Completion Cost (EACc): Estimates the cost to complete the project based on current cost performance [3]. EACc is computed as given in Eq. 6.

$$\text{EACc} = \text{BAC} \div \text{CPI}. \tag{6}$$

Estimate of Completion (ETC): Indicates the estimated cost of the remaining work [3] and it is computed as given in Eq. 7.

$$\text{ETC} = (\text{BAC-KD}) \div \text{CPI}. \tag{7}$$

CASE STUDY

The EVA is implemented on a simple construction project whose construction schedule is given in Table 1.

Table 1: Hypothetic activity durations and cost values for the case study problem

Activity Name	Planned Duration (Day)	Cost (TL/Day)	Total Cost (TL)
Excavation	5	7500	37500
Formwork of Foundation-Reinforcement	4	12000	48000
Concreting of Foundation	3	10000	30000
Formwork & reinforcement (Entrance Floor)	4	12000	48000
Concreting (Entrance Floor)	3	10000	30000
Masonry	6	8000	48000
MEP	5	6000	30000
Plastering	4	5000	20000
Door & Window Frames	4	8000	32000
Painting	2	4000	8000

Method and the early start and finish times of the activities are also given in Table 1. Hypothetical actual activity durations and costs are represented in Table 2. Given figures are hypothetic and they are provided to execute EVA computations day by day.

Table 2: Hypothetic realized activity duration and cost values

Activity Name	Actual Duration (Day)	Actual Cost (TL/Day)	Actual Total Cost (TL)
Excavation	6	7530	45180
Formwork of Foundation-Reinforcement	4	12000	48000
Concreting of Foundation	4	10050	40200
Formwork & reinforcement (Entrance Floor)	4	12000	48000
Concreting (Entrance Floor)	3	10000	30000
Masonry	7	8040	56280
MEP	4	5970	23880
Plastering	5	5025	25125
Door & Window Frames	3	7960	23880
Painting	2	4000	8000

The schedule performance index calculated by Equation 4 is given in Figure 1. As stated in the literature review the trend of the index is unclear in the beginning of the construction project and it presents sharp rises and falls.
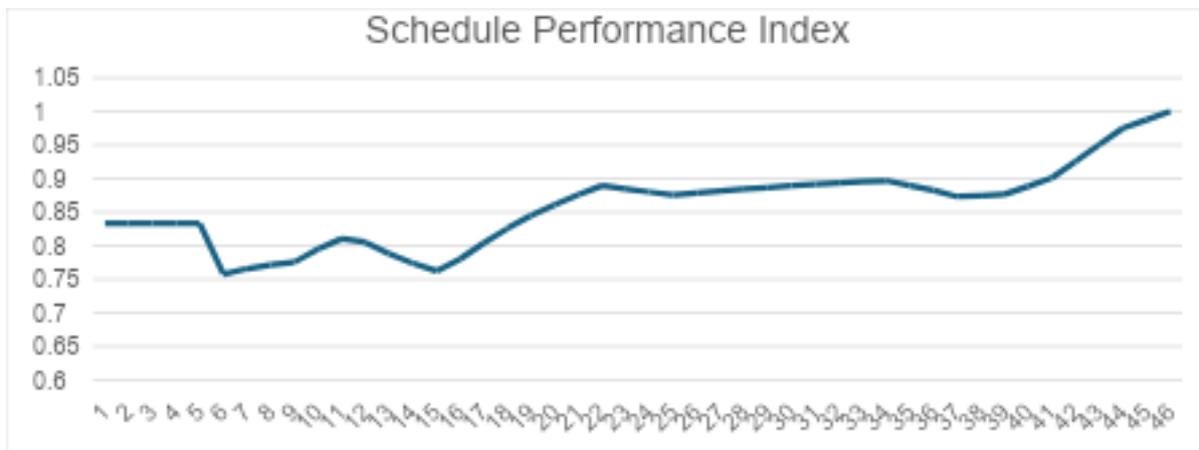

Figure 1: Computed schedule performance index throughout the construction period

The cost performance index calculated by Equation 3 is given in Figure 2. The fluctuations of the CPI index are significantly high at the initial phase.

Figure 2: Computed cost performance index throughout the construction period

The completion estimate EAC calculated by Equation 5 is given in Figure 3. Figure 3 is reversely correlated with Figure 2. As the CPI increases EAC decreases.
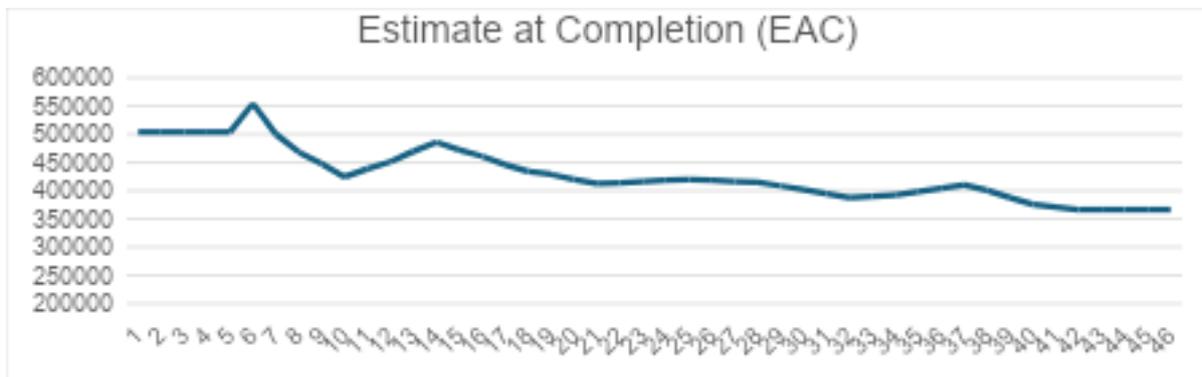

Figure 3: EAC index throughout the construction project.

The Completion Cost Estimate EACc, calculated by Equation 6, which estimates the cost to complete the project based on cost performance, is given in Figure 4.
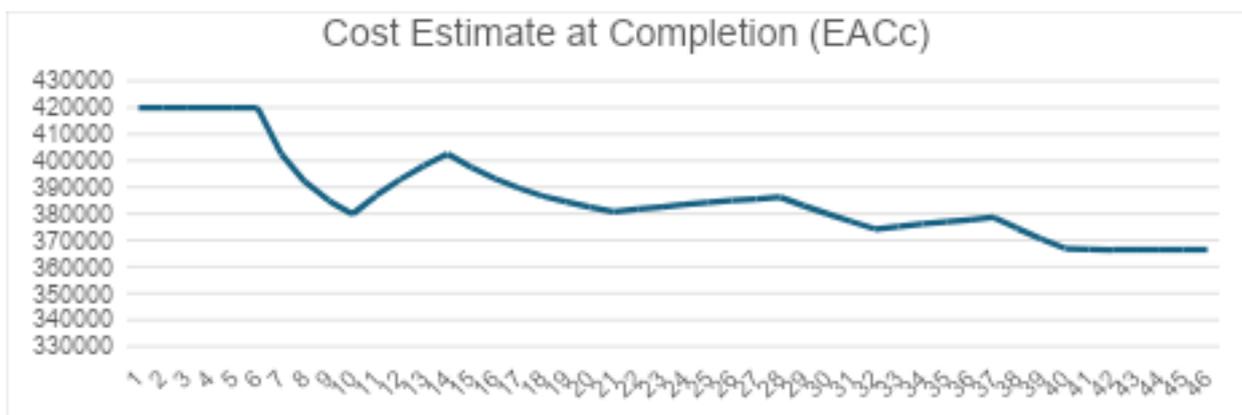

Figure 4: EACc index throughout the construction project.

The ETC calculated by Equation 7 and showing the estimated cost of the remaining work is given in Figure 5.
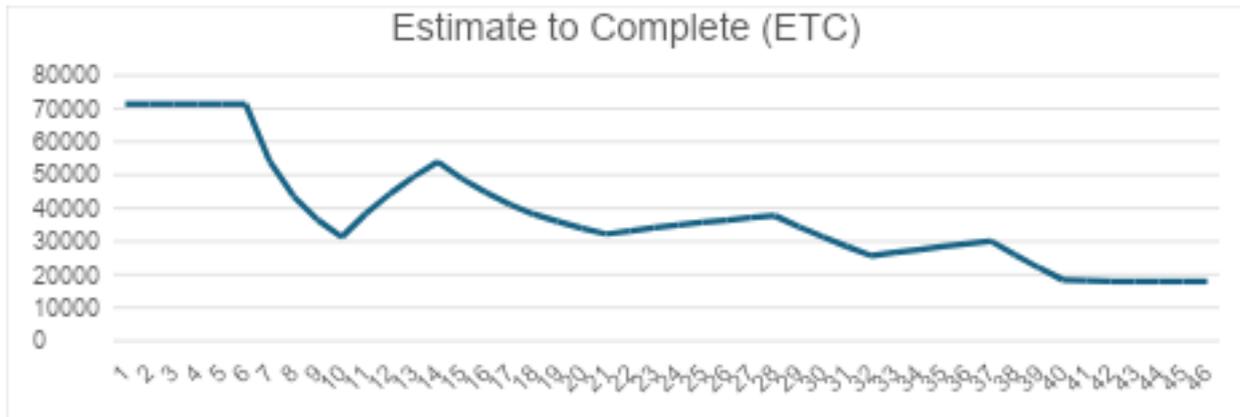
Figure 5: ETC index throughout the construction project

CONCLUSION

In this study, earned value analysis was performed on a 46-day construction project and analyzes were made on the duration and cost of the construction. As a result of the analysis, the estimated completion time and cost of the construction were obtained. It can be seen from the graphs presented in Figures 3, 4 and 5 that the estimated values fluctuate significantly depending on the data flow.

As stated in the literature review, time and cost performance indices are also subject to significant fluctuations at the beginning of the construction and the magnitude of fluctuation decreases as the construction progresses. It can be seen from Figures 1 and 2 that reaching full saturation requires the completion of 50% of the construction. As the time and cost performance indices reach saturation, the fluctuations of the forecast values decrease. In particular, it is seen that the fluctuations of the forecast values presented in Figures 3 and 4 decrease after Day 23, which corresponds to half of the construction period.

As an important tool of project management, the ability of EVA to make consistent and reliable forecasts, especially in the early stages of construction, is important for taking early precautions and avoiding financial difficulties. For this reason, it is necessary to develop models that can make time and cost estimates with lower fluctuations to overcome the shortcomings of this study.

ACKNOWLEDGMENT

REFERENCES

[1] Bhosekar, S. K., & Vyas, G. (2012). Cost controlling using earned value analysis in construction industries. International Journal of Engineering and Innovative Technology (IJEIT), 1(4), 324-332

[2] Noori, S., Bagherpour, M., & Zareei, A. (2008). Applying fuzzy control chart in earned value analysis: a new application. World Applied Sciences Journal, 3(4), 684-690.

[3] Waris, M., Khamidi, M. F., & Idrus, A. (2012). The cost monitoring of construction projects through earned value analysis. Journal of construction engineering and project management, 2(4), 42-45.

[4] Joseph, A., & Lukas, P. E. (2008). Earned value analysis—Why it doesn't work. Proc., AACE Int. Transactions, EVM, 1.

[5] Vargas, R. V. (2003). Earned value analysis in the control of projects: Success or failure. AACE International Transactions, 21(4), 211-214

[6] Bryde, D., Unterhitzenberger, C., & Joby, R. (2018). Conditions of success for earned value analysis in projects. International Journal of Project Management, 36(3), 474-484

[7] Gershon, M. (2013). Using earned value analysis to manage projects. Journal of Applied Business and Economics, 15(1), 11-14.

[8] Christensen, D. S. (1998). The costs and benefits of the earned value management process. Journal of Parametrics, 18(2), 1-16.

[9] Zhong, S., & Wang, X. (2011). Improvement and application of earned value analysis in coal project management. Procedia Engineering, 26, 1983-1989.

[10] Bagherpour, M., Zareei, A., Noori, S., & Heydari, M. (2010). Designing a control mechanism using earned value analysis: an application to production environment. The International Journal of Advanced Manufacturing Technology, 49, 419-429.

[11] Acebes, F., Pereda, M., Poza, D., Pajares, J., & Galán, J. M. (2015). Stochastic earned value analysis using Monte Carlo simulation and statistical learning techniques. International Journal of Project Management, 33(7), 1597-1609.

[12] Valle, J. A., & Soares, C. A. P. (2006). The use of earned value analysis (EVA) in the cost management of construction projects. In Proc., Project Management Institute Global Congress, Newtown Square, PA (pp. 1-11).

[13] Cioffi, D. F. (2006). Designing project management: A scientific notation and an improved formalism for earned value calculations. International journal of project management, 24(2), 136-144.

[14] Kim, B. C. (2015). Dynamic control thresholds for consistent earned value analysis and reliable early warning. Journal of Management in Engineering, 31(5), 04014077.

[15] Bettemir Ö. H. (2024) Modification to Remove the Anomaly of the Earned Value Analysis, 28th International Congress on Project Management and Engineering Jaén, 3rd-4th July 2024, 120 - 136

RESUME

Ayşe Nur ŞENGÜL

Ayşe Nur Şengül started her undergraduate education at Inonu University, Faculty of Engineering, Department of Civil Engineering and graduated in 2023. In the same year, she started her graduated program at Inonu University Institute of Sciences, Department of Construction. Her research interests focus on project time management and optimization. She is especially interested in time-cost analysis and optimization processes in construction projects.

Nezir DOĞAN

Nezir Doğan graduated from the Department of Civil Engineering at İnönü University in 2022. After his graduation, he started his master's degree at İnönü University and focused his studies on indoor positioning systems, project management and error analysis. In his research, he aims to develop innovative methods and integrate these methods into practical engineering applications.

Prof. Dr. Önder Halis BETTEMİR

Prof. Dr. Önder Halis Bettemir Graduated from Middle East Technical University, Department of Civil Engineering in 2003. Completed Master of Science studies at the Department of Civil Engineering in Geodesy and Photogrammetry discipline in 2006. Completed his Ph.D. studies at the Department of Construction Management at Middle East Technical University in 2009. Bettemir was assigned as Assistant Prof. to the Department of Civil Engineering of Yuzuncu Yil University in 2010 and studied in this university until 2014. Bettemir continued his academic career in İnönü University and obtained the degree of Assoc. Prof. in 2017. He became professor in 2023. His main research areas are Geographic Information Systems and Optimization of Construction Schedules. Moreover, exact quantity take-off by Building Information Modeling is his current study topic.

ASSESSMENT OF SCHEDULE VARIANCES BY MONTE CARLO SIMULATION

Merve KAYA[1], Önder Halis BETTEMİR[2], Mehmet Fatih ACAR[3]

INTRODUCTION

The construction industry is one of the key dynamics of economic growth and contains significant uncertainty and risk due to its complexity and dependence on a large number of variables. Uncertainties refer to the impact of unforeseen future events on the project, while risks are defined as the potential for these uncertainties to have negative impacts [1]. Uncertainties in construction projects usually arise from factors such as labor, material supply, environmental conditions and economic fluctuations. In addition to these, many other factors such as design errors, equipment failures, incorrect technical calculations, natural events, legal regulations, etc. have an impact on project duration and cost. These uncertainties and risks can cause cost overruns and delays by complicating time and budget management in projects. To illustrate, economic risks such as exchange rate fluctuations or sudden changes in material prices can increase project costs more than that of estimated, while factors such as labor shortages or weather conditions can lead to serious deviations from the planned time [2, 3]. Problems such as extending the project duration and exceeding the budget carry great risks that can lead to disputes between the parties and the failure of the projects. Companies operating in this sector are aware of the need for meticulous preparation and accurate evaluation of business proposals in a highly competitive environment. In this process, the high accuracy of the schedule and cost estimate is vital not only for the successful completion of the work, but also for the protection of the company's reputation and long-term profitability. However, it is not always easy to ensure the accuracy of time and cost estimates. Therefore, the data used in the preparation of time and cost proposals for construction work needs to be detailed, reliable and realistic. In particular, factors such as miscalculating the quantities of construction items or failing to take into account uncertainties in cost analyses can lead to inaccurate estimates of the construction cost, thereby affecting the company's profitability and causing losses. During the preparation of time and cost proposals for construction projects, it is also critical to establish an accurate and reliable construction schedule. Completing the construction process on time and within budget is an important factor affecting customer satisfaction as well as the company's reputation. However, if the duration of the work is erroneously underestimated, increases in project costs and quality problems may become inevitable.

In this context, it is necessary to develop new approaches to improve the accuracy of the data used in the preparation of time and cost proposals for construction work and the creation of

[1] İnönü University, Malatya/Turkey, Orcid: https://orcid.org/0009-0003-3580-1238 , kya.mrw@gmail.com
[2] İnönü University, Malatya/Turkey, Orcid: https://orcid.org/0000-0002-5692-7708, onder.bettemir@inonu.edu.tr
[3] İnönü University, Malatya/Turkey, Orcid: https://orcid.org/0009-0004-4850-6243, mehmetfatihacr@gmail.com

work schedules. In this study, a research was conducted to determine the effect of the variations of the construction duration estimates and to improve these estimates using advanced analytical methods such as Monte Carlo Simulation (MCS). Based on the results of the research, more effective and reliable methods can be developed for preparing work proposals and creating construction schedules. An executable application was developed to calculate the proposed duration according to the risk ratios determined by the planners. This application allows contractors to add a margin of safety and profit on top of the specified amount, aiming to minimize the negative situations that may occur at the end of the work. The scope of the study includes the determination of the uncertainties of the schedule estimates required for the planning of construction works by MCS. In the following sub-headings, brief information is given on the mentioned topics and their relevance to the study is explained.

Planning

Planning the execution phases of construction projects is vital for the successful completion of the project. This planning process must be done meticulously with realistic data. Construction planning provides critical information on which construction items will be executed on which day and how many workers will be employed at each stage of the project. It is also an important tool for determining the project completion time. This process focuses in detail on the requirements and processes of each construction item. During the project planning phase, the scope and objectives of the project are scrutinized. The complexity and scale of the work and the appropriate use of resources are taken into account when determining which construction items need to be completed as well as when they are needed to be completed. During labor and resource planning, the numbers of construction workers including their areas of expertise and required working hours for each construction item are evaluated in detail. In addition, the overall timeline of the project is created to determine the completion time and the total time required to complete each phase of the construction project is calculated. In this process, potential risks and uncertainties are also taken into account. Alternative plans are created against problems that may arise at any stage of the project and the process is ensured to proceed uninterruptedly. Construction scheduling and planning plays an important role in project management and is an important factor in determining the steps required for the successful completion of the project.

Calculation of Exact Bill of Quantities

The detailed and accurate preparation of the BOQ is a critical step in construction projects. This process ensures accurate estimation of amounts of construction materials as well as project cost estimates and provides a firmer basis for the duration of works. It also helps to determine when the materials needed by the teams need to be procured, which improves operational efficiency. This process also helps to minimize project risks by optimizing work processes and avoiding cost disruptions. Accurate preparation of estimates and actual quantities is therefore vital for the successful completion of construction projects.

Monte Carlo Simulation

Monte Carlo Simulation is a statistical method used to understand and predict the behavior of a system under uncertainty. This simulation is used to model different possible states of a system

through random number generation and statistical analysis. In complex and uncertain systems, such as construction projects, Monte Carlo simulation is an important tool because it provides realistic predictions by taking into account the interaction of many variables and their uncertainties. MCS takes into account the parameters of the project that are under uncertainty, selects random values for each parameter and runs simulations many times with the randomly generated numbers. The data obtained as a result of these simulations are statistically analyzed and provide information about the different possible situations and outcomes of the project. For example, when used in construction projects, factors such as time, cost, resource requirements and risks can be modeled with MCS. This method is used to evaluate different possible scenarios of the project and understand the risks. In construction projects, assessing uncertainties in factors such as duration and cost provides project managers and stakeholders with a more accurate picture. This enables more informed decisions to be made, resources to be managed more effectively and actions to be taken to ensure the successful completion of the project. MCS is an important tool in project management, providing realistic forecasts in complex and uncertain systems such as construction projects. This method helps to determine the strategies necessary for the successful completion of the project and to manage risks effectively.

LITERATURE REVIEW

Helton compared Differential Analysis, Monte Carlo Analysis, Response Surface Methodology, and Fourier Amplitude Sensitivity Testing techniques in performance evaluation for radioactive waste disposal. The models used in performance evaluation for radioactive waste disposal and other related operations are often overly complex in problems with a large number of both input and output variables. As a result of the comparison of the methods, it was determined that the Monte Carlo method is the approach with the widest applicability for use in performance evaluations for radioactive waste disposal [4].

Akgöl and Günay (2018) proposed a Monte Carlo-based alternative method for route rationality scales used in measuring the road network rationality of cities [5]. In order to obtain the data, software integrated with Google Maps was developed using Java programming language.

Köse and Yanmaz (2010) presented two empirical and semi-empirical methods for the temporal variation of fresh water scour around bridge abutments. The reliability of abutment foundation depth against scour tendency by empirical method and MCS are predicted [6]. In the study, reliability calculations for nonlinear equations were performed by Monte Carlo Simulation. The results of the empirical and semi-empirical methods were found to be consistent with each other and with the experimental data.

Since horizontal loads cause bending and significant shear stresses in tie beams, the structural behavior of hollow web systems is strongly influenced by the geometrical and mechanical properties of the tie beams and their behavior, which is closely related to the tie beam contribution ratio (r). Therefore, r-values can be considered as an important parameter in the nonlinear calculation of hollow web systems. Doran and Alacalı (2017) calculated the probabilities of collapse of these tie beams by Monte Carlo simulation by expressing the tie beam contribution ratios with the help of the extended bar theory by considering a sufficient

number of hollow wall systems with geometry frequently encountered in practice in order to determine the "r" values [7].

Aydemir and Zorbozan (2012) investigated the effects of uncertainties on the bending moment capacity by considering the statistical distributions of the parameters affecting the column bending moment capacity Mp. For this purpose, the variations of the analytical moment-curvature relationships using material strengths, section dimensions and different concrete behavior models with the experimental results were numerically modeled by MCS. With the help of a computer program, the coefficients of increase in bending moment bearing strength for sample column sections with different reinforcement arrangement, section size, concrete class, reinforcement class and lateral reinforcement mechanical index were calculated and the results obtained were subjected to statistical evaluation [8].

Şengül (2011) modeled the entry and transport of chlorine ions into concrete with Fick's 2nd law and prepared software that takes into account the distributions in structural properties and environmental effects with Monte Carlo analysis [9].

Birgönül and Dikmen (1996) performed cost estimation for a lump sum sample project using MCS in the light of expert opinions and records of previous years. A sample apartment building project was studied. In order to enable the risk analysis to be carried out, the project cost was divided into sub-cost groups and the risk factors covered by each sub-cost group were determined in line with the opinions of experts. At the stage of determining the probability distributions, the probability distributions and intervals that best characterize the behavior of each sub-cost group were determined. The project inputs were loaded and 2000 iterations were performed. The decision maker will be able to prepare an appropriate proposal by determining the level of risk he/she can assume depending on his/her attitude towards risk [10].

Camalak (2016) evaluated the overtopping reliability of Tanyeri Dam using probabilistic methods. Bivariate flood hydrographs and pre-flood reservoir water levels were randomly generated by Monte Carlo Simulation and the probability of the maximum flood wave level exceeding the crest level of the dam was estimated [11].

Çavuş and Yanıkoğlu (2003) performed reliability analysis of complex systems using Monte Carlo Simulation. During the analysis, a 5-element series and non-parallel complex sample system was analyzed. MCS gives results in a shorter time than analytical methods when computers are used, while numerical methods always give the same result for the same system, MCS gives approximate results depending on the random numbers generated, and it is important to choose the appropriate sampling number when using MCS [12].

Öztekin investigated the reliability of the distances between bolts and bolts to the joint plate edges in bolted joints. Loading type, bolt type, and joint plate thicknesses were taken as variables and reliability analysis was performed for all combinations of input parameters by MCS [13].

The duration of excavation task and the unit cost of excavation play an important role in the selection of excavators used in construction [14]. The factors affecting excavator efficiency and excavation speed were stated as excavator bucket volume, bucket fill rate, turning angle, cycle time, environmental and weather conditions, and type of excavated ground. In the study,

software was developed in C++ programming language that estimates the excavation time and the number of trucks required, taking into account the environmental conditions and different ground conditions. Contractors can determine the risks of time and cost estimates by the software which can run on Windows PC [14].

Bettemir (2006) [15] analyzed the sensitivity and error of the differential rectification method using images captured with a CCD camera and a pushbroom scanner. The sensitivity and efficiency of the algorithm are compared with other image rectification methods and the results are analyzed. In addition, software for rectification of satellite images was developed. It is shown that this software is very accurate and therefore can be particularly useful for producing images with very good resolution, especially for mountainous regions.

Bettemir (2020) applied the modified PERT method to determine the probability of criticality of work packages during the execution of construction activities by considering all possible path completion combinations. By calculating the probability that a path will be completed at a certain time and the remaining paths will be completed earlier than the relevant time, the probability of the relevant path being a critical path is calculated. For large networks, the relationship between path completion combinations and statistical intersection operations was derived and a macro code was created to perform intersection calculations. The study was conducted on 4 examples and the results were compared with Monte Carlo simulation. The study concludes that the method is significantly faster than MCS with similar probability estimates [16].

Bettemir (2012) estimated the georeferencing accuracy of pushbroom satellite imagery without using ground control points. The study includes scenario analyses for various sensitivities of star trackers, GPS receiver, timing of image acquisition, internal camera parameters, and digital elevation model (DEM). The probability distribution function of the input parameters was determined taking into account sensor characteristics, environmental influences and assumptions. Differential sensitivity analysis is used to examine the most probable geo-referenced coordinates and sensitivity for a given sensor and imaging geometry sensitivity. The study is based on the BilSAT orbit and SPOT 5 and Landsat optical system characteristics. A ShuttleRadar Topography Mission (SRTM) DEM was used for georeferencing. Only one acquisition sample of a pushbroom array was considered. It is noted that changes in satellite attitude and imaging geometry should also be taken into account [17].

Yan et al. (2010) studied the current applications of computer games in design, visualization and education [18]. In order to expand the study and adaptability of computer games in design education, they presented the solution addressing the interoperability between games and building models to enhance architectural visualization and education. The implementation of BIM-Game has provided several connections between various fields such as architecture, engineering, computer science, visualization and game development. Yan et al. presents a framework for integrating BIM and games and the possibility of real-time, interactive and photorealistic virtual walking with a virtual user model to demonstrate the use and potentials of the framework.

Bettemir investigated the uncertainties in project planning for predetermining the time and heavy equipment requirements for excavation of a portion of a construction project and the

effects of these uncertainties on the outcome. The estimations were carried out by MCS. Uncertainties and risks related to weather, ground conditions and operator productivity are modeled with MCS and their effects on the completion time and equipment requirements are analyzed. The uncertainty range given by the machine efficiency tables were utilized to estimate the efficiency of the construction machine by generating random numbers within the stated range. When the obtained results were analyzed, it was revealed that the excavation time was the most difficult item to predict accurately. However, excavator machine-hour is the item with the least deviation [19].

Rüppel et al. (2011) presented a game approach based on BIM to study the impact of building condition on human behavior during the evacuation process. Since it is impossible to conduct rescue tests in burning buildings to study human behavior in real life, a game modeling with human behaviors that do not contradict the reality of existing evacuation simulation models is used to study situations that can be encountered in real life risk situations. It offers a serious game approach to study human behavior under difficult conditions. The capabilities of BIM in terms of fire and smoke are demonstrated in engineering simulations in order to create realistic serious game scenarios in a new and efficient way. BIM can contribute to preventing conflicts and identifying safety risks in advance, as it allows building occupants as well as rescue teams to demonstrate different scenarios as soon as possible without modeling the game scenario [20].

Beccali et al. (2011) implemented an algorithm in Microsoft Visual Basic for Application (VBA) to facilitate the management of the weather database and to allow users to easily create the file. In this way, the implementation of ISO 15927 4 is possible using a popular software tool such as Microsoft Excel or Access, without the use of any other expensive specialized software [21]. The generated application makes implementation of MCS easy and costless.

Goulding et al. (2014) presented a gaming environment supported by a web-based virtual cloud platform for integrated projects. The details of the Unified Software Development Process organized according to the determination of the specified cloud platform are also described. Goulding et al. (2014) presented a cloud-based VR Construction Site Simulator to provide trainees with a simulation experience of a future construction site. The simulator provides a risk-free environment where trainees can evaluate how their decisions will affect their work [22].

Natephra et al. (2017) developed a BIM-based lighting design feedback (BLDF) prototype system by proposing a method for realistic visualization of lighting condition and calculation of energy consumption. The system developed uses an interactive and immersive virtual reality (VR) environment to simulate daylight and artificial light illumination in buildings and visualized realistic VR scenes using digital cameras [23].

Sandoval developed a simulator prototype that integrates a Discrete Event Simulation (DES) model of a construction project and a BIM model representing the physical characteristics of the construction project within a commercial game engine [24]. The prototype allows the user to visualize how the duration of a single construction activity is affected by different input parameters. The presented simulator prototype is a promising approach to integrate DES and BIM models, allowing users to interact with them in a 3D environment. In addition, it allows users to import their own BIM models and conduct their own DES studies, which has many

potential uses in industry. Using a BIM model of a facility during the planning of its construction to test different strategies regarding the use of resources, the location of materials and the hierarchy of tasks will help stakeholders in the decision-making process. They also aimed to integrate the system to construction management education, as visualizing the results of these plans in a 3D environment could assist to understand the projects faster and better. In addition, it is intended to provide an alternative way to digitally study construction activities, extending the model to cover the entire facility instead of focusing on a single construction activity, allowing the user to manage a complete construction project considering the interaction between activities [24].

Yu-Cheng Lin (2018) developed a Database-driven VR/BIM-based Communication and Simulation (DVBCS) system integrated with BIM, game engine and virtual environment technologies for custom healthcare design in semi-immersed virtual environment (VR). The DVBCS system was applied in a case study of a cancer center design project in Taiwan to validate the system and demonstrate its effectiveness in practice. The results show that the DVBCS system is an effective visual communication and simulation platform for healthcare design. The advantage of the DVBCS system is not only to improve the communication efficiency between design teams and healthcare stakeholders, but also to facilitate visual interactions and facilitate the decision-making process when communicating in a 3D VR/BIM environment. It is envisaged that the effective use of the proposed DVBCS system will assist design teams and stakeholders in addressing future healthcare design work in a systematic way [25].

Omaran et al. (2019) [26] applied BIM and life cycle cost (LCC) analysis in the development of a BIM-game prototype. In the study, a BIM-game prototype was created that integrates BIM into LCC analysis visualization with Unity 3D game engine simulation, and the BIM-game prototype allows users to navigate through building environments with the ability to simulate LCC. The developed system is a combination of many fields such as architecture, engineering, computer programming and game development. A questionnaire was developed to evaluate the effectiveness of the BIM-game prototype and the results suggest that the system has significant potential for use in future projects.

Liu et al. (2020) considered added owner's penalty intensity, incentive pool size and allocation rate, BIM cost, Integrated Project Delivery (IPD) gain, liquidated damage, loss from BIM data source sharing, and other variable parameters to the triad evolutionary game model [27]. The triadic behavioral strategies are investigated with evolutionary game theory and the impact of variable parameters and stability of system equilibrium on the stability of evolutionary games is analyzed. In the study, the paths are simulated mathematically using MATLAB program. The dynamic mechanism of owner control is elucidated in the study.

METHOD

The methodology of this study involves analyzing the duration and uncertainty of construction projects by developing and implementing MCS using the C++ programming language. MCS consists of five steps.

First, all activities in the project and their interdependencies were modeled as an activity network. In this modeling process, the durations of each activity and the uncertainties associated with the durations were determined. In order to reflect the uncertainty in the activity durations, the normal distribution, which is one of the most frequently used probability distributions in the literature, was used. This distribution was used to simulate the possible variations of activity durations. Then, based on the network structure and activity data, a Monte Carlo simulation model was coded in C++ language. During the development of the model, efficient data structures and algorithms were used to improve the accuracy and efficiency of the simulation. The main operation of the simulation involves calculating the project duration at each trial by random sampling. In each simulation run, the completion times of all activities are calculated to obtain the project end time. This process was repeated up to a certain number of iterations. In this study, 500, 2500, 5000, 10000, as well as 25000 trials are examined and the results were recorded. Running the simulation with different iteration numbers played a critical role in assessing the accuracy of the results and the reliability of the model. As the number of iterations increased, the statistical reliability of the results produced by the simulation also increased, resulting in a more consistent analysis of the uncertainties related to the project duration. By analyzing the results, outputs such as the average estimate of the project duration and confidence intervals were obtained. These results can contribute to decision support systems by providing more precise information to project managers and planners.

The mean value and the standard deviation of the recorded trial results are computed. The mean value represents the most expected result in other words the most likely project duration. The standard deviation characterizes the expected variation in the project duration. Apart from the project duration, the simulation also provides the likelihoods of the activities to be critical. Throughout the simulation number of being a critical and noncritical activity are counted for each activity and the probability of being a critical activity is computed by dividing the number of being a critical activity to total number of simulation trials.

CASE STUDY

The network consisting of 6 activities given in Figure 1 is analyzed by MCS. The calculation of the completion time of the specified network with Monte Carlo Simulation is coded in C++ programming language.
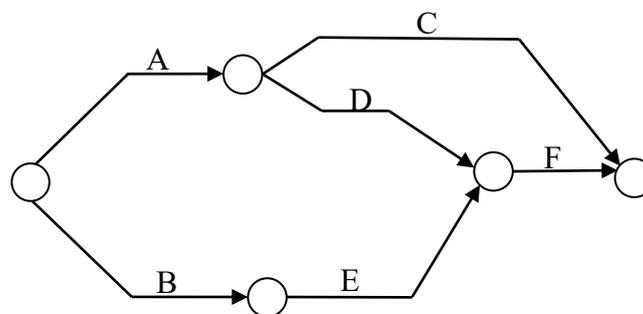

Figure 8: The activity on arow diagram of the analyzed project

A: 10 ± 3 days
B: 6 ± 2.5 days
C: 7 ± 4 days
D: 5 ± 1.5 days
E: 8 ± 3 days
F: 4 ± 2 days

X= A+C
Y= A+D+F
Z= B+E+F

The network given in Figure 1contains three paths. The first considered path represented as X is the A-C path, the second one illustrated as Y is the A-D-F path and the third one is demonstrated as Z is the B-E-F. Simulations were executed in user defined trial numbers. First of all, random numbers were generated for each activity as many times as the determined iteration number, and then the average of these numbers was calculated. By calculating the mean and standard deviations, the random values for each task were transformed according to the desired mean and standard deviation. To illustrate, since the duration of activity A is 10±3 days, the standard deviation of the generated numbers is set to 3.

Path 1: A + C

Path 2: A + D + F

Path 3: B + E + F

In each cycle, the completion times of these paths were compared and the path that took the longest time was selected. With these values, the average time required to complete the project and the standard deviation were calculated. The result is the expected duration of the project and the uncertainty of the estimated duration.
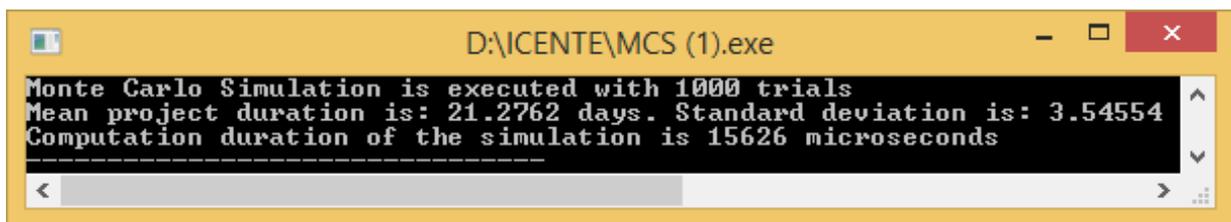

Figure 9: The screenshot of the developed simulator.

The simulation is conducted for 500, 2500, 5000, 10000, and 25000 trials. The software provides the computation duration which is slightly longer than 15 milliseconds for 1000 trials. The computation duration increases to 50 milliseconds for the 25000 trials but instead of duration, memory demand becomes more challenging issue when the number of trials is increased. The analysis results are presented in Table 1.

Table 2: Analysis results of the Monte Carlo Simulation

Number of Trials	Mean	St Dev
500	21.361	3.45915
2500	21.321	3.44499
5000	21.361	3.43962
10000	21.3588	3.48641
25000	21.3712	3.45542

The same project is analyzed by Program Evaluation and Review Technique (PERT) and the expected duration and its standard deviation is computed as 19 ± 3.90 days. As stated by [16] PERT provides shorter expected durations but higher standard deviation values than MCS. In this case, the difference between the duration estimations of PERT and MCS is high since the mean durations of the paths A-D and B-E are very close to each other and each path might be the longer path depending on the generated random numbers.

CONCLUSION

In this study, duration of a construction project is estimated by MCS which is conducted by executable software developed in C++. The analyses are repeated by selecting different trial number and similar mean and standard deviation values are obtained for the project completion duration.

The developed software utilizes different randomly generated numbers therefore the simulation results may deviate. Table 1 illustrate that the magnitude of the deviation is negligible and the mean project duration converges to 21.36 days. On the other hand, percentage of deviation is relatively higher for the standard deviation. Number of trials can be assigned as 5000 without losing significant computational accuracy for the construction projects with similar number of activity.

The developed software cannot form the network diagram itself. Formation of the network diagram by the activity precedence relationships can be added to the software as future study.

REFERENCES

[1] Kerzner, H. (2013). Project Management: A Systems Approach to Planning, Scheduling, and Controlling. John Wiley & Sons.

[2] PMI. (2017). A Guide to the Project Management Body of Knowledge (PMBOK Guide). Project Management Institute.

[3] Baccarini, D. (2006). The concept of project complexity–a review. International Journal of Project Management, 24(4), 313-326.

[4] Helton, J. C. (1993). Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliability Engineering & System Safety, 42(2-3), 327-367.

[5] Akgöl, K., & Günay, B. (2018). Kentiçi Yol Ağlarının Rasyonellik Hesabında Monte Carlo Yönteminin Kullanımı. Teknik Dergi, 29(1), 8153-8166.

[6] Köse, Ö., & Yanmaz, A. M. (2010). Köprü kenar ayaklarındaki oyulma güvenilirliği. Teknik Dergi, 21(101), 4919-4934.

[7] Doran, B., & Alacalı, S. (2017). Bağ Kirişlerinin Göçme Olasılıklarının Monte Carlo Simülasyonu ile Belirlenmesi. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 19(55), 92-99.

[8] Aydemir, C., & Zorbozan, M. (2012). Betonarme Kolonların Olası Eğilme Momenti Kapasitelerinin Belirlenmesi İçin Bir Yöntem. Teknik Dergi, 23(112), 5903-5930.

[9] Şengül, Ö. (2011). Klor İyonu Etkisindeki Betonarme Yapı Elemanlarının Dayanıklılığı İçin Olasılığa Dayalı Tasarım. Teknik Dergi, 22(107), 5409-5423.

[10] Birgönül, M. T., & Dikmen, İ. (1996). İnşaat projelerinin risk yönetimi. İMO teknik Dergi, 97, 1305-1326.

[11] Çalamak, M. (2017). Tanyeri Barajı örneğinde barajların taşkın nedeniyle üstten aşılma güvenilirliğinin belirlenmesi. Journal of the Faculty of Engineering and Architecture of Gazi University, 32(3), 965-975.

[12] Çavuş, T. F., & Yanıkoğlu, E. (2003). Karmaşık Sistemlerin Monte Carlo Yöntemi ile Güvenilirlik Analizi. Sakarya University Journal of Science, 7(3), 99-102.

[13] Öztekin, E. (2015). Cıvatalı Çelik Yapı Birleşimlerinde Cıvata Yerleşim Mesafelerinin Güvenilirliklerinin Monte Carlo Simülasyon Metodu İle Araştırılması. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 21(6), 213-223.

[14] Sağlam, B., & Bettemir, Ö. H. (2018). Estimation of duration of earthwork with backhoe excavator by Monte Carlo Simulation. Journal of Construction Engineering, 1(2), 85-94.

[15] Bettemir, Ö. H. (2006). Sensitivity and error analysis of a differential rectification method for CCD frame cameras and pushbroom scanners (Master's thesis, Middle East Technical University).

[16] Bettemir, O. (2020). Computation of Critical Path Probabilities by Modified PERT. Gazi University Journal of Science, 1-1.

[17] Bettemir, Ö. H. (2011). Prediction of georeferencing precision of pushbroom scanner images. IEEE transactions on geoscience and remote sensing, 50(3), 831-838.

[18] Yan, W., Culp, C., & Graf, R. (2011). Integrating BIM and gaming for real-time interactive architectural visualization. Automation in Construction, 20(4), 446-458.

[19]   Bettemir, Ö. H. (2009). Kazı ve hafriyat sürelerindeki belirsizliğin Monte Carlo Analizi ile tahmini. Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 14(2), 165-173.

[20]   Rüppel, U., & Schatz, K. (2011). Designing a BIM-based serious game for fire safety evacuation simulations. Advanced engineering informatics, 25(4), 600-611.

[21]   Beccali, M., Bertini, I., Ciulla, G., & Di, P. (2011, November). Software for weather databases management and construction of reference years. In Building Simulation 2011 (Vol. 12, pp. 1182-1186). IBPSA.

[22]   Goulding, J. S., Rahimian, F. P., & Wang, X. (2014). Virtual reality-based cloud BIM platform for integrated AEC projects. Journal of Information Technology in Construction, 19, 308-325.

[23]   Natephra, W., Motamedi, A., Fukuda, T., & Yabuki, N. (2017). Integrating building information modeling and virtual reality development engines for building indoor lighting design. Visualization in Engineering, 5, 1-21.

[24]   Sandoval, C. A. O., Tizani, W., & Koch, C. (2018). A method for discrete event simulation and building information modelling integration using a game engine. Advances in Computational Design, 3(4), 405-418.

[25]   Lin, Y. C., Chen, Y. P., Yien, H. W., Huang, C. Y., & Su, Y. C. (2018). Integrated BIM, game engine and VR technologies for healthcare design: A case study in cancer hospital. Advanced Engineering Informatics, 36, 130-145.

[26]   Omaran, S. M., Alghamdi, A. A., Alharishawi, S. C., & Hains, D. B. (2019, June). Integrating BIM and game engine for simulation interactive life cycle analysis visualization. In ASCE International Conference on Computing in Civil Engineering 2019 (pp. 120-128). Reston, VA: American Society of Civil Engineers.

[27]   Liu, K., Liu, Y., Zhou, H., Kou, Y., Ji, Q., & Li, D. (2021). Evolutionary game and numerical simulation of participants' collaborative behavior in integrated project delivery project. Alexandria Engineering Journal, 60(1), 373-385.

RESUME

Merve KAYA

Merve Kaya obtained her Bachelor's degree in Civil Engineering from İnönü University, Malatya, Turkey in 2020 and commenced her Master's studies at the same institution in 2023. Her academic research is centered on the integration of digital twin technology into the construction sector, with a particular emphasis on its role in project management and methodologies for model development. By advancing the understanding of digital twin concepts, she aims to elucidate its potential contributions to the construction industry and its transformative impact on engineering practices.

Prof. Dr. Önder Halis BETTEMİR

Prof. Dr. Önder Halis Bettemir Graduated from Middle East Technical University at the department of civil engineering in 2003. Completed master of science studies at the department of civil engineering on geodesy and photogrammetry discipline in 2006. Performed PhD studies at Construction Management division at the Middle East Technical University in 2009. Bettemir assigned as Assistant Prof. to the department of Civil Engineering of Yuzuncu Yil University in 2010 and studied in this university until 2014. Bettemir continued his academic career in İnönü University and gained Assoc. Prof. Degree in 2017. He became Professor in 2023. Main study subjects are Geographic Information Systems and optimization of construction schedules. Moreover, exact quantity take-off by Building Information Modelling is his current study subject.

Mehmet Fatih ACAR

Mehmet Fatih Acar obtained his Bachelor's degree in Civil Engineering from İnönü University, Malatya, Turkey in 2017 and commenced his Master's studies at the same institution in 2019 and still ongoing. His academic research is related with the uncertainties of the construction cost and duration estimations. He is also working as site engineer at a private construction company.
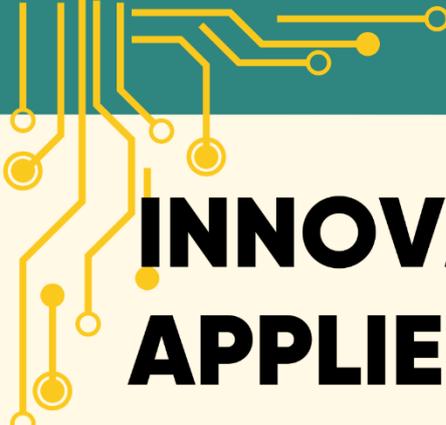
INNOVATIVE APPROACHES IN APPLIED ENGINEERING FIELDS

Engineering is the cornerstone of modern technological advancement, bridging creativity and functionality. "Innovative Approaches in Applied Engineering Fields" brings together the latest research, cutting-edge methodologies, and transformative solutions from leading professionals and academics.

FIELDS

- Artificial Intelligence
- Industry 4.0 Technologies
- Sustainable Engineering
- Real-world case studies in various engineering domains.

This book serves as a valuable resource for engineers, researchers, and students who aim to stay ahead in the rapidly evolving world of engineering. Uncover new horizons in innovation, efficiency, and sustainable solutions for a better future!